Bayesian analysis of incomplete time and cause of failure data

Chiranjit Mukhopadhyay[a,*], Asit P. Basu[b]

[a] *Indian Statistical Institute, 8-th Mile Mysore Road, R.V. College Post, Bangalore 560059, India*

[b] *Department of Statistics, University of Missouri, Columbia, MO 65201, USA*

Received 9 July 1993

Abstract

For series systems with k components it is assumed that the cause of failure is known to belong to one of the $2^k - 1$ possible subsets of the failure-modes. The theoretical time to failure due to k causes are assumed to have independent Weibull distributions with equal shape parameters. After finding the MLEs and the observed information matrix of $(\lambda_1, \ldots, \lambda_k, \beta)$, a prior distribution is proposed for $(\lambda_1, \ldots, \lambda_k)$, which is shown to yield a scale-invariant noninformative prior as well. No particular structure is imposed on the prior of β. Methods to obtain the marginal posterior distributions of the parameters and other parametric functions of interest and their Bayesian point and interval estimates are discussed. The developed techniques are illustrated using a numerical example.

Keywords: Series system; Competing risks; Weibull distribution; Masking; EM algorithm; Bayesian estimation

1. Introduction

Suppose a unit has k causes of failure $C_1, \ldots, C_k$. Here a unit may be a man-made machine or a biological system, a cause of failure may be a machine-component or a disease, and failure may mean a nonfunctional state or death. Let Y_i denote the theoretical time to failure of a unit due to C_i, $i = 1, \ldots, k$. Then the real time to failure of a unit is given by $U = \text{Minimum}\{Y_1, \ldots, Y_k\}$, if it fails as soon as one of the causes takes place.

This paper analyzes incomplete data that might arise while dealing with such kind of series systems. The most common form of incompleteness in observations on time to failure is right-censoring, in which one observes $U > u$. The other source of incompleteness is the deficiency of knowledge of the exact cause of failure C_i. The literature is rich when the cause of failure C_i is exactly known, which is the usual competing risks model, and we refer to David and Moeschberger (1978) for an overview and bibliographical references. But there are situations both in reliability theory and survival analysis in which the cause of failure might not exactly be known for the failed units.

Usher (1987) gives an example where the exact cause of failure may be unknown in an Engineering application. Think of a hardware failure of a large computer system in use. In such a situation, the objective being to get the system back on line as soon as possible, whenever the cause of failure is narrowed down to a small subset of components, say a circuit board for example, the entire subsystem is replaced by a new one. Thus the exact failing component may not be investigated any further. One can suppose similar scenarios to arise while dealing with complex mechanical (automobile for example) or electrical (power generator for instance) systems. Similarly the exact cause of death of a patient remains unknown without an autopsy in biomedical applications.

Following Guess et al. (1991), for a failed unit, we define the minimum random subset (MRS) as the set $\{C_{i_1},\dots,C_{i_l}\}\subseteq\{C_1,\dots,C_k\}$ of 'narrowed-down' possible cause(s) of failure.

With $k = 2$, Y_1 and Y_2 independent exponentials, and incomplete knowledge of the cause of failure, which is known as masking in the literature, MLEs are provided by Miyakawa (1984). He also deals with the problem nonparametrically for $k = 2$. Miyakawa does not address the issue of censoring. Dinse (1982) gives a nonparametric solution for an arbitrary k, but in his model either the cause of failure is exactly known or completely unknown, i.e. his MRS's are confined to either the whole set $\{C_1,\dots,C_k\}$ or the singletons. Usher and Hodgson (1988) solve the problem of finding the MLEs for the case of $k = 3$ independent exponentials with masking. There, however, they do not deal with censored observations. Doganaksoy (1991) extends the work of Usher and Hodgson (1988) (with $k = 3$ independent exponentials with masking) in the direction of finding approximate confidence intervals of the parameters, while Reiser et al. (1995) provide a Bayesian analysis for $k = 3$ independent exponentials. Berger and Sun (1993) render a Bayesian analysis for k independent Weibulls with unequal shape parameters, but they restrict themselves to the case of the cause of failure always being unknown (MRS $= \{C_1,\dots,C_k\}$).

In this paper, we assume that we have N k-component series systems with independent and identically distributed lifetime U. For a failed unit, together with its time to failure U we also observe that its cause of failure belongs to a MRS $\{C_{i_1},\dots,C_{i_l}\}\subseteq\{C_1,\dots,C_k\}$, while for a right-censored unit the only information we have is $U > u$. The assumed theoretical model is, the lifetime of the ith component $Y_i \sim \text{Weibull}(\lambda_i,\beta)$, for $i = 1,\dots,k$ and $Y_1,\dots,Y_k$ are independent. The analysis in the case of unequal shape parameters turns out to be quite different from the treatment given here. In particular, the choice of the prior distribution in the present problem, which is intuitively appealing, yields some analytical reduction in the posterior analysis and thus, although there are $k+1$ parameters, one never needs to numerically integrate in more than two dimensions to compute the marginal posterior quantities of the parameters and other parametric functions of interest, while in the case of unequal shape parameters the computations are done using a Gibbs sampler after an appropriate adjustment of Theorem 2.2 of Berger and Sun (1993). In Section 2, we derive the MLEs of $(\lambda_1, \ldots, \lambda_k, \beta)$ using the EM algorithm of Dempster et al. (1977), and give the expressions to compute the observed information matrix. This is primarily to lay the ground work for the computation of the joint posterior modes, and to obtain the asymptotic normal approximation of the posterior distribution of the parameters. For known β, the model is equivalent to that of k independent exponentials and serves to generalize the results of Reiser et al. (1995) and that of Usher and Hodgson (1988), who remark that the problem of finding the MLE for k independent exponentials with masking and an arbitrary k is intractable. In Section 3, we introduce the prior distribution for $(\lambda_1, \ldots, \lambda_k)$. A scale-invariant noninformative prior for $(\lambda_1, \ldots, \lambda_k)$ is also derived, which can be modeled by the proposed prior with some particular choices of the hyper-parameters. No specific form is assumed for the prior of β, but β and $(\lambda_1, \ldots, \lambda_k)$ are assumed to be independent a priori. In Section 4, we analyze the joint posterior distribution of the parameters and derive some of its useful properties, which are helpful in obtaining the marginal posterior distributions and the Bayesian estimates of the parameters and the parametric functions of interest. The parametric functions for which the marginal posterior distributions are obtained are π_i's, λ, $\bar{F}_i(\cdot)$'s, and $\bar{F}(\cdot)$; where π_i is the probability of failure due to cause C_i, $\lambda = \sum_{i=1}^{k} \lambda_i$, $\bar{F}_i(\cdot)$ is the survival function of Y_i, $i = 1, \ldots, k$, and $\bar{F}(\cdot)$ is the survival function of U. We illustrate the Bayesian methods to obtain the marginal posterior distributions, Bayes' estimates, posterior modes, and highest posterior density (abbreviated as HPD from now on) credible sets of the parameters and the parametric functions of interest using a set of simulated data in Section 5. Finally, we give some concluding remarks in Section 6.

2. Likelihood function

First note that for a failed unit with the observed time to failure u and MRS $\{C_i, \ldots, C_h\}$, its contribution to the likelihood function is given by

$$(\lambda_i + \cdots + \lambda_h)\, \beta\, u^{\beta-1}\, e^{-\lambda u^{\beta}},$$

while for a right-censored unit with $U > u$ its contribution to the likelihood function is $e^{-\lambda u^{\beta}}$. Let $u_1, \ldots, u_N$ denote the $N (= n - m)$ observations on U, of which m are right-censored. This notation of using u for both $U = u$ and $U > u$, although slightly abusive, is helpful for clarity.

The problem of writing the likelihood function lies in the algebraic formalism of ordering the MRSs. The Guess et al. (1991) form of the likelihood function does not yield a tractable form. Consider the problem of writing the contribution to the likelihood function for units having MRS of cardinality l, $l = 1, \ldots, k$. Let $q_l = \binom{k}{l}$.

For a fixed l there are q_l possible subsets of $\{C_1,\ldots,C_k\}$. These q_l possible subsets are ordered as follows.

Let

$$\delta_{jl}^{(i)} = \begin{cases} 1 & \text{if } C_i \in \text{MRS} \\ 0 & \text{if } C_i \notin \text{MRS} \end{cases} \quad \text{for } i = 1,\ldots,k;\ j = 1,\ldots,q_l;\ l = 1,\ldots,k.$$

Let $\kappa_{jl} = \delta_{jl}^{(1)} \cdots \delta_{jl}^{(k)}$ denote a k-bit binary integer. One can associate a unique κ_{jl} with each subset of $\{C_1,\ldots,C_k\}$ of cardinality l. Now order these q_l subsets according to the order $\kappa_{1l} > \cdots > \kappa_{q_l l}$, so that one can speak of the jth subset of cardinality l without any ambiguity.

Suppose there are n_{jl} units in the jth subset of cardinality l, $l = 1,\ldots,k$, with their observed times to failure being t_{jlr}, $r = 1,\ldots,n_{jl}$; $j = 1,\ldots,q_l$; $l = 1,\ldots,k$. Let $n = \sum_{l=1}^{k} \sum_{j=1}^{q_l} n_{jl}$, denote the total number of uncensored units $S = \sum_{l=1}^{k} \sum_{j=1}^{q_l} \sum_{r=1}^{n_{jl}} \log(t_{jlr})$; and $T(\beta) = \sum_{p=1}^{N} u_p^{\beta}$. Then the likelihood function can be written as (with the convention that $0^0 = 1$)

$$L(\lambda_1,\ldots,\lambda_k,\beta) = \left[\prod_{l=1}^{k} \prod_{j=1}^{q_l} \left(\sum_{i=1}^{k} \lambda_i \delta_{jl}^{(i)}\right)^{n_{jl}}\right] \beta^n \, e^{S(\beta-1)} \, e^{-\lambda T(\beta)}. \tag{2.1}$$

To find the MLEs of the parameters, if one differentiates log L and equates it to 0, the $(k+1) \times (k+1)$ system of equations do not yield a tractable form to solve, even numerically. So the EM algorithm is tried instead. In order to apply the algorithm, one should look at the complete data likelihood function. Since the incompleteness due to right-censoring does not pose any problem, the incompleteness is regarded as the lack of exact knowledge of the cause of failure. So for the complete data problem, let n_i denote the total number of units, failing due to C_i, $i = 1,\ldots,k$, so that $n = \sum_{i=1}^{k} n_i$. These n_i's are estimated by their expected values, given the incomplete data, in the E-step; and utilizing these expected values, the complete data likelihood is maximized in the M-step to yield the current estimates of the parameters; and one iterates between these two steps until the solution converges. With the other notations as above, the complete data likelihood is given by

$$L_C(\lambda_1,\ldots,\lambda_k,\beta) = \left[\prod_{i=1}^{k} \lambda_i^{n_i}\right] \beta^n \, e^{S(\beta-1)} \, e^{-\lambda T(\beta)}. \tag{2.2}$$

In order to formulate the iteration, a few notations are introduced. Let $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_k)'$; $\gamma_{jl}(\boldsymbol{\lambda}) = \sum_{i=1}^{k} \lambda_i \delta_{jl}^{(i)}$; $x_{jl}^{(i)} = n_{jl} \frac{\lambda_i \delta_{jl}^{(i)}}{\gamma_{jl}(\boldsymbol{\lambda})}$, and $T'(\beta) = \sum_{p=1}^{N} \log(u_p) u_p^{\beta}$.

E-Step:

$$n_i = \sum_{l=1}^{k} \sum_{j=1}^{q_l} x_{jl}^{(i)} \quad \text{for } i = 1,\ldots,k. \tag{2.3}$$

M-Step:

$$\frac{\partial \log L_C}{\partial \lambda_i} = 0 \;\Rightarrow\; \lambda_i = \frac{n_i}{T(\beta)} \quad \text{for } i = 1,\ldots,k, \tag{2.4}$$

$$(2.4) \text{ and } \frac{\partial \log L_C}{\partial \beta} = 0 \;\Rightarrow\; \beta = \frac{nT(\beta)}{nT'(\beta) - ST(\beta)}. \tag{2.5}$$

Note that the solution of (2.5) is independent of the values of n_i's obtained in (2.3) and the λ_i's as well. Let us denote (2.5) by $\beta = \varphi(\beta)$. One solves (2.5) iteratively; by first choosing an initial value β_0 of β, obtaining β_{l+1} as $\varphi(\beta_l)$, and continuing until a desired degree of accuracy is achieved. Let the solution of (2.5) be denoted by $\hat{\beta}$. Thus $\hat{\beta}$ does not depend on the EM iteration and is repeatedly used in (2.4) of the M-step. Once $\hat{\beta}$ is obtained, one starts with an initial value $\boldsymbol{\lambda}_0$ of $\boldsymbol{\lambda}$. In the iterative step; given $\boldsymbol{\lambda}_l$, first the n_i's are estimated using (2.3), and then these n_i's are utilized in (2.4) to obtain $\boldsymbol{\lambda}_{l+1}$, the current estimate of $\boldsymbol{\lambda}$. As usual the iteration continues until a satisfactory solution is reached. We suppress the iterative subscript l for the current estimate of $\boldsymbol{\lambda}$ and let the solution of (2.4) denote by $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1,\ldots,\hat{\lambda}_k)'$, at any given stage of iteration. Let us denote the parameter vector $(\boldsymbol{\lambda}', \beta)'$, by $\boldsymbol{\theta}$, and let its estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\lambda}}', \hat{\beta})'$.

In order to show that the solution of (2.4) and (2.5), at any given stage of iteration, indeed maximizes L_C, one must show that the matrix of the second partial derivatives evaluated at $\hat{\boldsymbol{\theta}}$, $[\partial^2 \log L_C/\partial \boldsymbol{\theta}^2 \,|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}] = \boldsymbol{H}$ (say), is negative definite. Denote the top left $l \times l$ submatrix of $\boldsymbol{H}$ by $\boldsymbol{H}_l$, $l = 1,\ldots,k$. It can be shown that

$$|\boldsymbol{H}_l| = (-1)^l \left[\prod_{i=1}^{l} \frac{n_i}{\hat{\lambda}_i^2}\right].$$

It can also be shown that

$$|\boldsymbol{H}| = (-1)^{k+1} \frac{1}{nT^2(\hat{\beta})}[(nT'(\hat{\beta}) - ST(\hat{\beta}))^2 + n^2(T(\hat{\beta})T''(\hat{\beta}) - T'^2(\hat{\beta}))] \times \left[\prod_{i=1}^{k} \frac{n_i}{\hat{\lambda}_i^2}\right],$$

where $T''(\beta) = \sum_{p=1}^{N} [\log(u_p)]^2 u_p^\beta$. Straightforward algebra yields

$$T(\beta)T''(\beta) - T'^2(\beta) = \sum\sum_{p<p'} (\log u_p - \log u_{p'})^2 u_p^\beta u_{p'}^\beta > 0 \quad \forall \beta > 0.$$

Hence $\hat{\boldsymbol{\theta}}$ maximizes (2.2) and is the MLE of $\boldsymbol{\theta}$.

For the convergence of the EM iteration, a sufficient condition on the n_{ij}'s is given below.

Condition C. $\exists$ an $1 \leqslant l \leqslant k \ni n_{lj} > 0 \;\forall j = 1,\ldots,q_l$.

Now we address the convergence of the EM algorithm involving the iteration between (2.3) and (2.4) only. The convergence issue of the EM iteration has been studied by Jeff Wu (1983). The incomplete data likelihood given in (2.1) above is unimodal if Condition C is satisfied. The likelihood of the complete data, given the incomplete observations, is the joint p.m.f. of k integer-valued random variables $N_1, \ldots, N_k$, where each N_i is the sum of $2^k - 1$ independent multinomial random variables, having multinomial probabilities of the type $\lambda_i \delta_{ji}^{(i)} / \gamma_{ji}(\lambda)$. Thus it is clear that the gradient vector of the expected log likelihood of this complete data given the incomplete observations is continuous. Thus Corollary 1 of Jeff Wu applies, and guarantees the convergence of the EM iteration.

We are interested in the asymptotic normal approximation of the posterior distribution of θ. For that purpose, we need to evaluate the second derivatives of the likelihood function. For $i, i' \in \{1, \ldots, k\}$,

$$\frac{\partial^2 \log L}{\partial \lambda_i \partial \lambda_{i'}} = \sum_{l=1}^{k} \sum_{j=1}^{q_l} \frac{n_{jl} \delta_{jl}^{(i)} \delta_{jl}^{(i')}}{\gamma_{jl}(\lambda)^2}, \qquad \frac{\partial^2 \log L}{\partial \lambda_i \partial \beta} = -T'(\beta),$$

$$\frac{\partial^2 \log L}{\partial \beta^2} = -\left(n\beta^{-2} + \lambda T''(\beta)\right).$$

Let $I(\hat{\theta})$ denote the observed information matrix, where

$$I(\hat{\theta}) = -\begin{bmatrix} \left(\frac{\partial^2 \log L}{\partial \lambda_i \partial \lambda_{i'}}\right)\Big|_{\theta=\hat{\theta}} & \left(\frac{\partial^2 \log L}{\partial \lambda_i \partial \beta}\right)\Big|_{\theta=\hat{\theta}} \\ \left(\frac{\partial^2 \log L}{\partial \lambda_i \partial \beta}\right)'\Big|_{\theta=\hat{\theta}} & \left(\frac{\partial^2 \log L}{\partial \beta^2}\right)\Big|_{\theta=\hat{\theta}} \end{bmatrix} \tag{2.6}$$

3. Prior distribution

In this section, we introduce a prior distribution $\pi(\lambda)$ for λ. In the analysis that follows, the prior on β is arbitrary and is denoted by $\pi(\beta)$. It is also assumed that the prior of λ is independent of the prior of β. The basic idea of the prior is taken from Peña and Gupta (1990), where they use it to find Bayesian estimates of the parameters of the bivariate exponential model of Marshall and Olkin (1967) under both series and parallel sampling. Algebraically the following prior is just a k-dimensional generalization of their prior. Although the motivation for this prior in this problem is quite different than theirs.

Let π_i denote the probability of failure due to cause C_i. In the present model, $\pi_i = \lambda_i/\lambda$. It is usually possible to express a prior opinion about $(\pi_1, \ldots, \pi_k)$ for an analyst of a series system. For instance, it is likely for a design engineer of a system to have a prior opinion about the probability of system failure due to a particular component or failure-mode of the system. A doctor may have a prior idea about the likelihoods of the different causes of death of a patient. We assume that $(\pi_1, \ldots, \pi_k)$ has a Dirichlet distribution with parameters $(\alpha_1, \ldots \alpha_k)$. This distribution has a density with respect to the Leabsgue measure in $\Re^{k-1}$ which is given by

$$\pi(\pi_1,\dots,\pi_{k-1}) = \begin{cases} \dfrac{\Gamma(\alpha_0)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\prod_{i=1}^{k}\pi_i^{\alpha_i-1} & \text{for } \pi_i \geqslant 0 \text{ and } \sum_{i=1}^{k-1}\pi_i \leqslant 1, \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

where $\alpha_0 = \sum_{i=1}^{k}\alpha_i$ and $\pi_k = 1 - \sum_{i=1}^{k-1}\pi_i$. The density in (3.1) is denoted by $\mathscr{D}(\alpha_1,\dots,\alpha_k)$. The prior on $(\pi_1,\dots,\pi_{k-1})$ is thus restricted to its natural domain, the simplex $\mathscr{S} = \{(x_1,\dots,x_{k-1})' \in \Re^{k-1}: x_1 \geqslant 0,\dots,x_{k-1} \geqslant 0, \sum_{i=1}^{k-1} x_i \leqslant 1\}$. Other than the flexibility of the shape and mathematical convenience, Dirichlet distribution also has an appealing property for appraising a prior distribution for $(\pi_1,\dots,\pi_{k-1})$. One might subjectively compare the relative likelihoods of failures due to different causes, and assign proportional values to the α_i's. For example, if C_i is felt to be x times as likely to be the cause of failure as compared to C_j one may take $\alpha_i = x\alpha_j$, as the expected prior guess. Then one must do a consistency check in a line similar to Berger (1985, p. 78). Since there are only finitely many causes, the problem of eliciting consistent α_i values is much easier in this case.

Observe that this prior is already scale-invariant. If the unit of measurements of the theoretical life distributions $\{Y_1,\dots,Y_k\}$ is changed, that is if instead $\{cY_1,\dots,cY_k\}$ are the theoretical life distributions for any $c > 0$, then π_i's remain unchanged and thus the same prior is applicable. Moreover a noninformative state of mind can be modeled by giving a uniform distribution on the natural domain of $(\pi_1,\dots,\pi_{k-1})$ viz. $\mathscr{S}$. This amounts to assigning $\alpha_1 = \cdots = \alpha_k = 1$.

Now assume that $(\pi_1,\dots,\pi_k)$ is independent of λ, and put a Gamma prior on λ with parameters (α,γ). That is,

$$\pi(\lambda) = \frac{\gamma^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda\gamma}, \quad \lambda > 0. \tag{3.2}$$

The density in (3.2) is denoted by $\mathscr{G}(\alpha,\gamma)$. At this point it should be mentioned that given β, a scale-invariant noninformative prior for λ can be modeled with $\alpha = 0$ and $\gamma = 0$ in (3.2) above (ignoring the normalizing constant $\gamma^{\alpha}/\Gamma(\alpha)$), as is explained below.

The following argument is conditional on β. The relevant information about λ arises through the p.d.f. (or c.d.f.) of U only. Because in this model, the cause of failure, the MRS of any unit, is independent of the failure time U, and the probability distribution of any MRS depends only on $(\pi_1,\dots,\pi_k)$. So consider the p.d.f. of U and $V = cU$ for an arbitrary $c > 0$ for scale-invariance.

$$f_U(u) = \lambda\beta u^{\beta-1}e^{-\lambda u^{\beta}},$$

$$f_V(v) = \eta\beta v^{\beta-1}e^{-\eta v^{\beta}}, \quad \text{where } \eta = \lambda/c^{\beta}.$$

Notice that this means, if the unit of measuring lifetime is changed, still the problem has the same formal structure. Thus if $\pi(\lambda)$ denotes the prior density for the (U,λ) setup and $\pi^*(\eta)$ denotes the prior density for (V,η), then the equality $P_\pi(\lambda \in A) =$

$P_{\pi^*}(\eta \in A)$ must hold $\forall A \subseteq (0,\infty)$. Since $\eta = \lambda/c^\beta$ it should also be true that $P_{\pi^*}(\eta \in A) = P_\pi(\lambda \in c^\beta A)$, where $c^\beta A = \{c^\beta z : z \in A\}$. Hence $P_\pi(\lambda \in A) = P_\pi(\lambda \in c^\beta A)$. This should hold $\forall c > 0$. The last equality can be written as

$$\int_A \pi(\lambda)\,\mathrm{d}\lambda = \int_{c^\beta A} \pi(\lambda)\,\mathrm{d}\lambda = \int_A \pi(c^\beta \lambda) c^\beta\,\mathrm{d}\lambda.$$

Now for the above equality to hold $\forall A$, $\pi(\lambda) = c^\beta \pi(c^\beta \lambda)$ must be true $\forall \lambda > 0$ and $\forall c > 0$. So in particular, taking $\lambda = c^{-\beta}$ one gets $\pi(c^{-\beta}) = c^\beta \pi(1)$. Now since the equality must hold $\forall c > 0$, given β, take $c = \lambda^{-1/\beta}$. Hence $\pi(\lambda) \propto \lambda^{-1}$.

With the above priors on $(\pi_1,\ldots,\pi_{k-1})$ and λ given in (3.1) and (3.2), respectively, the prior on $\boldsymbol{\lambda}$ is obtained as follows:

$$\pi(\pi_1,\ldots,\pi_{k-1},\lambda) = \left[\frac{\Gamma(\alpha_0)}{\prod_{i=1}^k \Gamma(\alpha_i)} \frac{\gamma^\alpha}{\Gamma(\alpha)}\right] \left[\prod_{i=1}^k \pi_i^{\alpha_i - 1}\right] [\lambda^{\alpha-1} e^{-\gamma\lambda}],$$
$$(\pi_1,\ldots,\pi_{k-1}) \in \mathscr{S}, \quad \lambda > 0. \tag{3.3}$$

Note that (3.3) is a density with respect to Lebesgue measure in $\Re^k$. Performing a nondegenerate transformation from $(\pi_1,\ldots,\pi_{k-1},\lambda) \to (\lambda_1,\ldots,\lambda_k)$ one can obtain the prior for $\boldsymbol{\lambda}$. The Jacobian of this transformation with $\pi_i = \lambda_i/(\lambda_1 + \cdots + \lambda_k)$ and $\lambda = \lambda_1 + \cdots + \lambda_k$ is $\lambda^{-(k-1)}$, which can be seen by finding the differential matrix of the transformation, and its determinant is found by lower-triangularizing it by replacing its ith row by ith row$+\lambda_i/\lambda^2 \times k$th row, for $i = 1,\ldots,k-1$. Hence the prior density of $\boldsymbol{\lambda}$ is given by

$$\pi(\boldsymbol{\lambda}) = \left[\frac{\Gamma(\alpha_0)}{\prod_{i=1}^k \Gamma(\alpha_i)} \frac{\gamma^\alpha}{\Gamma(\alpha)}\right] \left[\prod_{i=1}^k \lambda_i^{\alpha_i - 1}\right] [\lambda^{\alpha-\alpha_0} e^{-\gamma\lambda}], \quad \lambda_1 > 0,\ldots,\lambda_k > 0. \tag{3.4}$$

The prior density in (3.4) is called the Dirichlet Gamma density with parameters $(\alpha_1,\ldots,\alpha_k,\alpha,\gamma)$ and is denoted by $\mathscr{DG}(\alpha_1,\ldots,\alpha_k,\alpha,\gamma)$. As a reference prior, one can use the scale-invariant noninformative prior $\mathscr{DG}(1,\ldots,1,0,0)$ (ignoring the normalizing constant), denoted by $\pi_I(\boldsymbol{\lambda})$, which is $\propto 1/\lambda^k$.

For later reference, for $\boldsymbol{\lambda} \sim \mathscr{DG}(\alpha_1,\ldots,\alpha_k,\alpha,\gamma)$ we give the expression for the moments of $\boldsymbol{\lambda}$ below.

$$E[\lambda_i] = \frac{\alpha_i \alpha}{\alpha_0 \gamma} \quad \text{for } i = 1,\ldots,k, \tag{3.5}$$

$$V[\lambda_i] = \frac{\alpha_i \alpha}{\alpha_0 \gamma^2}\left[\frac{\alpha_i + 1}{\alpha_0 + 1}(\alpha + 1) - \frac{\alpha_i}{\alpha_0}\alpha\right], \qquad \mathrm{Cov}(\lambda_i,\lambda_j) = \frac{\alpha_i \alpha_j \alpha}{\alpha_0 \gamma^2}\left[\frac{\alpha+1}{\alpha_0+1} - \frac{\alpha}{\alpha_0}\right].$$

4. Posterior analysis

As before, let $\boldsymbol{\theta} = (\boldsymbol{\lambda}', \beta)'$ denote the parameter vector. The prior density of $\boldsymbol{\theta}$ is denoted by $\pi(\boldsymbol{\theta})$ and is given below.

$$\pi(\theta) = \pi(\lambda)\pi(\beta), \tag{4.1}$$

where $\pi(\lambda)$ is as given in (3.4) above and $\pi(\beta)$ is an arbitrary prior on β. We denote the data by $\boldsymbol{D}$ and are interested in the posterior distributions of the parameters and some parametric functions given $\boldsymbol{D}$. For any parameter or parametric function ϕ, its posterior density is generically denoted by $\pi(\phi|\boldsymbol{D})$. Multiplying the likelihood function given in (2.1) by the prior given in (4.1), we get

$$\pi(\theta|\boldsymbol{D}) \propto \left[\prod_{i=1}^{k} \lambda_i^{n_{1i}+\alpha_i-1}\right] \left[\prod_{j=2}^{k-1} \prod_{l=1}^{q_j} \left(\sum_{i=1}^{k} \lambda_i \delta_{jl}^{(i)}\right)^{n_{jl}}\right] \lambda^{n_{1k}+\alpha-\alpha_0} e^{-(T(\beta)+\gamma)\lambda} g(\beta), \tag{4.2}$$

where $g(\beta) = \beta^n e^{S\beta}\pi(\beta)$. We first find the marginal posterior density of β by integrating (4.2) above with respect to $d\lambda_1 \cdots d\lambda_k$. Observe that since n_{jl}'s are non-negative integers, one can expand $(\sum_{i=1}^{k} \lambda_i \delta_{jl}^{(i)})^{n_{jl}}$ using a multinomial expansion, and then take the two products on j and l respectively, and finally end up with a $2^k - (2+k)$ fold summation. Now if each term within this summation is multiplied by $[\prod_{i=1}^{k} \lambda_i^{n_{1i}+\alpha_i-1}]\lambda^{n_{1k}+\alpha-\alpha_0} e^{-(T(\beta)+\gamma)\lambda}$, each one of them gets a form of a $\mathscr{DG}$ density. Integrating these terms one can get the marginal posterior density of β. But notice that it is not necessary to evaluate the integrals of each term explicitly. Let a typical term be denoted by $\mathscr{DG}(\alpha_1^{(p)}, \ldots, \alpha_k^{(p)}, \alpha^{(p)}, T(\beta)+\gamma)$, where $\alpha^{(p)}$ and $\alpha_i^{(p)}$, $i = 0, \ldots, k$, denote the corresponding posterior parameter values of a $\mathscr{DG}$ density. For $i = 1, \ldots, k$, $\alpha_i^{(p)}$ depends on the dummy indices of the $2^k - (2+k)$ fold summations, but that dependence is suppressed for the sake of clarity. Also notice that $\alpha_i^{(p)}$'s are free of θ. Thus the only factor in the integral involving β is $(T(\beta)+\gamma)^{-\alpha^{(p)}}$. Hence we need to evaluate $\alpha^{(p)}$. From (3.4), for a $\mathscr{DG}(\alpha_1, \ldots, \alpha_k, \alpha, \gamma)$ density, the exponent of $\lambda = \alpha - \alpha_0$. But for each term in the summation the exponent of λ is constant, viz. $n_{1k} + \alpha - \alpha_0$, and $\alpha_0^{(p)} = \sum_{i=1}^{k} \alpha_i^{(p)} = \sum_{l=1}^{k-1} \sum_{j=1}^{q_l} n_{jl} + \alpha_0$. Hence $\alpha^{(p)} = \alpha_0^{(p)} + n_{1k} + \alpha - \alpha_0 = n + \alpha$. Thus we get

$$\pi(\beta|\boldsymbol{D}) \propto (T(\beta)+\gamma)^{-(n+\alpha)}\beta^n e^{\beta S}\pi(\beta). \tag{4.3}$$

From the preceding argument, it is also clear that $\pi(\lambda|\beta, \boldsymbol{D})$ is a finite mixture of $\mathscr{DG}$ densities. Notice further that in each of the terms in the mixtures, the term involving λ is $\lambda^{n_{1k}+\alpha-\alpha_0} e^{-(T(\beta)+\gamma)\lambda}$. Hence with the transformation $\pi_i = \lambda_i/\lambda$, $i = 1, \ldots, k-1$ and $\lambda = \sum_{i=1}^{k} \lambda_i$ one can get (in a similar way the $\mathscr{DG}$ density in (3.4) is derived from (3.3)) that $\pi(\pi_1, \ldots, \pi_{k-1}|\boldsymbol{D})$ is a mixture of Dirichlet densities and is independent of λ, and hence of β. The marginal posterior density of λ, conditional on β, is $\mathscr{G}(n+\alpha, T(\beta)+\gamma)$. In order to get the marginal posterior density of λ_i for $i = 1, \ldots, k$, one first obtains the joint posterior density of (π_i, λ) conditional on β. This is accomplished by noticing that π_i is independent of λ and β, the posterior density of π_i is a mixture of Beta densities, and that of λ conditional on β is a Gamma density. Now after making the transformation, $(\lambda_i, \pi_i) = (\pi_i\lambda, \pi_i)$ and numerically integrating π_i out from this bivariate density one gets the posterior density of λ_i conditional on β. Finally this conditional density is multiplied by $\pi(\beta|\boldsymbol{D})$ and is numerically integrated with respect to β in order to obtain

$\pi(\lambda_i|\boldsymbol{D})$. The above mentioned properties of the posterior distribution, for the purpose of marginalization of (4.2), which is a major problem with a parameter space of arbitrary dimension due to multidimensional numerical integrations, is summarized below.

P1. $\pi(\beta|\boldsymbol{D})$ is as in (4.3) but a one-dimensional numerical integration is required to evaluate the normalizing constant.
P2.

$$\pi(\lambda|\boldsymbol{D}) = \frac{\lambda^{n+\alpha-1}}{\Gamma(n+\alpha)} \int \left[(T(\beta)+\gamma)^{(n+\alpha)} e^{-(T(\beta)+\gamma)\lambda} \pi(\beta|\boldsymbol{D})\right] \mathrm{d}\beta \quad \text{for } \lambda > 0$$

and thus is obtained by a one-dimensional numerical integration.
P3. $\pi(\pi_1,\ldots,\pi_{k-1}|\boldsymbol{D})$ is a finite mixture of $\mathscr{D}(\alpha_1^{(p)},\ldots,\alpha_k^{(p)})$ densities. In particular, for $i = 1,\ldots,k$, $\pi(\pi_i|\boldsymbol{D})$ is a mixture of $\mathscr{B}(\alpha_i^{(p)}, (n - n_{1k} + \alpha_0) - \alpha_i^{(p)})$ (Beta) densities, and thus no numerical integration is necessary. Moreover, π_i is independent of (λ,β) a posteriori.
P4. For $i = 1,\ldots,k$ and $\lambda_i > 0$,

$$\pi(\lambda_i|\boldsymbol{D}) = \frac{\lambda_i^{n+\alpha-1}}{\Gamma(n+\alpha)} \int \left[(T(\beta)+\gamma)^{(n+\alpha)} \pi(\beta|\boldsymbol{D}) \int \left[\frac{\pi(\pi_i|\boldsymbol{D})}{\pi_i^{n+\alpha}} e^{-(T(\beta)+\gamma)(\lambda_i/\pi_i)}\right] \mathrm{d}\pi_i\right] \mathrm{d}\beta$$

and thus is obtained by a two-dimensional numerical integration.

Since the model is regular, the usual asymptotic arguments of Johnson (1970) or Ghosh et al. (1982) to approximate $\pi(\boldsymbol{\theta}|\boldsymbol{D})$ by a $(k-1)$-variate normal distribution via a Laplace approximation hold true. Note that we have already shown that $\hat{\boldsymbol{\theta}}$, the MLE of $\boldsymbol{\theta}$, exists in Section 2. Suppose that $\boldsymbol{I}(\hat{\boldsymbol{\theta}}) = \boldsymbol{T}'(\hat{\boldsymbol{\theta}})\boldsymbol{T}(\hat{\boldsymbol{\theta}})$, (where $\boldsymbol{I}(\hat{\boldsymbol{\theta}})$ is as given in (2.6)); $(\xi_1,\ldots,\xi_{k+1})' = \boldsymbol{\xi} = \boldsymbol{T}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$; $\boldsymbol{\theta}_0$ the true unknown value of $\boldsymbol{\theta}$ from which the data $\boldsymbol{D}$ are generated; and $\mathbf{P}_{\boldsymbol{\theta}_0}$ the corresponding probability measure. Also let $\Phi(\cdot)$ denote the standard normal c.d.f.. Then following Johnson (1970) the following probability statement can be made:

Suppose the prior $\pi(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$ is twice differentiable in a neighborhood of $\hat{\boldsymbol{\theta}}$. Then given $\varepsilon > 0$ and $\eta > 0$, $\exists N_0 \ni \forall N \geqslant N_0$

$$\mathbf{P}_{\boldsymbol{\theta}_0}\left[\left|P(\xi_1 \leqslant \zeta_1,\ldots,\xi_{k+1} \leqslant \zeta_{k+1}|\boldsymbol{D}) - \prod_{i=1}^{k+1} \Phi(\zeta_i)\right| < \eta \text{ uniformly in } \zeta_1,\ldots,\zeta_{k+1}\right] > 1 - \varepsilon.$$

Note that the above statement simply means that for large n one can approximate $\pi(\boldsymbol{\theta}|\boldsymbol{D})$ by a $\mathcal{N}_{k+1}(\hat{\boldsymbol{\theta}}, [\boldsymbol{I}(\hat{\boldsymbol{\theta}})]^{-1})$ p.d.f., for any regular prior $\pi(\boldsymbol{\theta})$, as is given in Berger (1985, p. 224). Under such circumstances, the marginalization of $\pi(\boldsymbol{\theta}|\boldsymbol{D})$ takes place in the line of usual multivariate normal theory and the HPD credible sets for the λ_i's or β are available directly, for example, in Berger (1985, p. 143), without any numerical methods.

Now we turn to the problem of estimating $\boldsymbol{\theta}$. The two most popular Bayesian estimates are the posterior mean, which is the Bayes' estimate for any symmetric quadratic loss, and posterior mode. The posterior mean of β is found by numerical integration using (4.3). To find the posterior mean of the λ_i's, one must first find it conditional on β, which from (3.5) can be seen to be $[(n+\alpha)/[(n-n_{1k}+\alpha_0)(T(\beta)+\gamma)]] \sum w\alpha_i^{(r)}$, where the '$\sum$' is on the $2^k-(2+k)$ fold summation and w is a constant depending on the dummy indices of the summation. One then numerically integrates out β with respect to $\pi(\beta|\boldsymbol{D})$. Notice that all the numerical integrations are in one dimension. To find the joint posterior mode of $\boldsymbol{\theta}$, one follows the same EM iteration as in Section 2. The E-step as given in equation (2.3) remains unchanged but the M-step is to be modified as follows:

M-step for the posterior mode:

$$\lambda_i = \frac{(n-k+\alpha)(n_i+\alpha_i-1)}{(n-k-\alpha_0)(T(\beta)+\gamma)}, \tag{4.4}$$

$$\beta = \frac{n(T(\beta)+\gamma)\pi(\beta)}{(n-k+\alpha)T'(\beta)\pi(\beta)-S\pi(\beta)(T(\beta)+\gamma)-(T(\beta)+\gamma)\pi'(\beta)}, \tag{4.5}$$

where $\pi'(\beta)$ is the derivative of $\pi(\beta)$. Note that like (2.5), (4.5) is also solved iteratively and is independent of (2.3) and (4.4). The maximization and convergence issues can be carried out in a line similar to Section 2. The HPD credible sets of the parameters are found numerically after obtaining the marginal posterior densities.

Finally, we give the expression for the posterior densities of the theoretical survival functions due to the k different causes, and the real survival function of a unit. In this model, the survival function of Y_i is given by $\bar{F}_i(t) = e^{-\lambda_i t^\beta}$ for $t>0$ and $i=1,\ldots,k$, and that of U is given by $\bar{F}(t) = e^{-\lambda t^\beta}$ for $t>0$. For a fixed $t>0$, let us denote the component reliabilities $\bar{F}_i(t)$'s by ρ_i, and the system reliability $\bar{F}(t)$ by ρ. Then using **P4**, for $i=1,\ldots,k$, and $0<\rho_i<1$,

$$\pi(\rho_i|\boldsymbol{D}) = \frac{(-\log\rho_i)^{(n+\alpha-1)}}{\rho_i\Gamma(n+\alpha)} \times \int\left[(T(\beta)+\gamma)^{(n+\alpha)}t^{-(n+\alpha)\beta}\pi(\beta|\boldsymbol{D}) \times \int\left[\frac{\pi(\pi_i|\boldsymbol{D})}{\pi_i^{n+\alpha}} e^{(T(\beta)+\gamma)(t^{-\beta}\log\rho_i)/\pi_i}\right]\mathrm{d}\pi_i\right]\mathrm{d}\beta \tag{4.6}$$

and is obtained by a two-dimensional numerical integration. Similarly using **P2**, for $0<\rho<1$,

$$\pi(\rho|\boldsymbol{D}) = \frac{(-\log\rho)^{(n+\alpha-1)}}{\rho\Gamma(n+\alpha)} \times \int[(T(\beta)+\gamma)^{(n+\alpha)}t^{-(n+\alpha)\beta}\pi(\beta|\boldsymbol{D})e^{(T(\beta)+\gamma)(t^{-\beta}\log\rho)}]\mathrm{d}\beta \tag{4.7}$$

and is obtained by a one-dimensional numerical integration. The posterior mean, mode and the HPD credible sets are found numerically and is illustrated in the next section.

5. Numerical illustration

In this section, we illustrate the above methods by using a set of simulated data. It is assumed that we have a three-component series system with Y_1 ~Weibull(2,2), Y_2 ~Weibull(3,2), and Y_3 ~ Weibull(4,2). Thirty observations are generated independently on each of Y_1, Y_2 and Y_3. To incorporate censoring, 30 observations are also generated independently from a Weibull(1.875,2) distribution to accommodate about 17% censoring. Let this censoring random variable be denoted by C. If for a unit, Minimum$\{Y_1, Y_2, Y_3\} > C$, then it is censored. However, as in Section 2, we denote the observations on U as Minimum$\{Y_1, Y_2, Y_3, C\}$, with the understanding that if the observation is censored, then C is observed (i.e. $U > C$), while for an uncensored observation $U =$ Minimum$\{Y_1,\ Y_2,\ Y_3\}$. In order to distinguish between a censored observation with an uncensored one, a censoring indicator variable I is introduced, with the interpretation of a value of 1 to be censored, while a value of 0 to indicate an uncensored observation. In order to mask the cause of failure, 30 observations on a masking variable M is generated independently of Y_1, Y_2, Y_3, and C, with M taking values 1, 2 and 3 with the respective probabilities 0.5, 0.4 and 0.1. The value of M is ignored for the censored observations. The value of M determines the cardinality of the MRS of a unit. For $M = 1$, the MRS is a singleton and the cause of failure is taken to be i, so that $Y_i =$ Minimum$\{Y_1, Y_2, Y_3\}$. For $M = 2$, the MRS is a doubleton and to get the exact sets, 30 discrete uniform random variables, say J, taking values in $\{1,2,3\}$ are also independently generated, and the corresponding MRS is taken to be $\{1,2,3\} - \{J\}$ for the observations with $I = 0$ and $M = 2$. This is done to ensure that the cause of failure masking is independent of the lifetime random variables. For $M = 3$, the MRS=$\{1,2,3\}$. The generated data, as described above, is given in Table 1.

In order to find the MLE, observed information matrix, joint posterior modes of the parameters, and the marginal posterior distributions of the parameters and the parametric functions (and thus their posterior estimates, like the posterior mean, posterior variance, marginal posterior mode and the HPD credible set), one only needs to input the observations on U, I, the $\delta_{jl}^{(i)}$'s, the a_{jl}'s, and the prior hyper-parameters. The rest of the computation is automated.

Let $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \lambda_3, \beta)'$. First $\hat{\boldsymbol{\theta}}$, the MLE of $\boldsymbol{\theta}$, and the observed information matrix $\boldsymbol{I}(\hat{\boldsymbol{\theta}})$, are found to obtain the asymptotic normal approximation to $\pi(\boldsymbol{\theta}|\boldsymbol{D})$, as given below,

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \\ \hat{\lambda}_3 \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} 1.4892 \\ 3.6729 \\ 4.2312 \\ 2.3643 \end{pmatrix}$$

Table 1
Simulated data

Unit no	Y_1	Y_2	Y_3	C	T	δ	MRS
1	0.62516	0.72725	0.23981	0.63620	0.23981	0	{3}
2	0.83342	0.63805	0.28854	1.76707	0.28854	0	{2,3}
3	0.67538	0.38329	0.55443	0.49215	0.38329	0	{2}
4	0.39754	0.46925	0.15951	1.17536	0.15951	0	{3}
5	0.47185	0.61409	0.38786	0.35490	0.35490	1	
6	0.82108	0.57064	0.64139	0.66914	0.57064	0	{1,3}
7	0.77629	0.09656	0.17585	0.98902	0.09656	0	{1,2}
8	0.46431	0.31356	0.11789	0.61577	0.11789	0	{1,2}
9	0.60673	0.69988	0.39525	0.47402	0.39525	0	{3}
10	0.65536	0.43368	0.93710	0.29639	0.29639	1	
11	0.58622	0.32305	0.64160	0.63806	0.32305	0	{2}
12	0.31186	0.30005	0.81246	0.43811	0.30005	0	{2}
13	0.72570	0.78889	0.35105	0.23408	0.23408	1	
14	0.52170	0.42158	0.85580	0.27910	0.27910	1	
15	0.37055	0.51686	0.58171	0.82427	0.37055	0	{1}
16	0.63458	0.29595	0.18717	0.52330	0.18717	0	{1,3}
17	0.16483	0.63969	0.38590	0.55446	0.16483	0	{1}
18	0.56549	0.23557	0.79635	0.75109	0.23557	0	{2,3}
19	1.40931	0.95601	0.49288	0.35631	0.35631	1	
20	0.80152	0.93125	0.48870	0.64894	0.48870	0	{1,2,3}
21	1.02141	0.41525	0.42413	1.12451	0.41525	0	{2}
22	0.70456	0.56495	0.34267	0.60937	0.34267	0	{3}
23	0.54958	0.49645	0.06495	0.25302	0.06495	0	{3}
24	0.58731	0.68379	0.27730	1.04177	0.27730	0	{3}
25	0.57147	0.47265	0.20099	0.30349	0.20099	0	{1,3}
26	0.53278	0.42817	0.60841	1.04648	0.42817	0	{2,3}
27	0.42270	0.37907	0.75178	0.96011	0.37907	0	{2}
28	0.34043	0.52601	0.07556	0.60491	0.07556	0	{2,3}
29	0.57099	0.60735	0.72170	0.58987	0.57099	0	{1,2,3}
30	0.30955	0.42283	0.36376	0.40119	0.30955	0	{1,2}

and

$$I(\hat{\theta}) = \begin{bmatrix} 1.1287 & 0.1352 & 0.1143 & -2.0873 \\ 0.1352 & 0.5699 & 0.0867 & -2.0873 \\ 0.1143 & 0.0867 & 0.5135 & -2.0873 \\ -2.0873 & -2.0873 & -2.0873 & 24.9998 \end{bmatrix}$$

and thus $\pi(\theta|D)$ can be approximated by a

$$\mathcal{N}_4\left(\begin{pmatrix} 1.4892 \\ 3.6729 \\ 4.2312 \\ 2.3643 \end{pmatrix}, \begin{bmatrix} 1.0685 & 0.1561 & 0.2291 & 0.1214 \\ 0.1561 & 2.7036 & 0.7256 & 0.2994 \\ 0.2291 & 0.7256 & 3.1756 & 0.3449 \\ 0.1214 & 0.2994 & 0.3449 & 0.1039 \end{bmatrix}\right)$$

density, for any regular prior $\pi(\theta)$ on θ.

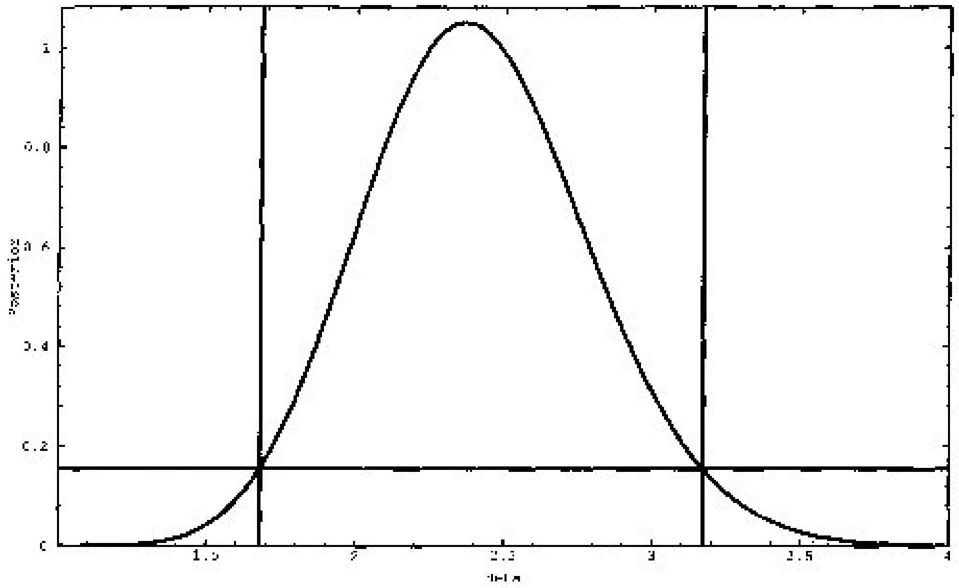

Fig. 1.

For the purpose of illustration, we take the scale-invariant noninformative prior π_I of Section 3 for $(\lambda_1, \lambda_2, \lambda_3)'$, and take a uniform prior for β on $[1, \infty)$. Note that for a given problem, one will usually have some prior idea about the failure rates of the Y_i's. If they are felt to be increasing (as is the case in this illustration), the support of the prior of β should be restricted to $[1, \infty)$, while for decreasing failure rates the prior support for β should be taken as $(0, 1]$. In either case, a uniform prior will serve the purpose of a reference prior. Thus the prior on $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta}) = \frac{1}{(\lambda_1 + \lambda_2 + \lambda_3)^3}, \quad \lambda_1 > 0, \quad \lambda_2 > 0, \quad \lambda_3 > 0, \quad \beta \geqslant 1.$$

Next, the joint posterior mode of $\boldsymbol{\theta}$, the maxima of $\pi(\boldsymbol{\theta}|\boldsymbol{D})$, is found again by using the EM algorithm, which is also utilized in obtaining $\hat{\boldsymbol{\theta}}$, the MLE of $\boldsymbol{\theta}$, above. The marginal posterior density of β, $\pi(\beta|\boldsymbol{D})$, is obtained numerically using **P1**, and is plotted in Fig. 1. Similarly, the marginal posterior densities, of λ_1, λ_2 and λ_3, $\pi(\lambda_1|\boldsymbol{D})$, $\pi(\lambda_2|\boldsymbol{D})$, and $\pi(\lambda_3|\boldsymbol{D})$, are numerically evaluated using **P4**, and are plotted in Figs. 2, 3 and 4, respectively. The posterior mean, posterior median, posterior variance, marginal posterior mode and the exact 95% HPD credible set are obtained numerically by using the marginal posterior densities. The asymptotic 95% HPD credible sets are found using the normal approximation to the posterior, described above. These quantities for the four parameters are summarized in Table 2.

The marginal posterior densities of π_1, π_2, and π_3 are obtained using **P3**, and are plotted in Figs. 5, 6, and 7, respectively. The marginal posterior density of λ is obtained

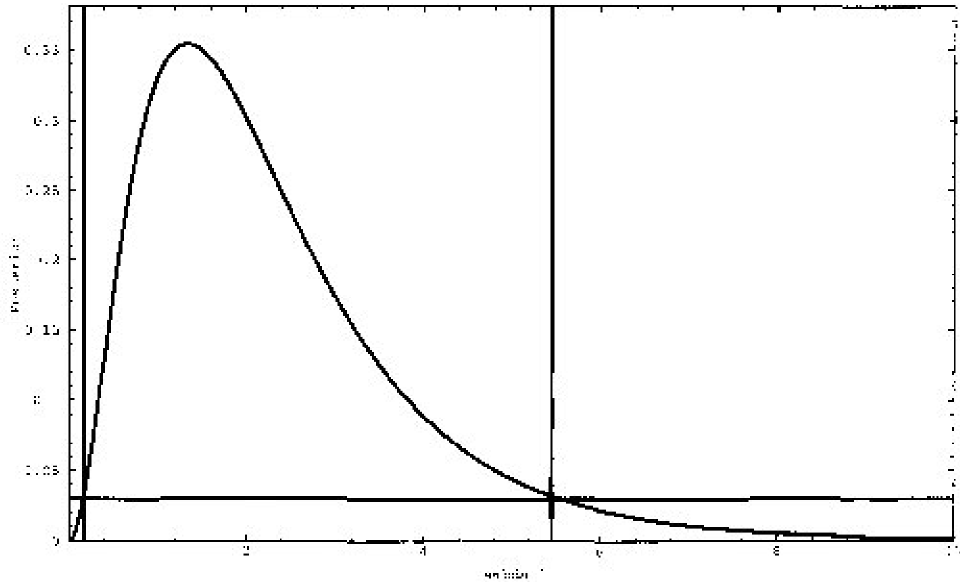

Fig. 2.

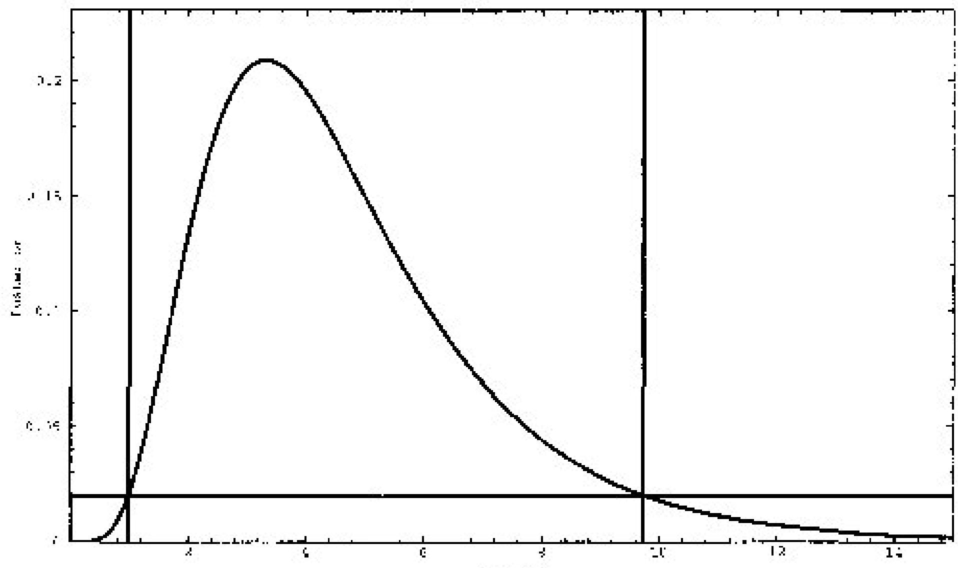

Fig. 3.

using **P2**, and is plotted in Fig. 8. The posterior estimates of these four parametric functions are obtained numerically and are summarized in Table 3.

Along with the marginal posterior densities of the parameters and the parametric functions, the above mentioned exact 95% HPD credible sets are also indicated in

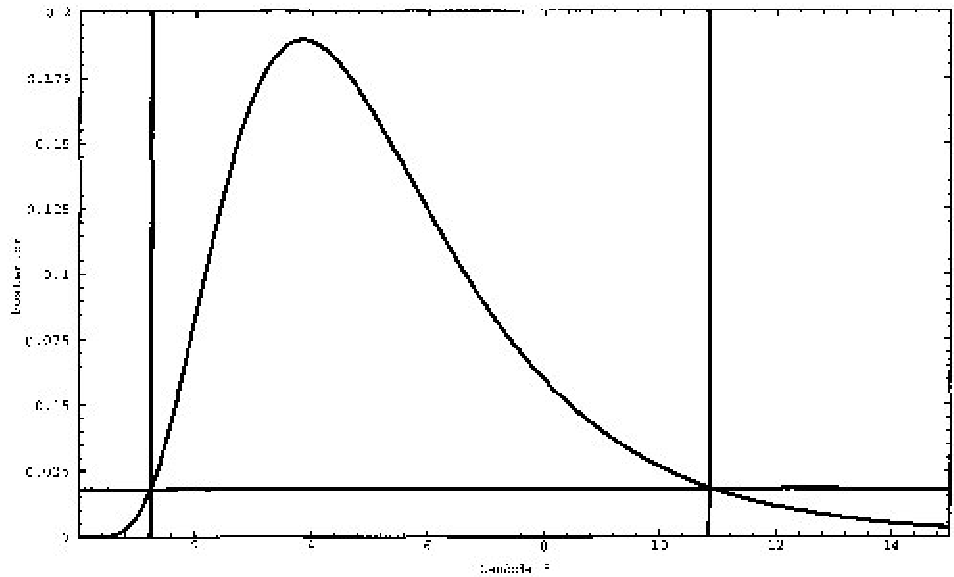

Fig. 4.

Table 2
Posterior estimates of parameters

Parameter	Joint modes	Posterior mean	Posterior median	Posterior mode	Marginal variance	Exact 95% HPD set	Asymptotic 95% HPD set
λ_1	0.9162	2.3535	1.9796	1.34	2.5495	[0.17,5.4563]	[0.0,3.5153]
λ_2	2.2595	4.7338	4.2090	3.31	6.2644	[0.99,9.7329]	[0.4501,6.8956]
λ_3	2.6029	5.3676	4.8088	3.83	7.4067	[1.24,10.848]	[0.7384,7.7239]
β	1.9913	2.4114	2.3958	2.36	0.1459	[1.68,3.1692]	[1.7324,2.9961]

Figs. 1–8, with the help of two vertical lines (indicating the two end-points), and a horizontal line (indicating the critical value of the posterior density).

In order to estimate the component reliability functions $\overline{F}_1(t)$, $\overline{F}_2(t)$, $\overline{F}_3(t)$, the marginal posterior densities of these functions are obtained at $t = 0.05, 0.10, \ldots, 0.85, 0.90$, using (4.6), while for estimating the system reliability function $\bar{F}(t)$, the marginal posterior densities are obtained at $t = 0.05, 0.10, \ldots, 0.55, 0.60$, using (4.7). The posterior mean, marginal posterior mode and the 95% HPD credible intervals are obtained at each of the above t's for each of the reliability functions. The point estimates (posterior mean and marginal posterior mode) and the two end-points of the HPD credible interval are plotted at each of the above t's and then joined by linear interpolation, to obtain the point estimates and the HPD credible band for the functions. Figs. 9–11 give these posterior estimates for $\overline{F}_1(t)$, $\overline{F}_2(t)$, and $\overline{F}_3(t)$, respectively, while Fig. 12 gives the estimates for $\bar{F}(t)$. In each of the Figs. 9–12, the 'dark' line indicates the

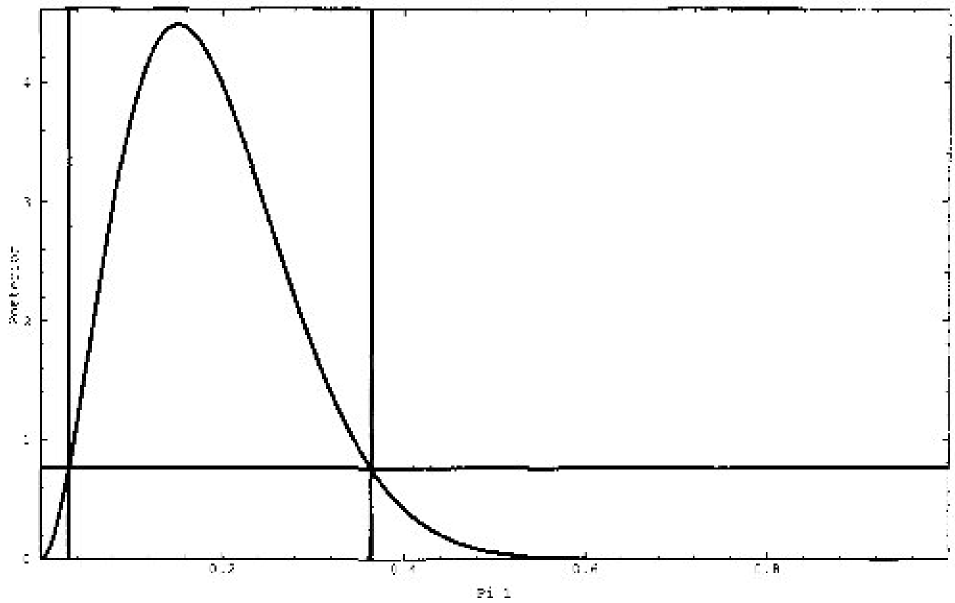

Fig. 5

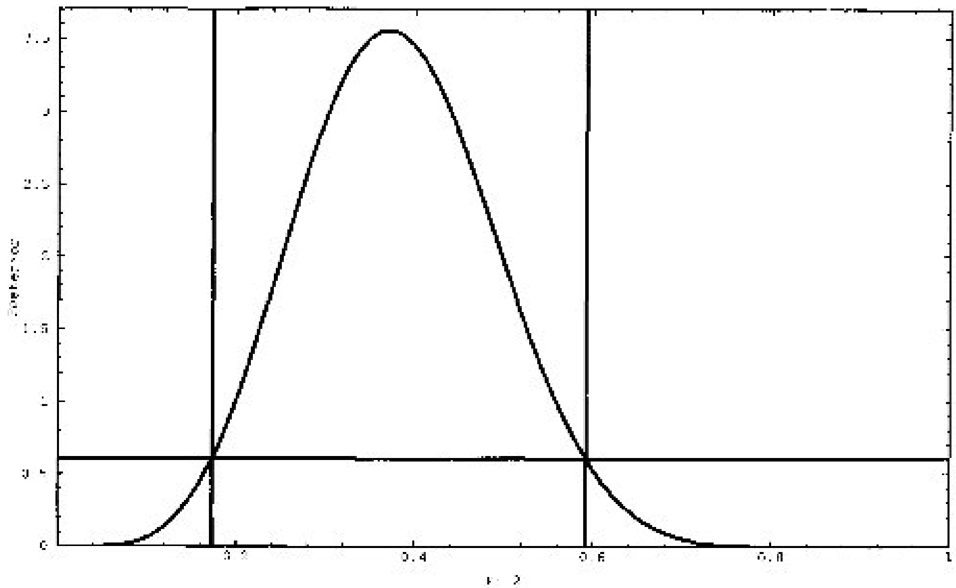

Fig. 6.

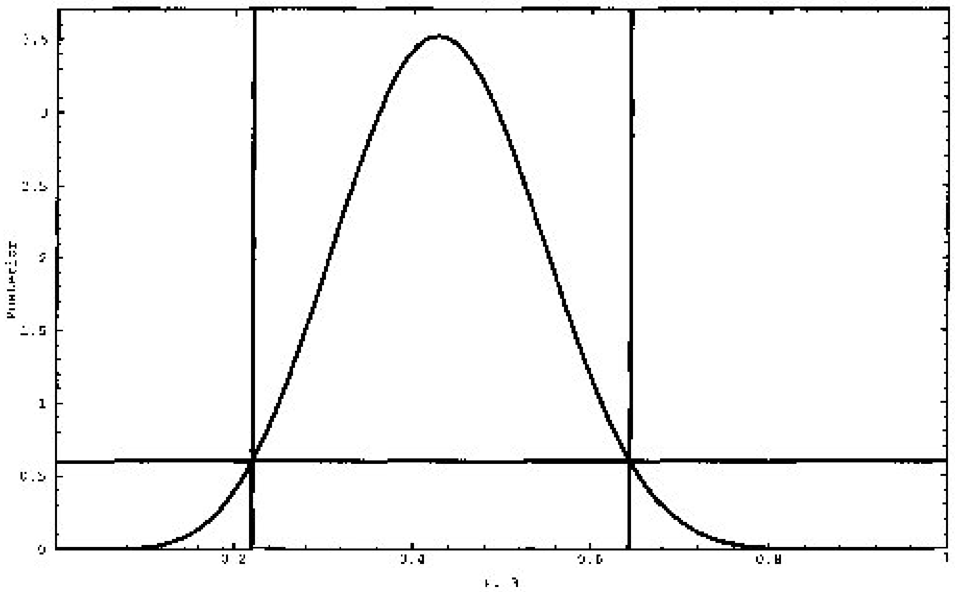

Fig. 7.

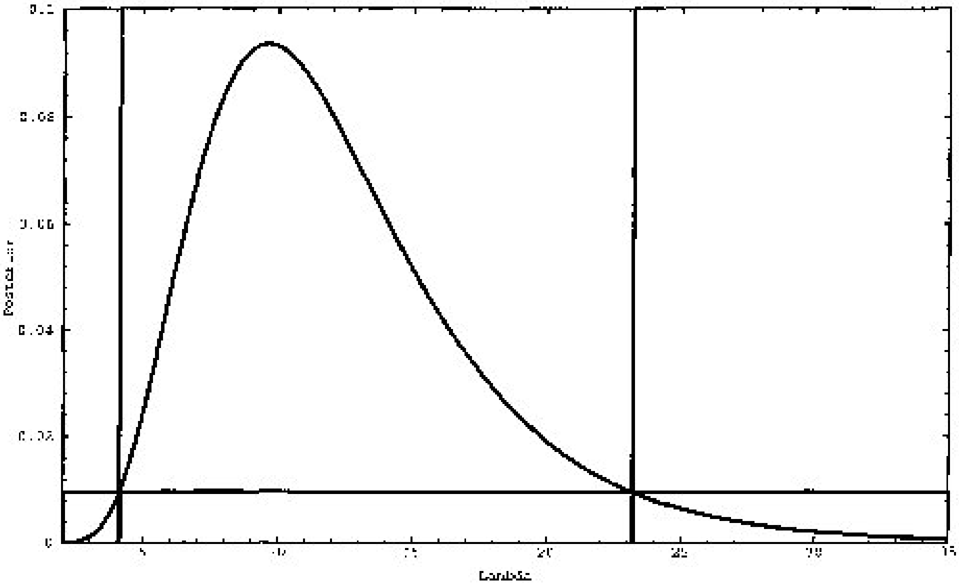

Fig. 8.

Table 3
Posterior estimates of parametric functions

Parametric function	Posterior mean	Posterior median	Marginal mode	Posterior variance	Exact 95% HPD set
π_1	0.1885	0.1777	0.152	0.0082	[0.031,0.3646]
π_2	0.3800	0.3767	0.370	0.0118	[0.173,0.592]
π_3	0.4315	0.4300	0.427	0.4315	[0.220,0.6443]
λ	12.4526	11.4191	9.585	28.5183	[4.125,23.2071]

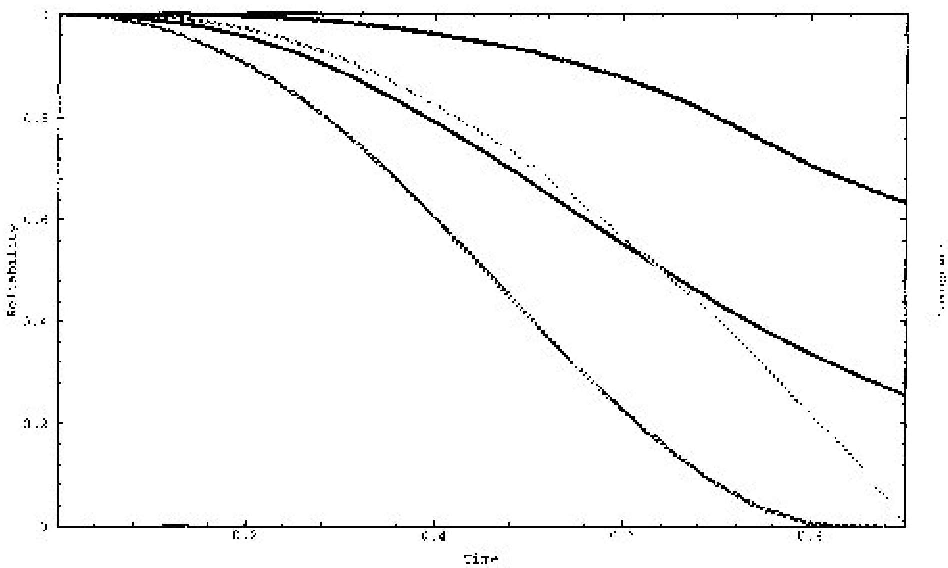

Fig. 9.

posterior mean, the 'light grey' line indicates the posterior mode, while the two 'dark grey' lines give the 95% HPD credible band for the reliabilities.

6. Concluding remarks

To conclude, we first observe a few computational details. The data are simulated using MINITAB. The remainder of the programs are written in C, and are run in a DEC-Alpha workstation. The graphics is done using Mathematica. In order to obtain both the MLE and the joint posterior modes, the initial value of β is taken to be 1. Convergence of (2.5) was obtained after 22 iterations and (4.5) converged after 24 iterations. For the EM iteration between (2.3) and (2.4) or (2.3) and (4.4), the initial values for λ_1, λ_2 and λ_3 are all taken to be 1, and both the EM iterations converged after 13 iterations. All the above iterations are continued until the consecutive estimates agree up to 4 decimal places after the decimal point. The program yielded these estimates along with the observed information matrix almost instantaneously. All the numerical

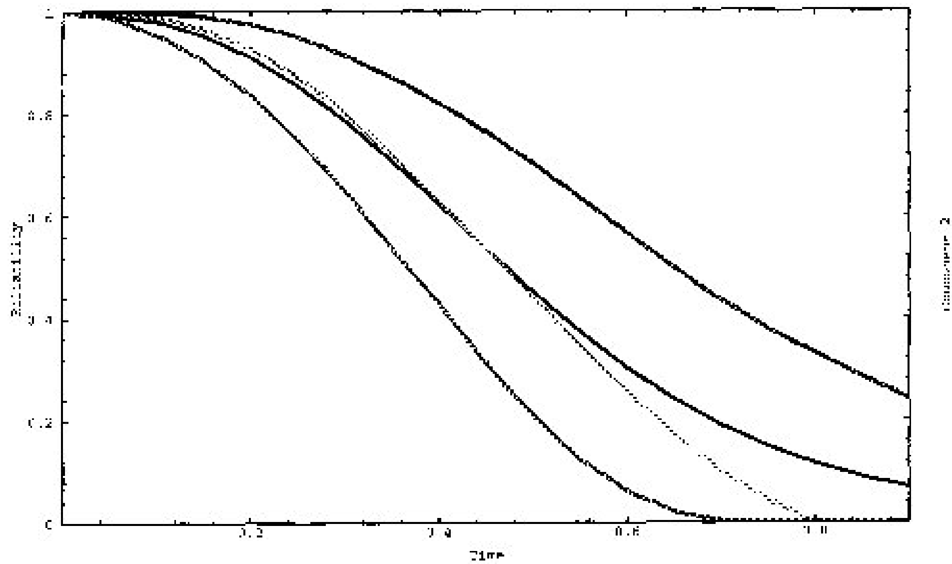

Fig. 10.

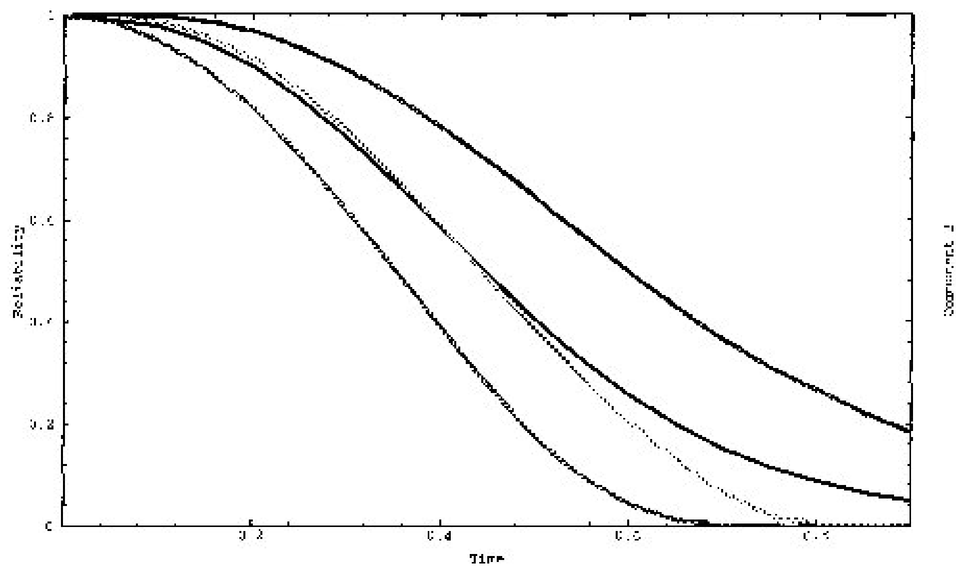

Fig. 11.

integrations are done using Simpson's one-third rule. Also all the improper integrals of the posterior densities are carried out till a finite value until the integral > 0.9999. The computation of $\pi(\beta|\boldsymbol{D})$ is done first, which is used in the computation of all the subsequent posterior densities (except of course the $\pi(\pi_i|\boldsymbol{D})$'s), and is obtained

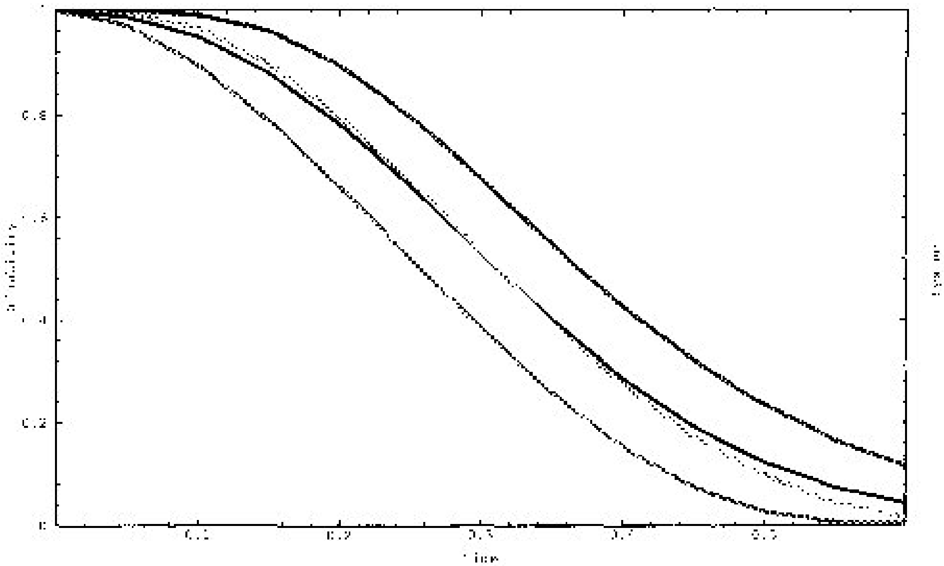

Fig. 12.

within 30 CPU minutes. The computation of $\pi(\lambda|\boldsymbol{D})$ and $\pi(\pi_i|\boldsymbol{D})$'s are equally fast. The evaluation of $\pi(\lambda_i|\boldsymbol{D})$'s is done parallely and takes more than 500 CPU minutes. The posterior distribution of $\overline{F}(t)$ for a fixed t, is obtained within 60 CPU minutes, but that of the $\overline{F}_i(t)$'s, which are again computed parallely, takes more than 1000 CPU minutes. A general purpose program is written to obtain the Bayesian estimates like the posterior mean, posterior quantiles, posterior variance, marginal posterior mode and any $100(1-\alpha)\%$ HPD credible set, once a univariate posterior density is given as input. This program is applied on the posterior densities of all the parameters and the parametric functions to obtain the Bayesian estimates, and yielded the results almost instantaneously in each case.

The purpose of the paper is to illustrate that a noninformative Bayesian analysis or a robust Bayesian analysis, (via the asymptotic normal approximation to the posteriors for any regular priors), of masked system failure data is computationally feasible in real time, when the component lifetimes are assumed to be independent Weibull's with proportional hazard. If the component lifetimes are independent exponentials, analysis may proceed exactly in a similar manner and the expressions of the posterior densities simplify further and the dimensionalities of the numerical integrations never exceed 1. While the numerical integrations of arbitrary dimensions may be avoided via Gibbs sampling, when the component lifetimes are Weibulls with unequal shape parameters; in this case of equal shape parameters, the solution is partially analytical, and thus reducing the dimension of numerical integrations to at most 2, even with a $(k+1)$-dimensional parameter space.

Acknowledgements

The authors would like to thank the referees for valuable comments. The research was supported in part by US Air Force Grant AFOSR F49620-95-1-0094. The first author would also like to thank the Department of Economics, Bilkent University, Ankara, Turkey, for their academic and computational support for the research.

References

Berger, J.O. (1985). *Statistical Decison Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

Berger, J.O. and D. Sun (1993). Bayesian analysis for the poly-Weibull distribution. *J. Amer. Statist. Assoc.* **88**, 1412–1418.

David, H.A. and M.L. Moeschberger (1978). *The Theory of Competing Risks*. Griffin's Statistical Monograph Series Vol. 39, Macmillan, New York.

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

Dinse, G.E. (1982). Nonparametric estimation for partially complete time and type of failure data. *Biometrics* **38**, 417–431.

Doganaksoy, N. (1991). Interval estimation from censored & masked system-failure data. *IEEE Trans. Reliab.* **R-40**, 280–285.

Ghosh, J.K., B.K. Sinha and S.N. Joshi (1982). Expansions for posterior probability and integrated bayes risk. In: S.S. Gupta and J.O. Berger, Eds., *Statistical Decision Theory and Related Topics III*. Academic Press, New York, 403–456.

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

Jeff Wu, C.F. (1983). On the convergence of the EM algorithm. *Ann. Statist.* **11**, 95–103.

Johnson, R.A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41**, 851–864.

Marshall, A.W. and I. Olkin (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **62**, 30–44.

Miyakawa, M. (1984). Analysis of incomplete data in competing risks model. *IEEE Trans. Reliab.* **R-33**, 293–296.

Peña, E.A. and A.K. Gupta (1990). Bayes estimation for Marshall–Olkin exponential distribution. *J. Roy. Statist. Soc. Ser. B* **52**, 379–389.

Reiser, B., I. Guttman, D.K.J. Lin, F.M. Guess and J.S. Usher (1995). Bayesian inference for masked system lifetime data. *Appl. Statist.* **44**, 79–90.

Usher, J.S. (1987). Estimating component reliabilities from incomplete accelerated life test data. Ph.D. dissertation, Department of Industrial Engineering, North-Carolina State University, Raleigh, NC.

Usher, J.S. and T.J. Hodgson (1988). Maximum likelihood analysis of component reliability using masked system life-test data. *IEEE Trans. Reliab.* **R37**, 550–555.