UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 03-32
Statistics and Econometrics Series 08
July 2003

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

USING WEIBULL MIXTURE DISTRIBUTIONS TO MODEL HETEROGENEOUS SURVIVAL DATA

J.M. Marín, M.R. Rodríguez Bernal and M.P. Wiper*

Abstract

In this article we use Bayesian methods to fit a Weibull mixture model with an unknown number of components to possibly right censored survival data. This is done using the recently developed, birth-death MCMC algorithm. We also show how to estimate the survivor function and the expected hazard rate from the MCMA output.

Keywords: Bayesian, Weibull, Survival Analysis, Mixtures, MCMC.

*Marín, Departamento de Informática, Estadística y Telemática, Universidad Rey Juan Carlos de Madrid; Rodríguez, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid; Wiper, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), e-mail: mwiper@est-econ.uc3m.es. Michael Wiper acknowledges assistance from the Spanish Ministry of Science and Technology via project BEC2000-0167.

Using Weibull mixture distributions to model heterogeneous survival data

J.M. Marín[1], M.T. Rodríguez-Bernal[2] and M. Wiper[2]
[1]Departamento de Informática, Estadística y Telemática
Universidad Rey Juan Carlos, Madrid, España
[2]Departamento de Estadística
Universidad Carlos III, Madrid, España

Abstract

In this article we use Bayesian methods to fit a Weibull mixture model with an unknown number of components to possibly right censored survival data. This is done using the recently developed, birth-death MCMC algorithm. We also show how to estimate the survivor function and the expected hazard rate from the MCMC output..
Key words: Bayesian, Weibull, Survival Analysis, Mixtures, MCMC.
AMS Classification: 62N05, 62H12.

1 Introduction

Historically, survival analysis has usually been carried out using non-parametric methods or via the classical statistical analysis of parametric survival models. However, during recent years, mainly due to the appearance of new computational algorithms, Bayesian methods have been increasingly applied in this field. See, for example, Ibrahim et al (2001).

One of the most popular parametric models used in the survival and reliability context is the Weibull distribution, see e.g. Dodson (1994).. However, where observations are taken from a possibly heterogeneous population, for example when patients have been treated using different methods, the simple Weibull model may not always be appropriate and mixture models may be considered. For example, in a Bayesian context, Chen et al (1985) used a two component mixture model for the analysis of cancer survival data generalizing an earlier idea of Berkson and Gage (1952). Quiang (1994) considered a similar model of a mixture of a Weibull component and a surviving fraction in the context of a lung cancer trial. It may be however that we still cannot capture fully the form of the survival distribution with mixtures of just two components. A more general idea is to

use semi-parametric models based on mixtures of a possibly large number of components. In this paper, we consider the modelling of lifetimes via Bayesian analysis of a mixture of Weibulls survival model with a possibly unknown number of components. This model could be appropriate for the analysis of clinical trial data, where several sub-populations may show different behaviour and the observed data consist of both complete and right censored lifetimes. Such a model can also be considered as an intermediate approach between simple parametric modeling and non-parametric methods. For an example of the latter in a Bayesian context, see Kottas (2002).

The paper is organized as follows. In Section 2, we define the mixture of Weibulls model that will be considered. In Section 3, we consider how to undertake Bayesian inference for this model assuming that the number of mixture components, k, is known, using a Gibbs sampling algorithm as in Diebolt and Robert (1994). In Section 4, we extend the situation to the case when k is unknown, by applying a birth-death algorithm developed by Stephens (2000). In Section 5, we illustrate the model using both simulated and real data sets and finally, in Section 6 we summarize our results and consider some possible extensions.

2 The Weibull Mixture Model and Data Observation

In this section, we define a Bayesian mixture of Weibulls survival model for analyzing survival data from clinical trials. Survival time is the time until an event occurs. For example, survival time may be the lifetime of a patient or time until recurrence of some disease of the patient.

We will suppose firstly that we shall observe the lifetimes S of a number of patients from a possibly heterogeneous population. Thus, we shall consider modeling the population distribution as a mixture of Weibull distributions. The basic Weibull density function is given by:

$$W(s|\theta, a) = \theta a s^{a-1} \exp\left\{-\theta s^a\right\},$$

where θ is a scale parameter and a is the shape parameter. A mixture of k Weibull densities is defined by:

$$f(s|k, \mathbf{w}, \theta, \mathbf{a}) = \sum_{j=1}^{k} w_j W(s|\theta_j, a_j),$$

where $\theta = (\theta_1, \ldots, \theta_k)$, $\mathbf{a} = (a_1, \ldots, a_k)$, are the parameters of each Weibull distribution and $\mathbf{w} = (w_1, \ldots, w_k)$ is a vector of non-negative weights which sum to 1. Note that for

large values of k, essentially any density on the positive real line can be approximated by such a mixture form.

Given this model, the population survival function $S(s|k, \mathbf{w}, \theta, \mathbf{a})$ and hazard function $h(s|k, \mathbf{w}, \theta, \mathbf{a})$ are straightforward to calculate. We have

$$S(s|k, \mathbf{w}, \theta, \mathbf{a}) = \sum_{j=1}^{k} w_j \exp\{-\theta s^a\} \text{ and}$$
$$h(s|k, \mathbf{w}, \theta, \mathbf{a}) = \frac{f(s|k, \mathbf{w}, \theta, \mathbf{a})}{S(s|k, \mathbf{w}, \theta, \mathbf{a})}.$$

Given a sample of lifetime data, we shall now assume that we wish to estimate the parameters of the mixture density: i.e. the number of elements in the mixture, k, and the remaining parameters, θ, $\mathbf{a}$, and $\mathbf{w}$. Furthermore, we should like to estimate other quantities of interest such as the survivor function, mean lifetime etc.

We shall now assume that we observe possibly right censored lifetime data for n patients; $\mathbf{x} = (x_1, ..., x_n)$ where $x_i = (s_i, \delta_i)$, s_i is an observed time point and δ_i is an indicator function:

$$\delta_i = \left\{ \begin{array}{l} 1 \text{ if the lifetime is uncensored, i.e. } S_i = s_i. \\ 0 \text{ if the lifetime is censored, i.e. } S_i > s_i. \end{array} \right\}$$

Given this data, the likelihood function takes the form:

$$l(k, \mathbf{w}, \theta, \mathbf{a}|\text{data}) \propto \prod_{i=1}^{n} \sum_{j=1}^{k} w_j (\theta_j a_j)^{\delta_i} s_i^{(a_j - 1)\delta_i} \exp\{-\theta_j s^{a_j}\}$$

and thus, direct inferential methods become impossible for relatively large values of n. It is possible to consider classical statistical inference via the EM algorithm, see e.g. McLachlan and Peel (2000) for a general review of classical mixture modelling. However there are certain problems associated with estimation of the number of mixture components, k, and we here prefer a Bayesian approach.

In the following section, we illustrate how to undertake Bayesian inference for this model given a sample of data.

3 Bayesian inference for the Weibull mixture model

In this section, we shall first assume that the number of mixture components, k, is known. In order to carry out Bayesian inference, we must first introduce prior distributions for

the remaining model parameters $\mathbf{w}, \theta, \mathbf{a}$. We shall assume the following, relatively diffuse prior distribution structure:

$$\begin{aligned}
&\mathbf{w}|k \sim Dirichlet(\phi, ..., \phi), \\
&a_j|k \sim Gamma\left(\alpha_a, \beta_a\right), \text{ for } j = 1, ..., k. \\
&\theta_j|k \sim Gamma\left(\alpha_\theta, \beta_\theta\right), \text{ for } j = 1, ..., k.
\end{aligned}$$

where typically we might choose $\phi = 1$ and small positive values for $\alpha_a, \beta_a, \alpha_\theta, \beta_\theta$.

In order to carry out posterior inference, we can now set up a Gibbs sampling scheme following the general method introduced by Diebolt and Robert (1994). Firstly, we introduce indicator variables Z_i, for $i = 1, ..., n$, which define from which element of the mixture the i'th observation has been generated. Thus,

$$\begin{aligned}
P(Z_i = j|k, \mathbf{w}) = w_j \text{ and} \\
S_i|k, Z_i = j, \mathbf{a}, \theta \sim W(\cdot|\theta_j, a_j).
\end{aligned}$$

Conditional on the indicators, the likelihood function simplifies to

$$\begin{aligned}
l(k, \mathbf{w}, \theta, \mathbf{a}, \mathbf{z}|k, \text{data}) \propto \\
\propto \prod_{j=1}^{k} (\theta_j a_j)^{\tilde{n}_j} \exp\left\{ a_j \sum_{i:z_i=j}^{n} \log s_i - \theta_j \sum_{i:z_i=j}^{n} s^{a_{z_i}} \right\}
\end{aligned}$$

where $\tilde{n}_j = \#\left\{i : z_i = j \text{ and } \delta_i = 1\right\}$ for $j = 1, \ldots, k$, is the number of uncensored data assigned to element j of the mixture. Combining this with the prior distributions, it is straightforward to calculate the conditional posterior distributions as follows

(i)

$$\mathbf{w}|k, \mathbf{z}, \theta, \mathbf{a}, \text{data} \sim Dirichlet(\phi_1 + n_1, ..., \phi_k + n_k),$$

where $n_j = \#\left\{i : z_i = j\right\}$ is the number of data assigned to element j of the mixture for $j = 1, \ldots, k$.

(ii)

$$\theta_j|k, \mathbf{z}, \mathbf{a}, \text{data} \sim Gamma\left(\tilde{n}_j + \alpha_\theta, \beta_\theta + \sum_{\{i:z_i=j\}} s_i^{a_j}\right),$$

for $j = 1, \ldots, k$, and $\tilde{n}_j = \#\left\{i : z_i = j \text{ and } \delta_i = 1\right\}$.

(iii)

$$f\left(a_j|k,\mathbf{z},\theta,\text{data}\right) \propto g(a_j) \quad \text{where}$$

$$g(a_j) = a_j^{\tilde{n}_j+\alpha_\theta-1}\exp\left\{-a_j\left(\beta_\theta - \sum_{\{i:z_i=j\}}\delta_i\log s_i\right) - \theta_j\sum_{\{i:z_i=j\}}s_i^{a_j}\right\}$$

(iv)

$$\Pr\left(z_i=j|k,\theta,\mathbf{a},\mathbf{w},\text{data}\right) \propto w_j\left(\theta_j a_j s_i^{a_j-1}\right)^{\delta_i}\exp\left\{-\theta_j s_i^{a_j}\right\},$$

for $i=1,\ldots,n$.

Given these conditional distributions, we can now define the following Gibbs sampling algorithm to simulate a sample from the joint posterior distribution.

1. $t=0$. Set initial values $\mathbf{w}^{(0)},\theta^{(0)},\mathbf{a}^{(0)}$
2. $z_i^{(t+1)} \sim z_i|k,\mathbf{w}^{(t)},\theta^{(t)},\mathbf{a}^{(t)}$,data, for $i=1,\ldots,n$
3. $\mathbf{w}^{(t+1)} \sim \mathbf{w}|k,\mathbf{z}^{(t+1)},\theta^{(t)},\mathbf{a}^{(t)}$,data
4. $\theta_j^{(t+1)} \sim \theta_j|k,\mathbf{z}^{(t+1)},\mathbf{w}^{(t+1)},\mathbf{a}^{(t)}$,data, for $j=1,\ldots,k$.
5. $a_j^{(t+1)} \sim a_j|k,\mathbf{z}^{(t+1)},\mathbf{w}^{(t+1)},\theta^{(t+1)}$,data, for $j=1,\ldots,k$.
6. $t=t+1$. Go to Step 2

The only complicated step in this procedure is Step 5: that of sampling the conditional distribution of a_j, for $j=1,\ldots,k$ as in (iii).

Here, we use a slice sampling algorithm (see Neal, 2003). This algorithm proceeds by using the following scheme of simulation:

5(a) First, simulate a uniform random variable; $y \sim U\left(0, g\left(a_j^{(t)}\right)\right)$ where $a_j^{(t)}$ is the current value of a_j and second,

5(b) simulate $a_j^{(t+1)}$ from a uniform distribution with support $S(y)=\{a_j : g(a_j)\geq y\}$.

where $g(a_j)$ is as in (iii).

In practice, the only difficulty with this algorithm is in evaluating the support $S(y)$ although as indicated by Neal (2003), this is straightforward to do by simply sampling from a uniform distribution over a slightly larger space and then checking that the constraint in 5(b) is verified.

In the following section, we indicate how to incorporate uncertainty about the dimension of the mixture, k.

4 Inference when k is unknown

Suppose now, that the number of elements in the Weibull mixture, k, is unknown and that we define a priori density $P(k)$ is defined with support $1, 2, \ldots, k_{\max}$ where typically we will choose $k_{\max} < n$. For example, we might consider a truncated Poisson distribution.

$$P(k) \propto \frac{\gamma^k}{k!} \text{ for } k = 1 \text{ to } k_{\max} \tag{1}$$

For the examples of Section 5, we have used the values $\gamma = 3$ and $k_{max} = 10$ but, in principle, any values could be considered. Other prior structures such as a discrete uniform defined on $[1, k_{max}]$ could also be used.

We can now extend the Gibbs sampling algorithm of Section 3 using a method which allows us to sample over different dimensional spaces (if we change the value of k, the number of model parameters is also altered) such as the reversible jump algorithm (Green 1995, Richardson and Green 1997) or the birth-death MCMC algorithm (Stephens 2000). As shown in Cappé et al (2003), these methods are essentially equivalent. The birth death sampler is usually somewhat easier to implement and has better mixing properties but the reversible jump sampler usually requires a shorter execution time. Here, following Stephens (2000), we consider the use of a birth-death sampler.

In order to implement this sampler, we modify the Gibbs sampling algorithm of Section 3 by replacing k by $k^{(t)}$ throughout and changing step 6 of the algorithm to:

6. Generate $k^{(t+1)}$ and modify the remaining parameters via the following birth-death sampler.

7. $t = t + 1$. Go to Step 2.

In order to carry out step 6, the model parameters are considered as observations from a marked point process and the mixture size k changes so that births and deaths of the mixture components occur in continuous time. The birth rate of the process, $\beta(k, \mathbf{w}, \theta, \mathbf{a})$,

conditional on the current population size and parameter values, $k, \mathbf{w}, \theta, \mathbf{a}$, is preset to a fixed value $\beta(\cdot) = \beta$. Note that if the population has reached its maximum size, $k = k_{\max}$ then it is assumed that further births are impossible and we set $\beta(k_{\max}, \cdot) = 0$.

If a birth of a component occurs, then k is increased by 1 and we generate the missing parameters for the extra component as follows

$$\begin{aligned} w_{k+1} &\sim Beta(1, k), \\ a_{k+1}|k &\sim Gamma\left(\alpha_a, \beta_a\right), \\ \theta_{k+1}|k &\sim Gamma\left(\alpha_\theta, \beta_\theta\right) \end{aligned}$$

i.e. a_{k+1} and θ_{k+1} are generated from their prior distributions. The weights of all remaining components are then rescaled so that they sum to 1, that is $w_j \to w_j/(1 + w_{k+1})$ for $j = 1, \ldots, k$.

Given the current mixture size and parameters, $k, \mathbf{w}, \theta, \mathbf{a}$, a death occurs by selecting one of the components to kill with probability proportional to its death rate $\delta_j(k, \mathbf{w}, \theta, \mathbf{a})$ where

$$\delta_j(k, \mathbf{w}, \theta, \mathbf{a}) = \beta \frac{l(k, \mathbf{w}, \theta, \mathbf{a} \setminus j|\text{data})}{l(k, \mathbf{w}, \theta, \mathbf{a}|\text{data})} \frac{P(k-1)}{kP(k)} \text{ for } j = 1, ..., k$$

where $l(k, \mathbf{w}, \theta, \mathbf{a}|\text{data})$ is the likelihood function for the mixture of all k components and $l(k, \mathbf{w}, \theta, \mathbf{a} \setminus j|\text{data})$ is the likelihood function where the j'th component has been removed and the weights of the other components rescaled to sum to 1, i.e.$w_l \to w_l/(1 - w_j)$ for $l = 1, ..., k$, $l \neq j$. The component is then removed, and the remaining components are then relabeled if necessary and their weights rescaled. The total death rate of the process is then

$$\delta(k, \mathbf{w}, \theta, \mathbf{a}) = \sum_{j=1}^{k} \delta_j(k, \mathbf{w}, \theta, \mathbf{a}).$$

Note that if the current value of k is $k = 1$, it is assumed that deaths are impossible; i.e. the death rate is $\delta(k = 1, \mathbf{w}, \theta, \mathbf{a}) = 0$.

In order to simulate realizations from this birth-death process, we thus follow the following simple algorithm

a. Start from initial values $k, \mathbf{w}, \theta, \mathbf{a}$.

b. Calculate death rates $\delta_j(k, \mathbf{w}, \theta, \mathbf{a})$ for $j = 1, ..., k$.

c. Generate the time to first birth or death from an exponential distribution with mean $\frac{1}{\beta + \delta(k, \mathbf{w}, \theta, \mathbf{a})}$.

d. Elect a birth with probability $\frac{\beta}{\beta+\delta(k,\mathbf{w},\theta,\mathbf{a})}$ or the death of component j with probability $\frac{\delta_j(k,\mathbf{w},\theta,\mathbf{a})}{\beta+\delta(k,\mathbf{w},\theta,\mathbf{a})}$ for $j = 1, ..., k$.

e. Modify the model parameters $k, \mathbf{w}, \theta, \mathbf{a}$ accordingly.

f. Go to b.

Thus, in order to move from $k^{(t)}$ to$k^{(t+1)}$ in Step 6 of our Gibbs sampling algorithm, we simply start the birth-death process from the current values $k^{(t)}, \mathbf{a}^{(t+1)}, \theta^{(t+1)}, \mathbf{w}^{(t+1)}$ and then run the process during a fixed time period t_0, for example $t_0 = 1$. We then reset $k^{(t+1)}, \mathbf{a}^{(t+1)}, \theta^{(t+1)}, \mathbf{w}^{(t+1)}$ to be the values of the mixture size and model parameters after this time period has elapsed. For details of why this algorithm works and examples in the context of mixtures of normal and t distributions, see Stephens (2000).

Given the Gibbs sample output, we can estimate various quantities of interest, such as the probability that the true distribution is a single Weibull distribution,

$$P(k = 1|\text{data}) \approx \frac{1}{N}\#\{k^{(t)} = 1\}$$

or th population survival function,

$$S(s|\text{data}) \approx \frac{1}{N}\sum_{t=1}^{N}\left[\sum_{j=1}^{k(t)} w_j^{(t)} \exp\left\{-\theta_j^{(t)} s^{a_j^{(t)}}\right\}\right].$$

Other quantities such as the mean survival time of a patient or the expected value of the hazard function can be estimated similarly.

5 Examples

5.1 Simulated Data

Firstly we simulated a sample of size 150 with 10% censoring from the three component mixture:

$$f(s|\cdot) = 0.6\ W(s|0.1, 0.5) + 0.3\ W(s|0.3, 1) + 0.1\ W(s|0.5, 2).$$

12 of the sampled data were right censored and the remaining 138 were completely observed. The prior distribution scheme outlined in the previous sections was used and we ran the birth death MCMC algorithm (4) for 60000 observations (10000 to burn-in)

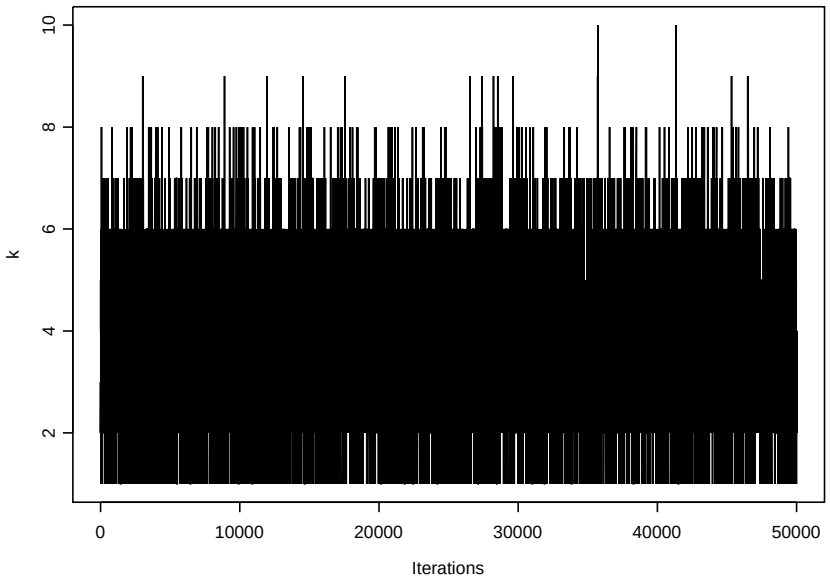

Figure 1: Plot of mixture size k versus iteration of the MCMC algorithm

with birth rate $\beta = 3$. In Figure 1, we show a plot of mixture size $k^{(t)}$ against the iteration t of the MCMC algorithm.

The plot illustrates that the mixing of the algorithm seems to be quite good; the mixture size moves between the values of 1 and 10 without remaining in the same place for too long. Note that we also produced a graph of the estimated mean of k against the number of MCMC iterations which also suggested that the algorithm was in equilibrium after around 10000 iterations.

Note also that we also ran the birth death algorithm with different values of the birth parameter β and in general we noted that large values gave better mixing but slower convergence while with small values of β, the algorithm mixed more slowly. Thus, the value used here seems to be a reasonable choice.

In Figure 2 we illustrate the estimated posterior distribution of the mixture size k. It can be seen that the posterior mode is the true value $k = 3$ and a 95% highest posterior density interval for k is given by $[2, 6]$ illustrating that there is some posterior uncertainty about the true mixture size.

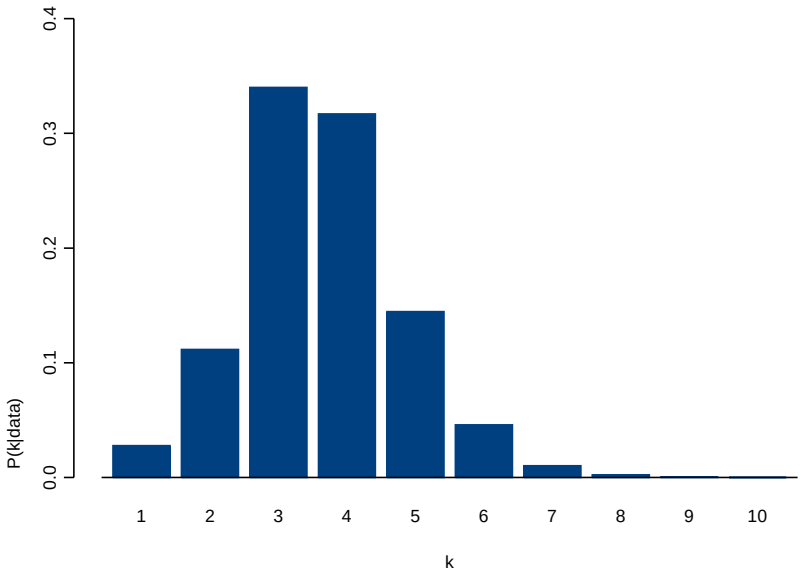

Figure 2: Posterior distribution of the mixture size k.

In Figure 3 we compare the *Kaplan-Meier* estimator and the posterior mean of the

estimated survival function with the true survivor function The fitted curve is somewhat closer to the true suvivor function than is the Kaplan Meier estimator.

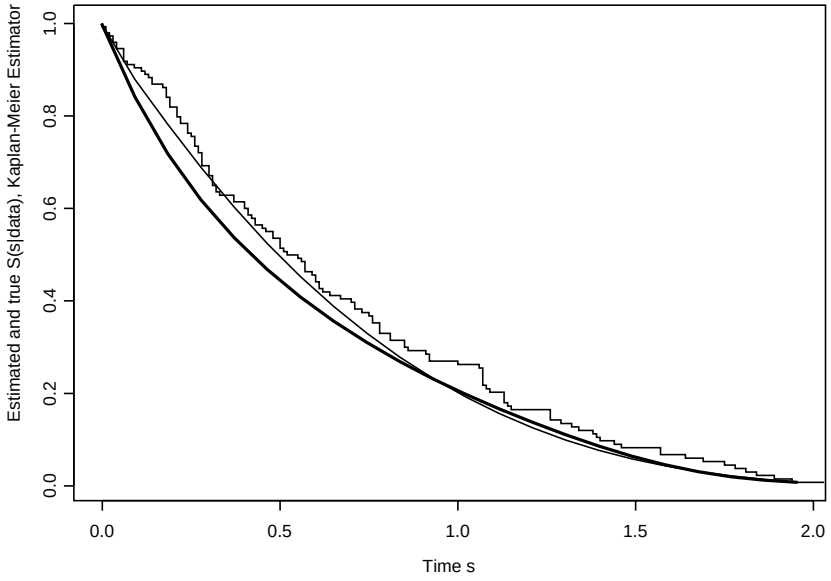

Figure 3: Expected Survivor Function (thin black line), True Survivor Function (thick black line) and Kaplan Meier Estimate (polygonal line).

In the following section, we consider the analysis of some real data.

5.2 Real Data Problem

Here we analyse data from 87 persons with *lupus nephritis (see Abrahamowicz et al., 1996)* These patients were studied over a 15 year time period, during which 35 deaths were recorded. As in the simulated example, we used the same prior distributions and a birth death MCMC algorithm with 60000 iterations (10000 to burn-in) to fit the data. In this example, there was again some uncertainty concerning the number of mixture components with a posterior credible interval for the mixture size being [3,8]. In Figure 4 we illustrate the predicted survivor function and the Kaplan Meier estimator. There appears to be a good correspondence between the two.

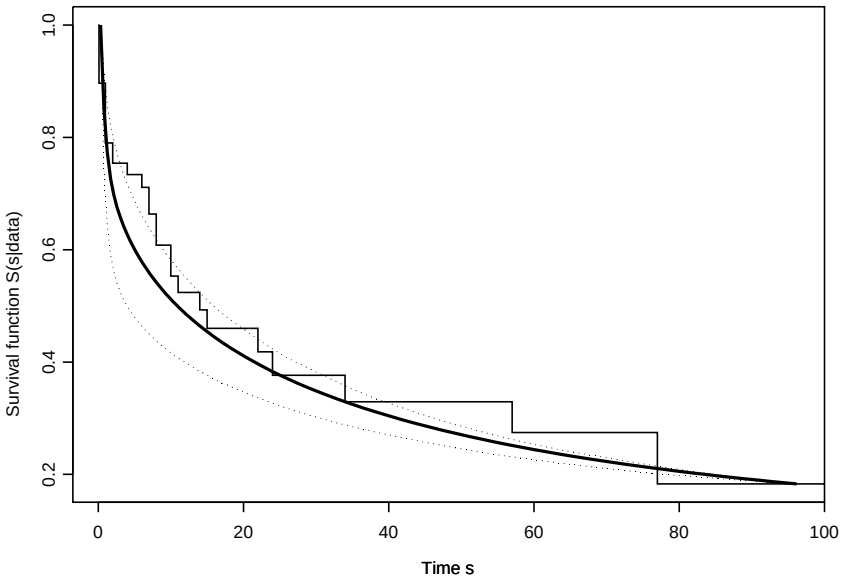

Figure 4: Fitted survival curve and Kaplan Meier estimator for the Lupus data.

Finally, in Figure 5, we illustrate the expected hazard function for this data set. The expected hazard falls quite rapidly towards 0.05.

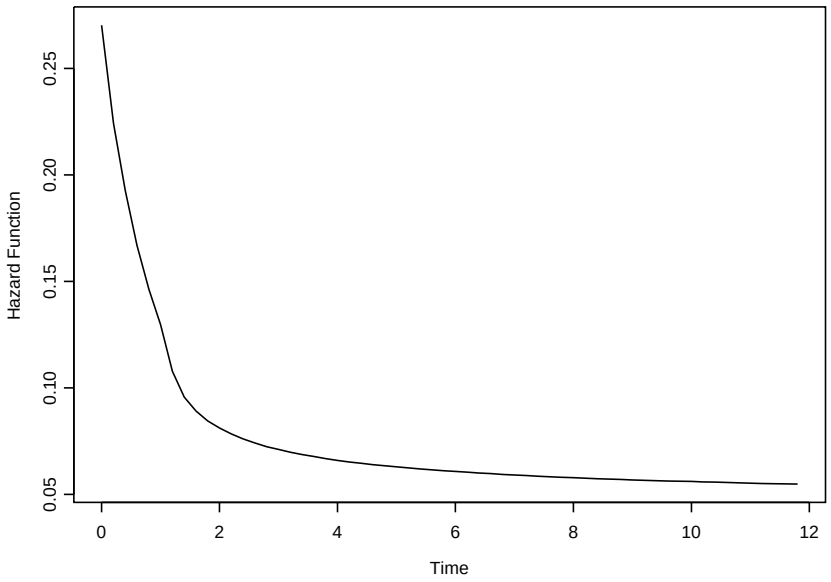

Figure 5: Expected hazard function for the Lupus data.

6 Conclusions and further developments

In this article, we have illustrated how Bayesian methods can be used to fit a mixture of Weibulls model with an unknown number of components to heterogeneous, possibly right censored survival data using a birth death MCMC algorithm. Some extensions and modifications are possible.

Firstly, given the symmetrical prior distribution structure used here, even for fixed mixture size k, the model is unidentifiable in the sense that the posterior distribution for the remaining mixture parameters has $k!$ modes. This is not a problem in terms of prediction of the reliability or hazard functions, but would be a problem if we wished to make inference about the individual elements of the mixture. In this case, an alternative would be to place a restriction on the prior parameter space, e.g. $w_1 > w_2 > ... > w_k$

which would make the model identifiable. Given such a restriction, we would then need to use a different algorithm to perform the MCMC sampling over different values of the mixture size. This could easily be done by using the reversible jump algorithm of Green (1995) and Richardson and Green (1997). Investigation is currently underway on comparing the performance of the two algorithms.

Secondly, we have assumed here that although we have a possibly heterogeneous population, no covariate information is available. One extension would be to consider the inclusion of covariate information to help predict the element of the mixture from which each observation comes.

Finally, although we have here used Bayesian methods to fit the mixture model, it would also be possible to consider classical mixture modelling via the EM algorithm; see e.g. McLachlan and Peel (2000) for an introduction. This would be relatively straightforward to implement assuming that the mixture size is known but, for unknown k, some type of information criterion needs to be used to select the mixture size.

Bibliography

Abrahamowicz, M., Mackenzie, T. and J. Esdaile (1996). Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. *J. Amer. Statist. Assoc.*, **91**, 1432-1439.

Berkson, J. and R. P. Gage (1952). Survival curves for cancer patients following treatment. *J. Amer. Statist. Assoc.*, **47**, 501-515.

Cappé, O., Robert, C.P. and Rydén, T. (2002). Reversible jump birth-and-death and more general continuous time samplers. *Unpublished Research Report*, CEREMADE, Université de Paris

Chen, W. C., Hill, B. M., Greenhouse, J. B. and J. V. Fayos (1985). Bayesian Analysis of Survival Curves for Cancer Patients Following Treatment. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds., *Bayesian Statistics 2*, 299-328,.North Holland: Amsterdam..

Diebolt, J. and C.P. Robert (1994). Estimation of finite mixture models through Bayesian sampling. *J. Royal Statist. Soc., B*, **56**, 363-375.

Dodson, B. (1994). *Weibull Analysis*. ASQ, Milwaukee.

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

Green, P. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, **82**, 711-732.

Ibrahim, J.G., Chen, M.H. and D. Sinha. (2001). *Bayesian Survival Analysis*. Springer Verlag: Berlin.

Kottas, A. (2002). Nonparametric Bayesian Survival Analysis using Mixtures of Weibull Distributions. Unpublished manuscript. Available from: `http://www.ams.ucsc.edu/~thanos/research.html`

McLachlan, G.J. and D. Peel (2000). *Finite Mixture Modeling.* John Wiley: New York.

Neal, R. (2003). Slice Sampling. *Ann. Stat.* In press.

Quiang, J. (1994). *A Bayesian Weibull Survival Model.* Unpublished Ph.D. Thesis. ISDS, Duke University.

Reynolds, J.F. (1973). On estimating the parameters of a birth-death process. *Australian J. Statistics*, **15**, 35-43.

Richardson, S. and P. Green. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Royal Statist. Soc., B*, **59**, 731-792.

Robert , C. (1996). Mixtures of distributions : inference and estimation. In W.R. Gilks, S.Richardson and D. J. Spiegelhalter, eds., *Markov Chain Monte Carlo in Practice*, 441-464. Chapman & Hall: London.

Stephens, M. (2000). Bayesian analysis of mixtures with an unknown number of components — an alternative to reversible jump methods. *Ann. Stat.*, **28**.

Wiper, M.P., Rios Insua, D. and F. Ruggeri. (2001). Mixtures of gamma distributions with applications. *J. Computational and Graphical Statist.*, **10**, 440-454.