Bayesian Analysis of the Mixing Function in a Mixture of Two Exponential Distributions

Sujit K. GHOSH & Nader EBRAHIMI

North Carolina State University, Raleigh

Northern Illinois University, DeKalb

Institute of Statistics Mimeo Series No. 2531

May 2001

SUMMARY. In this paper we propose a semiparamteric Bayesian approach to estimate the mixing function in a mixture of two exponential distributions. Unlike, the traditional mixture of two distributions in this paper we assume that the mixing parameter changes with time. Such models arise naturally in many applications such as software reliability engineering and other related fields. Our proposed models are different from the usual mixture of two or more probability distributions in the sense that we consider mixing two or more hazard rates instead of probability densities. This makes the underlying parameters of the model more interpretable from an engineering application point of view as most engineers think in terms of hazard rates instead of probability densities. The Bayesian posterior distributions of the parameters are analytically intractable. We use Markov Chain Monte Carlo (MCMC) methods to fit our proposed models to observed data. Another advantage of the simulation based Bayesian methods for our models is that censoring can be handled routinely. We also use a decision-theoretic model choice criteria to compare our proposed semiparametric model with a corresponding fully parametric model. We illustrate our approach with an application involving insulating fluid failure data.

AMS (1991) *subject classification.* primary 62F35, secondary 62G05, 62H12

Key words and phrases. Bayesian inference, hazard rates, MCMC, mixture distributions.

1 Introduction

In experiments with life testing, it has been found that the lifetime may often be reasonably described by an exponential distribution. The exponential distribution has been used as a model, at least as a first approximation, in areas ranging from studies on the lifetimes of manufactured items, the time interval between failures of software systems to research involving survival or remission times in chronic diseases. See Balakrishnan and Basu (1995) and Meeker and Escobar (1998) for more applications of the Exponential distribution. However, in some applications we may need to extend the exponential distributional assumption to more general class of distribution that meets the specific need of the possibly complicated structure of the data. For instance, consider two populations with exponential distributions that have been mixed in unknown proportions. In this article we develop a model which is a mixture of only two hazards, where the mixing proportion varies over time. In our subsequent development we restrict our models to only constant hazards. Similar methods can be extended for mixtures with more than two hazards and possibly for any other lifetime distributions such as Weibull, Gamma etc. However when considering other probability distributions, one must be careful about the identifiability of the parameters when allowing for time-variant mixing function. In Section 2 we find that even in the case of mixing two exponential hazards with time-dependent mixing function,

our models are sufficiently flexible to model different shapes of hazard function. In fact, we will show that our approach models any bounded hazard function, where the bounds are itself unknown parameters.

Consider that two populations having exponential distributions with constant hazard rate λ_1 and λ_2 respectively, have been mixed in the unknown proportions $p(t)$ and $1-p(t)$ respectively, where $p(\cdot) \in [0,1]$ and varies continually with time t. The resulting hazard function of the process is given by

$$\lambda(t) = \lambda_1 p(t) + \lambda_2(1-p(t)) \tag{1}$$

Notice that, this hazard rate mixture model in (1) is different from the usual mixture models, which typically considers a mixture of two densities. The purpose of this formulation is that in some experiments engineers usually think in terms of failure rates as opposed to probability densities. We present couple of real life examples at the later part of this section. It is clear that even when $p(t) = p$, $\forall t$, the model as described in (1) does not reduces to the usual mixture of two exponential distributions. Moreover, for non-constant mixing function $p(t)$, our hazard rate mixture models can be significantly different from the usual density mixture models.

The hazard rate mixture model is related to the usual mixture model in the following way. Instead of mixing two constant hazard rates we now consider the

classical mixture of two distributions,

$$f(t) = pf_1(t) + (1-p)f_2(t),\ 0 \leq p \leq 1,$$

where $f_1(t)$ and $f_2(t)$ are two densities on $[0, \infty)$. For this case the corresponding hazard function is obtained by

$$\lambda_f(t) = p(t)\lambda_1(t) + (1-p(t))\lambda_2(t) \tag{2}$$

where $p(t) = pS_1(t)/S_p(t)$, $S_i(t) = \int_t^\infty f_i(s)ds$, $\lambda_i(t) = f_i(t)/S_i(t)$, for $i = 1, 2$ and $S_p(t) = pS_1(t) + (1-p)S_2(t)$. Notice that by choosing $\lambda_i(t) = \lambda_i$ for $i = 1, 2$, we obtain

$$p(t) = p\exp(-\lambda_1 t)/[p\exp(-\lambda_1 t) + (1-p)\exp(-\lambda_2 t)].$$

Thus, a particular choice of $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$ leads to a special form of the "dynamic mixing function"; $p(\cdot)$. In fact, from (1) it appears that the hazard rate mixture model produces a general class of bounded hazard rate models, where the bounds are unknown. Thus, even with exponential hazards our models are sufficiently flexible to entertain different shapes for hazard rates. We take a Bayesian viewpoint and develop techniques to obtain finite sample estimates of not only the parameter of interest but also the finite sample measures of uncertainty of our proposed estimates. Next we describe two real life situations where the proposed hazard rate mixture models arise quite naturally.

The model described in (1) arises in practice quite often. For example, components may be manufactured over a period of time, using two different machines.

Physical characteristics and thus reliability of components from the two different machines may be different, but it may be impossible otherwise to distinguish between components made with the different machines. Given that the proportions of units in a population manufactured by the first machine at time t is $p(t)$, then the population hazard function can be described by the model (1) The above model is also applicable in software engineering, for example, in software reliability engineering one can handle an operational profile, the environment or the operational profile is established by combining the input states, that varies continually in time simply by expressing the operational profile as the weighted sum of two component operational profiles. If $p(t)$ is the weight for the first component (the other weight being $1-p(t)$) and if the failure rates for the two different component operational profiles are λ_1 and λ_2 respectively, then the overall failure rate is given by the equation (1). For more details about operational profile see Musa (1999). In these cases, often it is of interest to obtain estimates of the constant hazard rates by treating the mixing function as a nuisance (infinite-dimensional) parameter.

Our goal in this article is to make inference for λ_1, λ_2 and $p(\cdot)$. To our knowledge a fully semi-parametric approach (with $p(\cdot)$ being unspecified) to this problem has not been addressed in the literature. Classical methods based on likelihood theory is almost infeasible. However given a parametric specification of the mixing function $p(\cdot)$ likelihood based approach (such mle's etc.) can be ob-

tained. However even in such problems a measure of uncertainty of the parameter estimates (e.g. mle's) are typically obtained via asymptotic distribution of the estimates. However for our application such asymptotic methods may not be suitable due to scarcity of data and resources. In some applications, due to destructive nature of the experiment large samples can not be obtained. We adopt a fully Bayesian semiparametric approach for this problem, avoiding asymptotics to provide inference conditionally only upon the data actually observed. Bayesian inference is in many ways a tempting approach to problems in reliability and survival analysis. In particular, the direct probabilistic interpretation of the posterior distribution is possible as well as problems formulated in terms of integrals with respect to the posterior.

The outline of the paper is as follows. In Section 2, we formulate the Bayesian model for a sample of lifetimes collected under (1). In Section 3, we discuss prior specifications for the parameters λ_1, λ_2 and $p(\cdot)$ and fitting using simulation-based methods. In this section we also illustrate our method by an example based on simulated data. We apply our technique to a real data set in section 4.

2 The hazard rate mixture model

Let $T_1, \ldots, T_n$ be a random sample from the probability distribution with hazard rate as given in (1). We wish to estimate the hazard rate $\lambda(t)$. To keep the

method and notations simpler, we assume that there is no censoring and we observe all the failure times. In the presence of censoring the likelihood becomes product of some complicated terms depending on the type of censoring and might involve integrals that can not be evaluated analytically. However from a Bayesian point of view, such difficulties can be avoided by using sampling based methods, which treats the censored data as missing observation. We discuss these issues in Section 3. The log-likelihood function takes the form

$$l(\theta) = \sum_{i=1}^{n} \left\{ \log[\lambda_2 - \delta p(t_i)] + \delta P(t_i) - \lambda_2 t_i \right\}, \tag{3}$$

where $\theta = (\lambda_1, \lambda_2, p(\cdot))$ denotes the parameters of the model and $\delta = \lambda_2 - \lambda_1$ and $P(t) = \int_0^t p(u)du$.

With $p(\cdot)$ being unspecified, the likelihood analysis based on the MLE is extremely complicated. To our knowledge, no one has attempted to obtain MLE of θ for the likelihood given in (3). A standard likelihood based methodology approximates the unspecified $p(\cdot)$ by using a step function approximation. However the constraints that $0 \leq p(t) \leq 1$ and the dimensionality of the problem may cause problem in maximizing the likelihood. Splines or wavelets based approach may be used as an alternative. However, we have not explored such possibilities in this article. From (3), it also appears that $p(\cdot)$ is not identifiable without further restrictions. This is because when $\delta = 0$, the likelihood doesn't depend on $p(\cdot)$. So from now on, without loss of generality we assume that $\delta > 0$.

We take a Bayesian perspective and obtain the posterior distribution of θ conditional on the data actually observed. In addition to being complicated the classical method does not provide finite sample measures of uncertainty (e.g. standard error) of the estimates. In most of the engineering data where data are usually scarce and are expensive to obtain, such large-sample estimates are often unreliable. The Bayesian methods are often the only alternative to make statistical inference in real time and they are generally computationally expensive as they depend on large simulation based methods. However with new generation fast computers such problems are becoming increasingly minor ones. So we take a Bayesian approach to obtain estimates of the hazard rates. This approach allows us to obtain smooth posterior estimate of $p(\cdot)$ with a continuous credible (confidence) band around it. For Bayesian inference, we complete the probability specification by considering a prior distribution on θ.

2.1 A Bayesian Approach

In order to obtain a full probability model we specify a prior process for the infinite-dimensional parameter θ. The prior distribution along with the sampling distribution as in (3) provides the joint probability model for the data and the parameters. We then use Bayes Theorem to obtain the posterior distribution of the parameters conditional on the observed data.

We start with the problem on how to specify the prior for the infinite-dimensional

parameter $p(\cdot)$. We propose an independent increment process as prior for $p(\cdot)$. Let $0 = a_0 < a_1 < \ldots < a_m < a_{m+1} = \infty$ be a partition of $[0, \infty)$. Assume that $p(t) = p_k$ for $t \in (a_k, a_{k+1}]$ for $k = 0, \ldots, m$. It should be noted that if $\sup_k(a_{k+1} - a_k) \to 0$, this step function can approximate any smooth function. Notice that instead of the step function we could also consider some smooth function (say linear or quadratic) within the interval $[a_k, a_{k+1})$. As we don't have much information about the p_k's, we assume that $\mu_k = logit(p_k)$ $i.i.d. N(\mu, \sigma^2)$. In other words, we assume that $p_1, \ldots, p_m$ are *exchangeable*. Alternatively, we may also consider a correlated prior process for $p(\cdot)$. For instance, we may consider an AR(1) process for the sequence $\mu_1, \mu_2, \ldots$. Notice that using suitable transformation on $p(\cdot)$, we may consider a Dirichlet process prior or Gamma process priors. However, in this article we limit our applications to only logit Gaussian independent increment processes. We can control the prior information by making σ large or small. For instance, if not much prior information is available we would make σ large to express our high uncertainty about $p(\cdot)$. However, if we have enough prior information, we would choose σ accordingly. For our data set, we use a flat hyper-prior for μ and an Inverse Gamma for σ^2 with a large variance, to make the prior information rather vague.

Next we consider a semi-conjugate prior for (λ_1, λ_2). In particular, we assume that the joint distribution of (λ_1, λ_2) is given by,

$$\pi(\lambda_1, \lambda_2) \propto \lambda_1^{b_1 - 1} \lambda_2^{b_2 - 1} \exp\{-c_1\lambda_1 - c_2\lambda_2\} I(\lambda_1 < \lambda_2) \tag{4}$$

where the requirement that $\lambda_1 < \lambda_2$ is needed for identifiability of the parameters. In applications, we choose c_i, b_i in such a way that the prior information is rather vague. For example, choosing $c_i = b_i = 10^{-3}$ produces almost flat priors. However one must be careful in specifying the hyper-parameters as some specifications (even with proper priors) may give rise to instability of the Gibbs sampler that we plan to use to fit the model. We discuss the model fitting in the next section 3. As always, the choice of the prior should reflect the honest understanding which the analyst has about the problem at hand. If the choice of hyperparameters do not offer enough room for this, then one should modify the prior specification.

Combining the sampling distribution as in (3), with the above prior distribution for $\mu = (\mu_1, \ldots, \mu_k)$ and (4) provides the complete Bayesian model and thus the posterior distribution of θ given the data $t = (t_1, \ldots, t_n)$, is proportional to

$$L(\lambda_1, \lambda_2, \mu; t)\pi(\lambda_1, \lambda_2)\pi(\mu). \tag{5}$$

The marginal posterior distributions $\pi(\lambda_1|t)$ and $\pi(\lambda_2|t)$ enable us to learn about the hazards of sub-populations. The posterior $\pi(\mu|t)$ tells us about the mixing function.

The expression in (5) is not analytically tractable so we turn to simulation based approaches for fitting such a model. That is, we seek to draw samples from the the posterior distribution in (5) and use a Monte-Carlo method in order to learn about its features. Notice that, if some the observations are censored, the Bayesian inference proceeds with a data augmentation approach. For instance, if

T_i is right censored, i.e., we observe $Y_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$, where C_i denotes the censored value, then we sample from the conditional distribution of T_i given (Y_i, Δ_i, θ) and augment it to the observed data to form a complete data likelihood.

3 Model Fitting using MCMC

Customary iterative approaches, using, e.g., rejection method or weighted bootstrap, as presented in Smith and Gelfand (1992) are difficult to implement for our models. However, the form in (5) is nicely suited for Markov Chain Monte Carlo (MCMC) simulation using Gibbs sampler of Gelfand and Smith (1990). Using the independent prior process formulation we express the sampling distribution as usual mixture model. In order to see this, first we write the hazard function as

$$\lambda(t) = \sum_{k=0}^{m} \alpha_k I_{A_k}(t).$$

where $\alpha_k = \lambda_2 - \delta p_k$ and $I_{A_k}(t)$ is the indicator function of the set $A_k = [a_k, a_{k+1})$. So it follows that the cumulative hazard is given by

$$\Lambda(t) = \sum_{k=0}^{m} (\alpha_k t + \beta_k) I_{A_k}(t),$$

where $\beta_k = \delta\left[p_k a_k - \sum_{j=0}^{k-1} p_j(a_{j+1} - a_j)\right]$. It now follows that the density of T can now be expressed as

$$f(t) = \sum_{k=0}^{m} \omega_k f_k(t), \tag{6}$$

where $\omega_k = e^{-\beta_k}\left[e^{-\alpha_k a_k} - e^{-\alpha_k a_{k+1}}\right]$ and $f_k(t) = \frac{\alpha_k e^{-\alpha_k t} I_{A_k}(t)}{e^{-\alpha_k a_k} - e^{\alpha_k a_{k+1}}}$. It is easily verified that $\sum_{k=0}^{m} \omega_k = 1$. Thus it appears that the failure time T has a truncated exponential mixture distribution. However this mixture distribution is quite different from the usual mixture distribution in the sense that the mixing weights ω_k's and the parameters of the component distributions (in this case, these are truncated exponentials) are functionally related. We make use of the representation (6) to implement the Markov Chain Monte Carlo (MCMC) method for our model. We use the data augmentation approach to reduce the mixture model in (6) to sampling from the components of the mixture model. Then we use a random walk Metropolis method to sample from the full conditionals of $\theta = (\lambda_1, \lambda_2, p_0, \ldots, p_m)$. More precisely, following Bayesian hierarchical model is equivalent to fitting a density of the form as given in (6).

$$\begin{aligned} T_i|K_i, \theta &\sim Exp(\alpha_{K_i})I(a_{K_i}, a_{K_i+1}), \quad i = 1, \ldots, n, \\ K_i|\theta &\sim categorical(\omega_0, \ldots, \omega_m) \\ \text{i.e. } \Pr[K_i = k|\theta] &= \omega_k, \quad k = 0, \ldots, m. \end{aligned}$$

where $Exp(\alpha)I(a, b)$ denotes an Exponential distribution with mean $1/\alpha$ truncated at a and b. We use the software `WinBUGS` (see Speigelhalter et al., 1996)

Table 1: Breakdown inter-failure times of insulating fluid under two levels of voltage stress.

0.19	0.59	0.18	0.35	1.47	0.38	0.99	0.52	0.18	1.65	0.85	0.66
0.26	4.33	19.15	0.77	1.39	2.8	36.18	0.35	0.24	0.37	0.03	0.7
0.28	0.1	0.52	0.12	0.19	0.77	0.1	1.58	8.42	11.73		

to implement the MCMC method and then to obtain samples from the posterior distribution.

4 Application to a real data set

We illustrate our methods by computing the the posterior distribution of λ_1, λ_2 and $p(\cdot)$. The data set presented in Table 1, are the times (in minutes) to break down of an insulating fluid under two levels of voltage stress.

In figure 1, we plot the histogram and boxplot of the inter-failure times on original and log scales respectively. It appears that the inter-failure times follow a log-normal type distribution. However in our data analysis we would try to model this data set using mixture of exponentials.

This is because, we know that the data set is coming from two populations

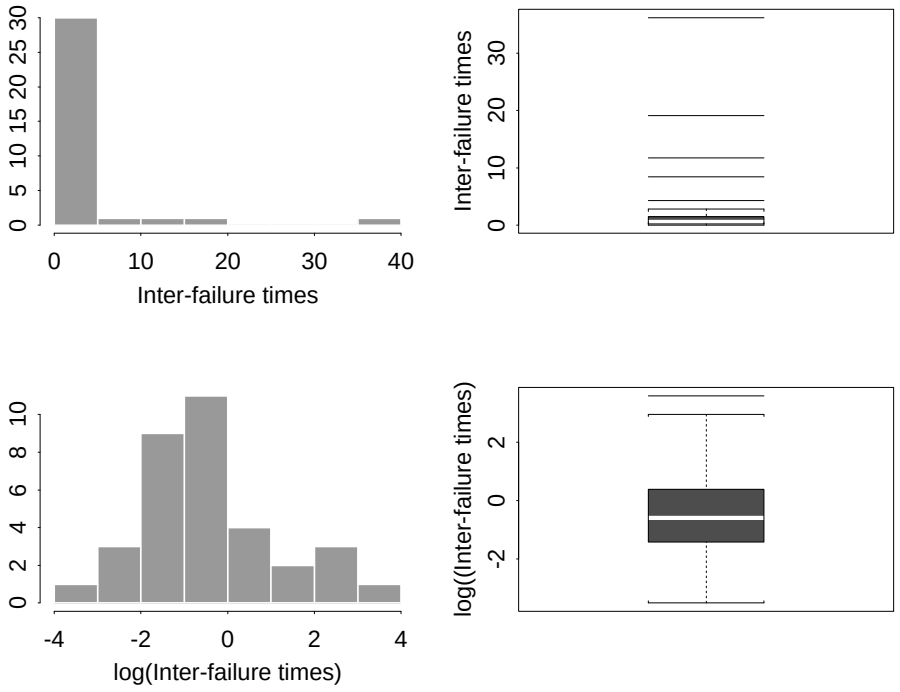

Figure 1: Histogram and boxplot of the data presented in Table 1.

(with different voltage stress), but we are not sure which observations are coming from what populations. As the voltage stress might change randomly over time, we assume that the mixing weight is a function of time. First, we fit an usual mixture of two exponential distributions. We call this a *mixture model-1.* In other words, we assume that failure times are coming from a density of the form

$$p\lambda_1 e^{-\lambda_1 t} + (1-p)\lambda_2 e^{-\lambda_2 t}.$$

We use locally flat priors, i.e. we assume that

$$\begin{aligned} \lambda_1 &\sim Gamma(0.001, 0.001)I(, \lambda_2), \\ \delta &\sim Gamma(0.001, 0.001)I(\lambda_1,) \\ \text{and } p &\sim Beta(0.5, 0.5) \end{aligned}$$

where $I(,)$ denotes the fact that we used a truncated Gamma distributions to ensure that $\delta = \lambda_2 - \lambda_1 > 0$. We used a `WinBUGS` code to obtain 5000 samples after a rejection of initial 5000 samples with three dispersed starting values. Thus for our final analysis we obtained 15000 samples from the joint posterior distribution of $(\lambda_1, \lambda_2, p)$. We used the trace plots and the autocorrelation functions plot to detect visual convergence of the chains. In addition we used the Gelman-Rubin diagnostic checks as a numerical support for the convergence of the chains. We ran the chains several times with different starting values and obtained very similar estimates, confirming no unusual behavior of the chains. The posterior summary values of the parameters $(\lambda_1, \lambda_2, p)$ are given in Table 2.

Table 2: Posterior estimates for the mixture model-1.

	mean	sd	2.5%	50%	97.5%
λ_1	0.0985	0.0401	0.0356	0.0932	0.1894
λ_2	1.8130	0.4413	1.0940	1.7610	2.8320
p	0.7840	0.0887	0.5829	0.7954	0.9244
$D(\theta)$	98.52	5.248	91.22	97.81	110.8
SPE	95.47	27.19	52.85	91.89	159.8

In Table 2 we present the posterior mean, sd, 2.5%, 50% and 97.5% percentiles of the three parameters of interest. We also report the posterior summary of the $D(\theta) = -2l(\theta)$, where $l(\theta)$ is as given in equation (3). In addition, we computed the model choice criteria proposed in Gelfand and Ghosh (1998), to see the predictive performance of our models. The Gelfand-Ghosh Criteria (GGC) is computed as:

$$GGC = \sum_{i=1}^{n} E[(\log(T_i^{rep} - \log(T_i^{obs})^2 | T^{obs}],$$

where T^{rep} represents a predicted value of T, under a given model, in our case the mixture model-1. The expectation is with respect to posterior predictive distribution of T^{rep} given the data $T^{obs} = (T_1^{obs}, \ldots, T_n^{obs})$. More details, on the motivation of the use of GGC and its connection to other information based criteria such as AIC and BIC can be found in Gelfand and Ghosh (1998). In

general, according to this criteria a model with lower GGC is preferred. In Table 2, we report the posterior predictive summary of the *sum of the predicted errors (SPE)*,

$$SPE = \sum_{i=1}^{n} (\log(T_i^{rep}) - \log(T_i^{obs}))^2.$$

Notice that $GGC = E[SPE|T^{obs}]$, which is the predictive mean (2nd column in Table 2) of the SPE.

From Table 2, it appears that the data is indeed coming from a mixture distribution as the 95% central posterior interval of p is not concentrated near to 0 or 1. Notice that the above models suggests a specific form of the mixing function for the hazard rate mixture model, viz., $p(t) = pe^{-\lambda_1 t}/[pe^{-\lambda_1 t} + (1-p)e^{-\lambda_2 t}]$. To obtain the posterior distribution of the mixing function, we use the samples of $(\lambda_1, \lambda_2, p)$ to obtain posterior mean and pointwise 95% posterior interval of the above function for each given value of t. We plot this functions at selected values of t in Figure 2. We notice that at initial period this function takes relatively lower value but approaches 1 as time increases. This function can serve as the baseline function of the semiparametric model that we fit next.

In order to fit the semi-parametric model we use $m = 5, 10, 15$ and 20 partitions of the interval $[0, \infty)$. We used the $D(\theta)$ and GGC as described above to choose $m = 15$ for our final analysis. We present the posterior summary of $D(\theta)$ and predictive summary of SPE in Table 3. In practice, one should try different partitions to see when the estimates of λ_1, λ_2 stabilizes. For our model, we

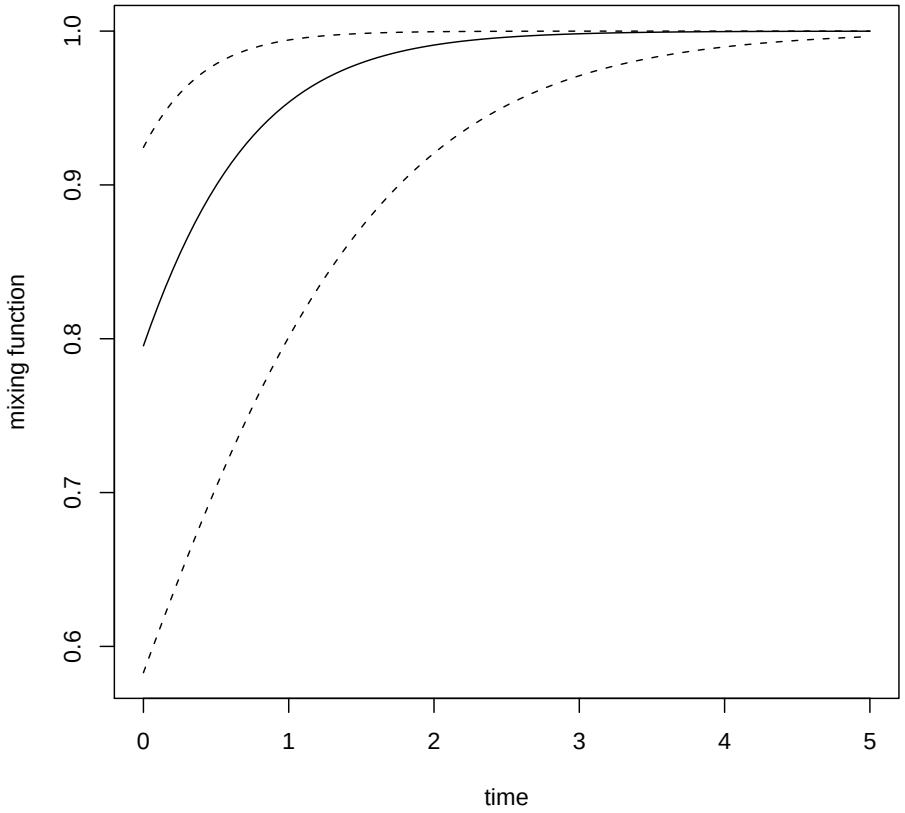

Figure 2: Posterior mean and 95% interval of the mixing function under model-1.

didn't find significant changes in the estimates for m beyond 20. We present the posterior estimates of λ_1, λ_2 for $m = 15$, in Table 3. Since for $p(\cdot)$ we have about 17 parameters, we present instead a plot of the smoothed version of the posterior mean of $p(\cdot)$ in Figure 3. Although posterior mean of the mixing function seems to be constant, but the lower and upper pointwise 95% posterior interval suggests that the mixing function is possibly time dependent at small inter-failure times but is essentially comes from of the first component as the inter-failure time increases. Also comparing the GGC of the model 1 and 2, we find that the gain in fitting the semiparameteric model to this data over the parametric model is not substantial. However the semiparametric model serves as an alternative to a class of parametric models and the GGC can be used to compare such non-nested models. In summary we find a strong evidence that the sampling distribution of the data is mixture of exponentials and the mixing function can be characterized by either by a parametric function as given in model 1 or by the step function as obtained in model 2.

Table 3: Posterior summary estimates for mixture model-2.

		mean	sd	2.5%	50%	97.5%
$m = 5$	$D(\theta)$	98.31	5.184	91.15	97.61	110.7
	SPE	95.17	27.25	52.84	91.3	159.6
$m = 10$	$D(\theta)$	98.28	5.064	91.26	97.68	110.3
	SPE	95.47	27.19	52.85	91.89	159.8
$m = 15$	$D(\theta)$	98.06	4.954	91.2	97.48	110.2
	SPE	95.07	27.74	52.47	91.44	160.3
	λ_1	0.0759	0.0347	0.0230	0.0708	0.1582
	λ_2	1.643	0.385	1.002	1.601	2.499
$m = 20$	$D(\theta)$	98.4	5.226	91.21	97.72	110.9
	SPE	95.77	27.74	52.95	92.16	159.5

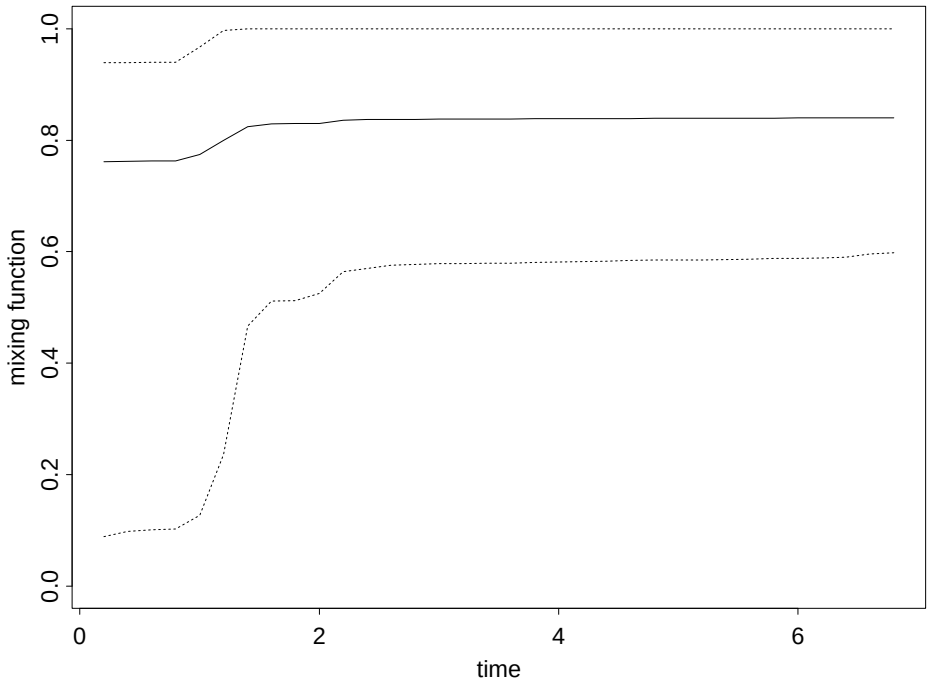

Figure 3: Posterior mean and 95% interval of the mixing function under model-2.

5 Discussion

In this article we have developed a Bayesian semiparamteric approach to estimating the mixing function of two constant hazard rate functions. It would be interesting to develop similar estimation method based on likelihood theory and derive the associated asymptotic distribution of the parameter estimates. To our knowledge such estimation methods have not been attempted for this problem. Another extension would be to see the sensitivity of the posterior distribution to the prior process of the mixing function. For instance we can assume that $p(t) = \sum_{k=1}^{m} p_k I_{A_k}(t)$, where both the bounds of the partitioning intervals A_k and the constant value of the function p_k in that partition are assumed unknown. For such prior processes, choosing the number of partitions m would be tricky! However we suspect the GGC may be still be used to select. In summary we think that the proposed class of models in (1) is very interesting for many engineering applications, in particular in the field of the software reliability. Further studies are required to understand the full flexibility of the proposed class of models.

Acknowledgments

We thank Speigelhalter, Thomas and Best of MRC Biostatistics unit, Institute of Public Health, Cambridge, UK for providing the program BUGS (Bayesian Inference Using Gibbs Sampling) and its PC version WinBUGS, available free of cost from their web site

`http://www.mrc-bsu.cam.ac.uk/bugs`

References

Balakrishnan, N. and Basu, A.P. (1995). *The Exponential Distribution*. Gordon and Breach Publishers, The Netherlands.

Gelfand, A.E. and Ghosh, S.K. (1998). Model Choice: A minimum posterior loss approach. *Biometrika*, **85**, 1–11.

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

Jewell, N.P. (1982). Mixtures of exponential distributions. *The Ann. Statist.*, **10**, 479–484.

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

Mendenhall, W. and Hader, R.J. (1958). Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. *Biometrika*, **45**, 504–520.

Musa, J.D. (1999). *Software Reliability engineering; More Reliable Software, Faster Development and Testing*, McGraw-Hill, New York.

Rider, P.R. (1961). The method of moments applied to a mixture of two exponential distributions. *Ann. Math. Statist.*, **32**, 143–147.

Smith, A.F.M. and Gelfand, A.E. (1992). Bayesian statistics without tears: A sampling-resampling perspective. *Amer. Statist.*, **46**, 84–88.

Speigelhalter, D., Thomas, A. and Best, N. (1999). *Winbugs*, Version 1.2 User Manual.

Wasserman, L. (2000). Asymptotic inference for mixture models using data-dependent priors. *JRSS-B*, 159–180.

Whittaker, J. (1973). The Bhattacharyya matrix for the mixture of two distributions. *Biometrika*, **60**, 201–203.