IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 13, Issue 2 Ver. IV (Mar. - Apr. 2017), PP 19-32
www.iosrjournals.org

Bayesian Estimation of Survivor Function for Censored Data Using Lognormal Mixture Distributions

Henry Ondicho Nyambega[1], Dr. George O. Orwa[2], Dr. Joseph K. Mung'atu[3] and Prof. Romanus O. Odhiambo[4]

[1] *Department of Mathematics and Actuarial Science, Kisii University, Kisii, Kenya*
[2,3,4] *Department of Statistics and Actuarial Science, JKUAT, Nairobi, Kenya*

Abstract: *We use Bayesian methods to fit a lognormal mixture model with two components to right censored survival data to estimate the survivor function. This is done using a simulation-based Bayesian framework employing a prior distribution of the Dirichlet process. The study provides an MCMC computational algorithm to obtaining the posterior distribution of a Dirichlet process mixture model (DPMM). In particular, Gibbs sampling through use of the WinBUGS package is used to generate random samples from the complex posterior distribution through direct successive simulations from the component conditional distributions. With these samples, a Dirichlet process mixture model with a lognormal kernel (DPLNMM) in the presence of censoring is implemented.*

Keywords: *Bayesian, Lognormal, Survivor Function, Finite Mixture models, Win BUGS*

I. Introduction

In many cases of statistical applications observed survival data may be censored. The data may also be generated from several homogenous subgroups regarding one or several characteristics (Singpurwalla, 2006), for example when patients are given different treatments. Furthermore, across the subgroups, heterogeneity may often be encountered. In such cases, the traditional methods of estimation may not sufficiently describe the complexity in these data. To produce better inferences, we consider mixture models (Stephens, 1997) which assume these data as represented by weighted sum of distributions, with each distribution defined by a unique parameter set representing a subspace of the population.

There has been an increased popularity of DP mixture models in Bayesian data analysis. According to Kottas, 2006, the Dirichlet Process (DP) prior for mixing portions can be handled by both a Bayesian framework through an MCMC algorithm. Furthermore, the DP prior fulfills the two properties proposed by Ferguson, 1973. First, it is flexible in support of prior distributions and the posteriors can be tractably analyzed. Second, it can capture the number K of unknown mixture components.

In the Bayesian context, Qiang, 1994 used a mixture of a Weibull component and a surviving fraction in the context of a lung cancer clinical trial. Tsionas, 2002 considered a finite mixture of Weibull distributions with a larger number of components for capturing the form of a particular survival function. Marin et al., 2005a described methods to fit a Weibull mixture model with an unknown number of components. Kottas, 2006 developed a DPM model with a Weibull kernel (DPWM) for survival analysis. Hanson, 2006 modeled censored lifetime data using a mixture of gammas. More recently, Farcomeni and Nardi, 2010 proposed a two component mixture to describe survival times after an invasive treatment. We consider a lognormal mixture distribution.

The Bayesian approach considers unknown parameters as random variables that are characterized by a prior distribution. This prior distribution is placed on the class of all distribution functions, and then combined with the likelihood function to obtain the posterior distribution of the parameter of interest on which the statistical inference is based (Singpurwalla, 2006).

In this paper we carry out posterior inference by sampling from the posterior distribution using simulation employing Markov Chain Monte Carlo (MCMC) methods. We employ the Gibbs Sampler through the Win BUGS (WinBUGS, 2001) software.

The rest of this paper is organized as follows. In Section 2, we define the mixture of lognormal model that will be considered. 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 through WinBUGS software. In Section 3, we illustrate the model using both simulated and real data sets and finally, in Section 4 we summarize our results and consider some possible extensions.

II. The Lognormal Mixture Model

2.1 Review of Bayesian Estimation

Let $t_1,...t_n$ be a random sample taken from a population indexed by the parameter θ, and the prior distribution is updated using the information from the sample. Suppose $f(\theta)$ is the prior distribution of θ. Then, $f(\theta)$ expresses what is known about θ prior to observing the data $\mathbf{t} = \{t_i; i = 1,...,n\}$.

The Bayesian approach is based on four tenets. First, is to decide on the prior. Secondly, decide on the likelihood.

$$f(\mathbf{t} \mid \theta) = \prod_{i=1}^{n} f(t_i \mid \theta) \tag{1}$$

which describes the process giving rise to the data in terms of unknown θ. Accordingly,

$$f(\mathbf{t} \mid \theta) = \frac{f(t, \theta)}{f(\theta)} \tag{2}$$

The third step in Bayesian estimation is the derivation of the posterior distribution through Bayes theorem by combining information contained in the prior distribution with information about the observed data in the likelihood, as

$$f(\theta \mid \mathbf{t}) = \frac{f(\theta, \mathbf{t})}{f(\mathbf{t})} = \frac{f(\mathbf{t} \mid \theta) f(\theta)}{f(\mathbf{t})} \propto f(\mathbf{t} \mid \theta) f(\theta) \tag{3}$$

This expresses what is known about θ after observing data $\mathbf{t} = \{t_i; i = 1,...,n\}$ and results in the making of a specific probability statement about the unknown parameter, given the data. The term

$$f(\mathbf{t}) = \int f(\theta \mid \mathbf{t}) d\theta \tag{4}$$

given by the marginal density of the $t_i; i = 1,...,n$, is the normalizing factor (Lindley, 1961) which ensures that

$$\int f(\theta \mid \mathbf{t}) d\theta = 1 \tag{5}$$

Assuming $\mathbf{t} = \{t_i; i = 1,...,n\}$ are independent observations on a random variable T, then, equation (3) can be re-written as

$$f(\theta \mid \mathbf{t}) = \frac{\prod_{i=1}^{n} f(t_i \mid \theta) f(\theta)}{\int f(\mathbf{t} \mid \theta) f(\theta) d\theta} \tag{6}$$

which represents the posterior distribution when sampled observations are available.

The posterior Bayes estimator of θ is the mean of equation (6). That is

$$\hat{\theta} = E(\theta) = \int \theta f(\theta \mid \mathbf{t}) d\theta \tag{7}$$

The fourth and the last step is deriving inference from the posterior. For complex posterior distributions, equation (7) is not tractable. As an alternative, we use MCMC sampling algorithms to sample from the posterior distribution.

Thus before any data are available, only the prior distribution $f(\theta)$ is used for inference. When a set of data, say $t^{(1)}$, are observed, the posterior distribution

$$f(\theta \mid t^{(1)}) \propto f(t^{(1)} \mid \theta) f(\theta) \tag{8}$$

while when a second set of data is available, we use the posterior from the first instance as a prior and incorporate the new data in a new updated posterior distribution, to obtain the updated posterior distribution as

$$f(\theta \mid t^{(1)}, t^{(2)}) \propto f(t^{(2)} \mid \theta) f(\theta \mid t^{(1)}) \propto f(t^{(2)} \mid \theta) f(t^{(1)} \mid \theta) f(\theta) \tag{9}$$

For data collected in n different time instances equation (9) can be generalized as

$$f(\theta \mid t^{(1)},...,t^{(n)}) \propto f(t^{(n)} \mid \theta) f(\theta \mid t^{(1)},...,t^{(n-1)}) \propto \prod_{i=1}^{n} f(t^{(n)} \mid \theta) f(t^{(1)} \mid \theta) f(\theta) \tag{10}$$

2.2 Review of Mixture Models

Let $\mathbf{t} = \{t_i; i = 1,...,n\}$ be a vector of n observations. A mixture model can be written as

$$f(t) = \sum_{i=1}^{n} \omega_j f(t_j \mid \theta_j) \tag{11}$$

where $f(t)$ is a finite mixture density function with different parameters θ_j, ω_j are mixing weights satisfying, $0 \le \omega_j \le 1$ with $\sum_{i=1}^{K} \omega_j = 1$ and $f(t_j \mid \theta_j)$ are the component densities of the mixture.

For this finite mixture model, we treat the number of subgroups, K representing the data under study as known. As the data size grows and data become more complicated, an infinite number of prior information is theoretically assigned for growing with data, giving a hierarchical representation.

The proportion of data explained by a subgroup j is represented by the component weight ω_j, while each component is also described by its own distribution $f(t_j \mid \theta_j)$, defined by component specific parameter θ_j.

If the components come from a parametric family, $f(t_j \mid \theta_j)$ with unknown parameters θ_j, then the parametric mixture model is

$$f(\mathbf{t} \mid \Psi) = \sum_{j=1}^{K} \omega_j f(t_j \mid \theta_j) \tag{12}$$

where θ is the collection of all distinct parameters occurring in the component densities, and Ψ the complete collection of all distinct parameters occurring in the mixture model.

In the Bayesian analysis of the model, we assume that $\omega \mid K \sim Dir(\nu_1,...,\nu_n)$, where the $\nu_i; i = 1,...,n$'s are fixed constants. Also, the component parameters θ_j are assumed a priori independent, conditionally on K and, possibly, a vector of hyperparameters, ϕ

$$f(\theta \mid K, \phi) = \prod_{j=1}^{K} f(\theta_j \mid \phi_j) \tag{13}$$

If a prior distribution $f(\theta)$ is specified, then a sample from the joint posterior of (K, ω, θ) can be obtained by means of Markov chain Monte Carlo methods (Nobile and Fearnside, 2005). However, inference about ω and θ is not straightforward, because the likelihood is invariant with respect to permutations of the components' labels.

In the Bayesian framework, a DP prior is assigned to the mixture model with a kernel distribution, to form a DP Mixture Model (DPMM). We write the DP mixture model as

$$F(t, G) = \int f(t \mid \theta) G(d\theta) \tag{14}$$

where $f(t \mid \theta)$ is the probability density function (PDF) of a parametric kernel with parameter vector θ.

If we set G as a DP prior, then $G \sim DP(\nu, G_0)$ denotes a Dirichlet Process prior placed on the random distribution function G. Thus $DP(\nu, G_0)$ is the Dirichlet process with a base distribution G_0, an infinite-dimensional distributional parameter (McAuliffe et al., 2006), which makes the DPMM a nonparametric method and ν is a positive scalar precision parameter.

To allow additional modeling flexibility, independent prior distributions, $[\nu]$ and $[\phi]$ are placed on ν and the parameters of $G_0 = G_0(\bullet \mid \phi)$ are specified to ν and G_0 to give the full hierarchical model as

$$\begin{aligned} t \mid \theta &\sim f(t \mid \theta_i) \\ \theta_i \mid G &\sim G(\theta) \\ G \mid \nu, \phi &\sim DP(\nu G_0(\theta)) \\ \nu, \phi &\sim [\nu][\varphi] \end{aligned} \tag{15}$$

2.3 The DPLNMM in Bayesian Framework

Consider a vector of n survival times $\mathbf{t} = \{t_i; i = 1,...,n\}$ that takes values in a space $\Omega_{1,...,K}$. A DPLN mixture model for t can be written as

$$f(t_j) \sim \sum_{i=1}^{n} \omega_j f(t_j \mid \mu_j, s_j^{\,2}), i = 1,...,n \tag{16}$$

where ω_j are mixing weights satisfying, $\omega_j > 0$ with $\sum_{j=1}^{K} \omega_j = 1$ and $f(t \mid \mu_j, s_j^2),\ j = 1,...,K$ is a kernel density of the lognormal distribution given by

$$f(t \mid \mu, s^2) = \frac{1}{\sqrt{2\pi} st} \exp\left\{ - \frac{(\log(t) - \mu)^2}{2s^2} \right\}, t > 0 \tag{17}$$

where $\mu > 0$ is the scale parameter and $s^2 > 0$ is the shape parameter (Ibrahim et al., 2001b).

For the DPLN mixture model, equation (15) the number of components K is known while μ, s^2 and ω are subject to inference. Thus if we let

$$x_{ij} = \begin{cases} 1, \ if \quad ith \quad unit\, is\, drawn\, from\, the\, jth\, mixture component \\ 0, \quad elsewhere \end{cases} \tag{18}$$

then $\omega_j = p(x_{ij}) = 1$.

For a mixture model with K components, the likelihood of a single t_i is given by

$$f(t_i \mid \omega, \mu, s^2) = \sum_{j=1}^{K} \omega_j f(t_j \mid \mu_j, s_j^{\,2}), i = 1,...,n \tag{19}$$

and for a vector of observations $\mathbf{t} = \{t_i; i = 1,...,n\}$,

$$f(\mathbf{t} \mid \omega, \mu, s^2) = \prod_{i=1}^{n} \sum_{j=1}^{K} \omega_j f(t_j \mid \mu_j, s_j^{\,2}), i = 1,...,n \tag{20}$$

Thus the joint Likelihood is

$$f(\mathbf{t} \mid x, \omega, \mu, s^2) = \prod_{i=1}^{n} \sum_{j=1}^{K} \left(\omega_j(x_{ij}) f(t_j \mid \mu_j, s_j^{\,2}) \right)^{x_{ij}}, i = 1,...,n \tag{21}$$

For this lognormal distribution we conveniently choose the following prior distributions for the unknown parameters, and accordingly, write the DPLNM model hierarchically as

$$\begin{aligned} & t_j \mid \mu_j, s_j^2 \sim f(t_j \mid \mu_j, s_j^2) \\ & (\mu_j, s_j^2) \mid G \sim G \\ & G \mid \nu, \beta, \theta, \sigma^2 \sim DP(\nu G_0) \\ & \nu \sim Gamma(\alpha_\nu, \beta_\nu) \\ & G_0 \sim inverse - Gamma(s^2 \mid \alpha, \beta) \cdot Normal(\mu \mid \theta, \sigma^2) \\ & \theta \sim Normal(\mu_\theta, \sigma_\theta^{\,2}) \\ & \sigma^2 \sim Inverse - Gamma(\alpha_\sigma, \beta_\alpha) \\ & \beta \sim Gamma(\alpha_\beta, \beta_\beta) \end{aligned} \tag{22}$$

The joint prior can be expressed as

$$f(\mu, s^2, \omega \mid \theta, \sigma^2, \alpha, \beta, \nu) = f(\omega \mid \nu) f(\mu \mid s^2, \sigma^2) f(s^2 \mid \alpha, \beta) \tag{23}$$

Now according to Lindley, 1961 posterior distribution is calculated through Bayes Theorem as

$$f(\theta \mid t) = \frac{f(t \mid \theta) f(\theta)}{\int f(\theta \mid t) d\theta} \propto f(t \mid \theta) f(\theta) \tag{24}$$

Thus by combining the likelihood and the prior, the posterior of μ and s^2 is given as

$$f(\mu,s^2,\omega\,|\,t,x,\theta,\sigma^2,\alpha,\beta,\nu)\propto f(\mu,s^2,\omega\,|\,\theta,\sigma^2,\alpha,\beta,\nu)\,f(t\,|\,x,\mu,s^2,\omega)$$

$$=\frac{\Gamma(\nu_1+\ldots+\nu_K)}{\Gamma(\nu_1)\ldots\Gamma(\nu_K)}\omega^{\nu_1-1}\ldots\omega^{\nu_K-i}\left[\prod_{j=1}^{K}\frac{1}{st\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(\mu_j-\theta)^2\right\}\right]\times$$

$$\left[\prod_{j=1}^{K}\frac{\beta^\alpha}{\Gamma(\alpha)}(s_j)^{-(\alpha+1)}\exp\left\{\frac{-\beta}{s_j^2}\right\}\right]\sum_{j=1}^{K}\left[\omega_j\frac{1}{st\sqrt{2\pi}}\exp\left\{-\frac{1}{2s_j^2}\left(\log(t_i)-\mu_j\right)^2\right\}\right]^{\sum_{i=1}^{n}x_{ij}}$$

$$=\frac{\Gamma(\nu_1+\ldots+\nu_K)}{\Gamma(\nu_1)\ldots\Gamma(\nu_K)}\omega^{\nu_1-1}\ldots\omega^{\nu_K-1}\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{K}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{K}(\mu_j-\theta)^2\right\}\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^K\left[\prod_{j=1}^{K}\left(s_j^K\right)^{-(\alpha+1)}\right]\times$$

$$\exp\left\{-\beta\sum_{j=1}^{K}\frac{1}{s_j^2}\right\}\left[\sum_{j=1}^{K}\omega_j^{n_j}\left(\frac{1}{2\pi st}\right)\right]^{\frac{n_j}{2}}\exp\left\{-\frac{1}{2s_j^2}\sum_{i=1}^{n}x_{ij}\left(\log(t_i)-\mu_j\right)^2\right\}$$

$$=\frac{\Gamma(\nu_1+\ldots+\nu_K)}{\Gamma(\nu_1)\ldots\Gamma(\nu_K)}\omega^{\nu_1-1}\ldots\omega^{\nu_K-1}\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{K}{2}}\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^K\left[\prod_{j=1}^{K}\left(s_j^K\right)^{-(\alpha+1)}\right]\exp\left\{-\beta\sum_{j=1}^{K}\frac{1}{s_j^2}\right\}\times$$

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{K}\left(\mu_j-\theta\right)^2\right\}\left[\sum_{j=1}^{K}\omega_j{}^{n_j}\left(\frac{1}{2\pi st}\right)^{\frac{n_j}{2}}\right]\exp\left\{-\frac{1}{2s_j^2}\sum_{i=1}^{n}x_{ij}\left(\log(t_i)-\mu_j\right)^2\right\} \tag{25}$$

which is a mixture model. Setting the value of the normalizing factor $f(\mathbf{t})=d$ we have

$$d=\left(\int f(\mu,s^2,\omega\,|\,\theta,\sigma^2,\alpha.\beta,\nu)\cdot f(t\,|\,x,\mu,s^2,\omega)\right)^{-1} \tag{26}$$

From which

$$d^{-1}=\int_0^\infty\int_{-\infty}^{\infty}f(\mu,s^2,\omega\,|\,t,x,\theta,\sigma^2,\alpha,\beta,\nu)ds^2d\mu$$

$$=\int_0^\infty\int_{-\infty}^{\infty}\left(\frac{1}{s^2}\right)^{\frac{n}{2}}\exp\left\{\frac{\sum_{i=1}^{n}\left(\log(t_i)\right)^2-\frac{\left(\sum_{i=1}^{n}\log(t_i)\right)^2}{n}}{2s^2}\right\}\exp\left\{-\frac{n}{2s^2}\left(\mu-\frac{\sum_{i=1}^{n}\left(\log(t_i)\right)}{n}\right)^2\right\}ds^2d\mu$$

$$=\sqrt{\frac{2\pi}{n}}\int_0^\infty\frac{\exp\left\{\frac{\sum_{i=1}^{n}\left(\log(t_i)\right)^2-\frac{\sum_{i=1}^{n}\left(\log(t_i)\right)^2}{n}}{2s^2}\right\}}{\left(s^2\right)^{n-\frac{1}{2}}}ds^2$$

$$=\sqrt{\frac{2\pi}{n}}\frac{\Gamma\left(\frac{n-3}{2}\right)}{\left(\frac{\sum_{i=1}^{n}\left(\log(t_i)\right)^2-\frac{\sum_{i=1}^{n}\left(\log(t_i)\right)^2}{n}}{2s^2}\right)^{\frac{n-3}{2}}} \tag{27}$$

Therefore the value of d is

$$d=\sqrt{\frac{n}{2\pi}}\frac{\left(\sum_{i=1}^{n}(\log(t_i)^2)-\frac{\left(\sum_{i=1}^{n}(\log(t_i))^2\right)}{n}\right)^{\frac{n-3}{2}}}{2^{\frac{n-3}{2}}\Gamma\left(\frac{n-3}{2}\right)} \tag{28}$$

Thus the full conditional posterior for μ and s^2 then becomes

$$f(\mu,s^2,\omega\,|\,t,x,\theta,\sigma^2,\alpha,\beta,\nu)=\frac{\Gamma(\nu_1+\ldots+\nu_K)}{\Gamma(\nu_1)\ldots\Gamma(\nu_K)}\omega^{\nu_1-1}\ldots\omega^{\nu_K-1}\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{K}{2}}\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^K\left[\prod_{j=1}^{K}\left(s_j^K\right)^{-(\alpha+1)}\right]\exp\left\{-\beta\sum_{j=1}^{K}\frac{1}{s_j^2}\right\}$$

$$\times\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{K}(\mu_j-\theta)^2\right\}\left[\sqrt{\frac{n}{2\pi}}\frac{\left(\sum_{i=1}^{n}(\log(t_i))^2-\frac{\left(\sum_{i=1}^{n}(\log(t_i))^2\right)}{n}\right)^{\frac{n-3}{2}}}{2^{\frac{n-3}{2}}\Gamma\left(\frac{n-3}{2}\right)\left(s^2\right)^{\frac{n}{2}}}\right]\exp\left\{\frac{\left(\sum_{i=1}^{n}(\log(t_i))\right)-\sum_{i=1}^{n}(\log(t_i))^2}{2s^2}\right\}\times$$

$$\exp\left\{-\frac{n}{2s^2}\left(\mu-\frac{\sum_{i=1}^{n}(\log(t_i))}{n}\right)^2\right\} \tag{29}$$

which is Dirichlet Process mixture model.

For each observation t_i, we define an indicator variable as

$$\delta_i=\begin{cases}1,\ if\ t_i\ is\ an\ uncensored\ failure\ time,\\ 0,\ if\ t_i\ is\ a\ \ censoring\ (right)\ time\end{cases} \tag{30}$$

If t_i is an uncensored failure time, that is, $i=0$, the full conditional DP mixture model is as given by equation (29). For a rightly censored observation t_i, $i=1$, then the posterior is given as

$$f(\mu,s^2,\omega\,|\,t,x,\theta,\sigma^2,\alpha,\beta,\nu)\propto f(\mu,s^2,\omega\,|\,\theta,\sigma^2,\alpha,\beta,\nu)(1-f(t\,|\,x,\mu,s^2,\omega))$$

$$=\frac{\Gamma(\nu_1+\ldots+\nu_K)}{\Gamma(\nu_1)\ldots\Gamma(\nu_K)}\omega^{\nu_1-1}\ldots\omega^{\nu_K-i}\left[\prod_{j=1}^{K}\frac{1}{st\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(\mu_j-\theta)^2\right\}\right]\times \tag{31}$$

$$\left[\prod_{j=1}^{K}\frac{\beta^\alpha}{\Gamma(\alpha)}(s_j)^{-(\alpha+1)}\exp\left\{\frac{-\beta}{s_j^2}\right\}\right]\sum_{j=1}^{K}\left[1-\omega_j\frac{1}{st\sqrt{2\pi}}\exp\left\{-\frac{1}{2s_j^2}\left(\log(t_i)-\mu_j\right)^2\right\}\right]^{\sum_{i=1}^{n}x_{ij}}$$

so that the full conditional posterior distribution of the model can be written as

$$f(\mu, s^2, \omega \mid t, x, \theta, \sigma^2, \alpha, \beta, \nu) = \frac{\Gamma(\nu_1 + \ldots + \nu_K)}{\Gamma(\nu_1)\ldots\Gamma(\nu_K)} \omega^{\nu_1 - 1} \ldots \omega^{\nu_K - 1} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{K}{2}} \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^K \left[\prod_{j=1}^{K} \left(s_j^K\right)^{-(\alpha+1)}\right] \exp\left\{-\beta \sum_{j=1}^{K} \frac{1}{s_j^2}\right\} \times$$

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{K}(\mu_j - \theta)^2\right\} \left[\sqrt{\frac{n}{2\pi}} \frac{\left(\sum_{i=1}^{n}(\log(t_i))^2 - \frac{\left(\sum_{i=1}^{n}(\log(t_i))^2\right)}{n}\right)^{\frac{n-3}{2}}}{2^{\frac{n-3}{2}}\Gamma\left(\frac{n-3}{2}\right)\left(s^2\right)^{\frac{n}{2}}}\right] \left(s_j^2\right)^{-((\alpha+n_j)/2+1)} \exp\left\{\frac{\beta + 0.5\left(\sum_{i,j}(\log(t_i)) - \mu\right)^2}{s^2}\right\} \times$$

$$\prod_{i,j}\left(1 - \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}(\mu_j - \theta)^2\right\}\left(\frac{(l0g(t_i)) - \mu_j}{s_j}\right)\right) \tag{32}$$

where n_j are the number of uncensored failure times in the j^{th} cluster.

2.4 Model Implementation by Gibbs Sampling

2.4.1 Review of Gibbs Sampling

This section describes the Gibbs sampling. The overall aim of Gibbs sampling is to simulate from the complex posterior density by creating a Markov chain with the posterior density as its stationary distribution. This is done by direct successive simulations from the component conditional distributions. Giudici et al., 2009 have formulated the Gibbs Sampling algorithm as

$$\theta_1^{(t)} \text{ is sampled from } f(\theta_1) \mid \theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)}, D);$$

$$\theta_1^{(t)} \text{is sampled from } f(\theta_2 \mid \theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)}, D); \tag{33}$$

$$\theta_1^{(t)} \text{ is sampled from } f(\theta_3 \mid \theta_1^{(t)}, \theta_2^{(t)}, \theta_4^{(t-1)}, \ldots, \theta_d^{(t-1)}; D);$$

where $\theta_1, \ldots, \theta_d$ represent the parameter of the model, and the D, is the data. The values of iteration N would always be sampled from the previous values from iteration (N-1). The distribution $f(\theta_j \mid \theta_{-j}, D)$ where $\theta_{-j} = \theta_1, \theta_2, \ldots, \theta_{j-1}, \ldots, \theta_d$, is the full conditional distribution and is the proposal distribution required by Gibbs Sampling (Giudici et al., 2009).

We note from Escobar & West, 1995 that the Gibbs sampler and its various adaptations has been the most commonly used approach in Bayesian mixture estimation. This is because for many Bayesian models, its implementation is particularly convenient due to two properties. First, the conditional conjugacy property ensures that the posterior conditional distributions required by the Gibbs sampler are from the same family as the prior conditional distributions. Second, the property of conditional independence arises in hierarchical models.

Suppose that the likelihood for data $\mathbf{t}$ is $f(\mathbf{t} \mid \theta)$, the prior for θ is $f(\theta \mid \varphi)$ and the hyperprior for φ is $f(\varphi)$. Then φ is conditionally independent of t given θ, and the posterior conditional densities are given by

$$f(\theta \mid \varphi) \propto f(\mathbf{t} \mid \theta) f(\theta \mid \varphi)$$
$$f(\varphi \mid \theta) \propto f(\theta \mid \varphi) f(\varphi) \tag{34}$$

We note that from the Tanner & Wong, 1987 data augmentation method it is simpler and more efficient to sample from a distribution $f(\theta, \varphi \mid \mathbf{t})$ than from $f(\theta \mid \mathbf{t})$. The augmentation parameter, also called the auxiliary variable φ, can be anything. If we can sample from $f(\theta, \varphi \mid \mathbf{t})$, then the required $f(\theta \mid \mathbf{t})$ is simply a marginal distribution of the augmented distribution, and a sample from $f(\theta \mid \mathbf{t})$ consists of ignoring the φ components of the (θ, φ) sample.

2.4.2 DPLNM Model by Gibbs Sampling

Now from equation (32) conditional posterior density for μ is given by

$$f(\mu_j \mid \mu_{-j}, s^2, \omega, t, x, \theta, \sigma^2, \alpha, \beta, \nu) = \frac{f(\mu_j, \mu_{-j}, s^2, \omega \mid t, x, \theta, \sigma^2, \alpha, \beta, \nu)}{\int f(\mu_j, \mu_{-j}, s^2, \omega \mid t, x, \theta, \sigma^2, \alpha, \beta, \nu) d\mu_j} \tag{35}$$

where μ_{-j} can either be one of the μ_j or could be new values drawn from the prior.

In the Bayesian framework to derive the posterior distribution one important trick is to ignore terms that are constant with respect to the unknown parameters. Thus we note that any factor in the posterior that does not involve μ_j will banish. Therefore the conditional posterior density for μ is

$$f(\mu_j \mid \mu_{-j}, s^2, \omega, t, x, \theta, \sigma^2, \alpha, \beta, \nu) \propto \exp\left\{ -\frac{\frac{1}{2}\sum_{t_i \in t_j} \left(\log(t_i) - \mu_j\right)^2}{s^2} \right\} \times$$

$$\prod_{i,j} \left(1 - \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \left(\mu_j - \theta\right)^2 \right\} \left(\frac{\log(t_i) - \mu_j}{s_j} \right) \right)$$

$$\propto \exp\left\{ \mu_j - \frac{\left(\frac{s_j^2 \theta + \sigma^2 \sum_{t_i \in t_j} \left(\log(t_i)\right)^2}{s_j^2 + \sigma^2 n_j} \right)}{\left(\frac{2\sigma^2 s_j^2}{s_j^2 + \sigma^2 n_j} \right)} \right\} \prod_{i,j} \left(1 - \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ 1 - \frac{1}{2\sigma^2} \left(\mu_j - \theta\right)^2 \right\} \left(\frac{\left(\log(t_i)\right) - \mu_j}{s_j} \right) \right) \tag{36}$$

Then we draw μ_j from

$$Normal\left(\frac{s_j^2 \theta + \sigma^2 \sum_{t_i \in t_j} \left(\log(t_i)\right)^2}{s_j^2 + \sigma^2 n_j}, \frac{\sigma^2 s_j^2}{s_j^2 + \sigma^2 n_j} \right) \tag{37}$$

Once we complete this step for all the n observations, next we update the cluster locations $(\mu_j, s_j^2), j = 1, \ldots, n$, conditional on $\nu, \theta, \sigma^2, \alpha, \beta$ and t.

To update (μ_j, s_j^2), we first update s_j^2 by determining its conditional posterior of given by

$$f(s_j^2 \mid \mu_j, \mu_{-j}, \omega, t, x, \theta, \sigma^2, \alpha, \beta, \nu) = \frac{f(\mu_j, \mu_{-j}, s^2, \omega \mid t, x, \theta, \sigma^2, \alpha, \beta, \nu)}{\int f(\mu_j, \mu_{-j}, s^2, \omega \mid t, x, \theta, \sigma^2, \alpha, \beta, \nu) ds_j^2} \tag{38}$$

and noting that factors not involving s_j^2 banish when the ratio is evaluated. Then a value is drawn from $f(s_j^2 \mid \mu_j, \mu_{-j}, \nu, \theta, \sigma^2, \alpha, \beta, t)$, which is given by

$$f(s_j^2 \mid \mu_j, \mu_{-j}, \omega, t, x, \theta, \sigma^2, \alpha, \beta, \nu) \propto \left(s_j^2\right)^{-\left(\frac{\alpha + n_j}{2}+1\right)} \exp\left\{ - \frac{\beta + \frac{1}{2}\sum_{t_i \in t_j} \left(\log(t_i) - \mu_j\right)^2}{s^2} \right\} \times$$

$$\prod_{i,j} \left(1 - \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ 1 - \frac{1}{2\sigma^2}\left(\mu_j - \theta\right)^2 \right\} \left(\frac{(\log(t_i)) - \mu_j}{s_j} \right) \right) \tag{39}$$

Thus, using Gibbs sampling we draw s_j^2 from

$$Inverse - Gamma\left(\alpha + \frac{n_j}{2}, \beta + \frac{1}{2}\sum_{t_i \in t_j} (\log(t_i))^2 \right) \tag{40}$$

Once we draw the new value of s_j^2, next we draw σ^2 from

$$f(\sigma^2 \mid s_j^2, \mu_j, \mu_{-j}, \omega, t, x, \theta, \sigma^2, \alpha, \beta, \nu) \propto IG\left(\sigma^2 \mid \alpha_\sigma, \beta_\sigma\right) \cdot \prod_{i=1}^{n} f(t_i \mid \theta, \sigma^2)$$

$$\propto \prod_{i,j} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ - \frac{1}{2\sigma^2}\left(\mu_j - \theta\right)^2 \right\} \sigma^{2-(\alpha_\sigma+1)} \exp\left\{ \frac{-\beta_\sigma}{\sigma_j^2} \right\} \right) \tag{41}$$

$$\propto \left(\sigma_j^2\right)^{-\left(\alpha_\sigma + \frac{n}{2}+1\right)} \exp\left\{ - \frac{\beta_\sigma + \sum_{j=1}^{n}\left(\mu_j - \theta\right)^2}{\sigma^2} \right\}$$

From which a value of σ^2 is drawn from

$$Inverse - Gamma\left(\alpha_\sigma + \frac{n}{2}, \beta_\sigma + \sum_{j=1}^{n} (\mu_j - \theta)^2 \right) \tag{42}$$

The prior distribution for β is a Gamma distribution. Thus using the Bayes' Law, we write its posterior distribution as

$$f(\beta \mid s_j^2, \alpha_\beta, \beta_\beta) \propto Gamma\left(\beta \mid \alpha_\beta, \beta_\beta\right) \cdot \prod_{i=1}^{n} IG(s_j^2 \mid \alpha, \beta)$$

$$\propto \prod_{j=1}^{n} \left(\beta^\alpha \exp\left\{ - \frac{\beta}{s_j^2} \right\} \beta^{(\alpha_\beta+1)} \exp\left\{ - \beta\beta_\beta \right\} \right) \tag{43}$$

$$\propto (\beta)^{(\alpha_\beta + 1 + \alpha n)} \exp\left\{ - \left(\beta_\beta + \sum_{j=1}^{n} \left(\frac{1}{s_j^n} \right) \right) \beta \right\}$$

Then we draw a value of β from

$$Gamma\left(\alpha_\beta + \alpha n, \beta_\beta + \sum_{j=1}^{n} \frac{1}{s_j^2}\right) \tag{44}$$

The conditional posterior density for θ is given by

$$\begin{aligned} f(\theta \mid \sigma^2, s_j^2, \mu_j, \mu_{-j}, t, \theta, \sigma^2, \alpha, \beta, \nu) &\propto Normal\left(\theta \mid \mu_\theta, \sigma^2{}_\theta\right) \cdot \prod_{i=1}^{n} f(t_i \mid \theta, \sigma^2) \\ &\propto \prod_{j=1}^{n} \left(\exp\left\{ -\frac{1}{2\sigma^2}\left(\mu_j - \theta\right)^2 \right\} \exp\left\{ \frac{\left(\theta - \mu_\beta\right)^2}{2\sigma_\theta^2} \right\} \right) \\ &\propto \exp\left\{ -\frac{\theta - \dfrac{\sigma^2 \mu_\theta + \sigma_\theta^2 \sum_{j=1}^{n} \mu_j}{\sigma^2 + \sigma_\theta^2 n}}{2\left(\dfrac{\sigma^2 \sigma_\theta^2}{\sigma^2 + \sigma_\theta^2 n}\right)} \right\} \end{aligned} \tag{45}$$

from which we draw θ from

$$Normal\left(\frac{\sigma^2 \mu_\theta + \sigma_\theta^2 \sum_{j=1}^{n} \mu_j}{\sigma^2 + \sigma_\theta^2 n}, \frac{\sigma^2 \sigma_\theta^2}{\sigma^2 + \sigma_\theta^2 n} \right) \tag{46}$$

For ν, we introduce an auxiliary variable u and as in Escobar & West, 1995 and assign a Beta distribution prior to ν. Then we sample ν from a mixed Gamma posterior distribution

$$\begin{aligned} f(u \mid \nu) &\sim Beta(\nu + 1, n) \\ f(\nu \mid u) &= c \cdot Gamma\left(\alpha_\nu + n, \beta_\nu - \log(u)\right) + (1 - c) \cdot Gamma\left(\alpha_\nu + n - 1, \beta_\nu - \log(u)\right) \end{aligned} \tag{47}$$

where

$$c = \frac{\alpha_\nu + n - 1}{n\left(\beta_\nu - \log(u)\right) + \alpha_\nu + n - 1} \tag{48}$$

We draw ν from

$$Beta(1, \nu + 1) \tag{49}$$

Finally, the conditional posterior of the mixing weight ω, $f(\omega \mid \mu_j, \mu_{-j}, s^2, \omega, t, x, \theta, \sigma^2, \alpha, \beta, \nu)$ is drawn from

$$Dir\left(\nu_1 \ldots \nu_n\right) \tag{50}$$

The survival function can then be estimated from the unknown functions

$$\begin{aligned} f(t) &= \int f(t \mid \theta) G(d\theta) \\ F(t) &= \int F(t \mid \theta) G(d\theta) \end{aligned} \tag{51}$$

where $F(t \mid \theta)$ and $f(t \mid \theta)$ are the cumulative distribution function (CDF) and probability density function (PDF) of a parametric kernel with parameter vector θ and G is a prior distribution.

In each iteration of the Gibbs sampling, values for these functions are sampled using the current estimates of $\mu_j, s_j^2, \nu, \theta, \sigma^2, \beta$, which are approximated using finite mixtures with a large number of mixing components so that the survival function can be estimated by

$$S\left(t \mid K, \omega, \mu, s^2\right) = \frac{1}{N} \sum_{i=1}^{K} \omega_i \left(1 - \Phi\left(\frac{\log(t) - \mu_i}{s_i} \right) \right) \tag{52}$$

where Φ is CDF of the standard normal distribution and N are the iterations of the Gibbs sampling. Other survival quantities can be estimated similarly.

III. Results

3.1 Simulated Data

This section a simulation study is undertaken in order to compare the proposed DPLNM model with competing parametric and nonparametric models; and to determine the best fitting probability model for the distribution of survival times. The comparison is based on the lognormal model and the Kaplan Meier (KM) estimator.

Based on the nature of the survival data, a mixture of two Lognormal (LN) (Mclachlan & Peel, 2000) distributions is considered. Singpurwalla, 2006 has shown that this mixture has a long tail which can be controlled by dispersion parameters of each mixture component, and also corresponds to the mixture distribution that represents the probability distribution of observations in the overall population. We however note that the number of components need not be confined to two, but that as indicated by Farcomeni & Nardi, 2010 two is already sufficiently flexible.

Setting a sample of size n=100 with the model

$$0.4LN(4,0.16)+0.6LN(5,0.09) \tag{53}$$

we simulated 10% censoring from the two component mixture with 12 of the sampled data were right censored and the remaining 88 completely observed. We then run the Bayesian MCMC in WinBUGS to analyze these data and investigated the distribution of $f(t \mid \mu_j, s_j^2)$, treating θ, ν, ω, as random parameters in the posterior model.

Since we know very little about the true values of these parameters, we used vague Gamma priors, setting $\nu = 1$ (Marin et al., 2005a) as follows

$$\begin{aligned} \nu &\sim Gamma(1,0.001) \\ \theta &\sim Normal(0,10^6) \\ \sigma^2 &\sim IG(2,0.001) \\ \beta &\sim Gamma(1,0.009976) \end{aligned} \tag{54}$$

These non-informative prior distributions were deployed to generate lifetime data sets resembling the nature of complex models (Kottas, 2006), and each have a variance of 10^6, not to influence the posterior distribution. A large prior variance is indicative of a vague distribution and therefore reflects relative ignorance about the true parameters.

Figure 1 provides plots for the simulated data from mixture of lognormal distributions.

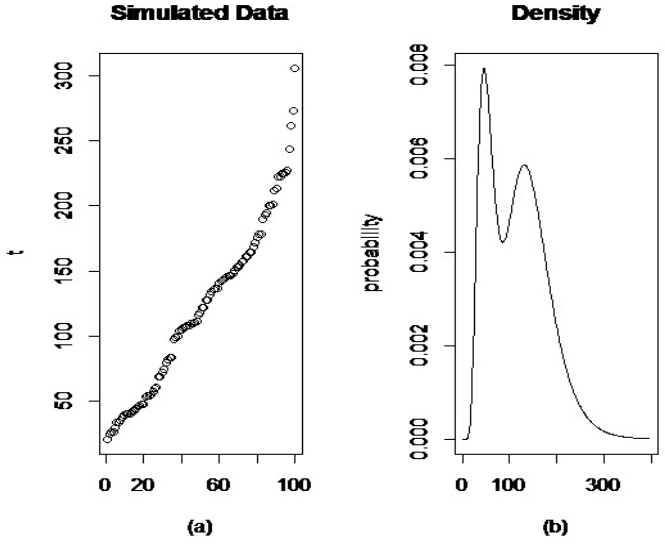

Figure 1: Simulated data from mixture of lognormal distributions

The figure shows that the mixture has a bimodal density, which cannot be captured by the regular lognormal distribution. We carried out a convergence diagnostic test to ensure convergence of the Markov Chains was used before results were taken from them by estimating the length of the burn-in period, before taking a sample from the converged chain. The plot in Figure 2 illustrates the trace history for μ and s^2.

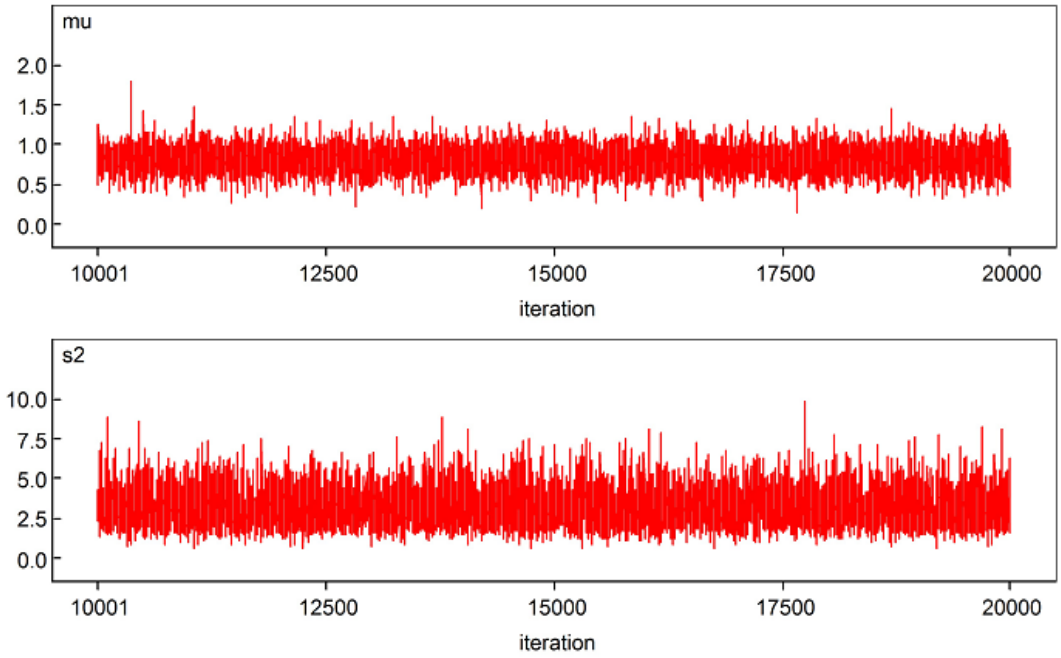

Figure 2: Trace history for μ and s^2.

The figure shows quite a good mixing of the algorithm, with the mixture size moves oscillating without remaining in the same place for too long.

We used the simulated data to illustrate the performance of the DPLNM model. We employed both graphical and quantitative methods to compare the parametric lognormal model, the non-parametric Kaplan Meier (KM) and the proposed model. Graphical comparison was through fitting the survival functions of the three models to the data and a visual inspection as to how similar shape and behavior of the survival functions (curves) are to the true model made. Figure 3 shows the survival curves (plots) obtained.

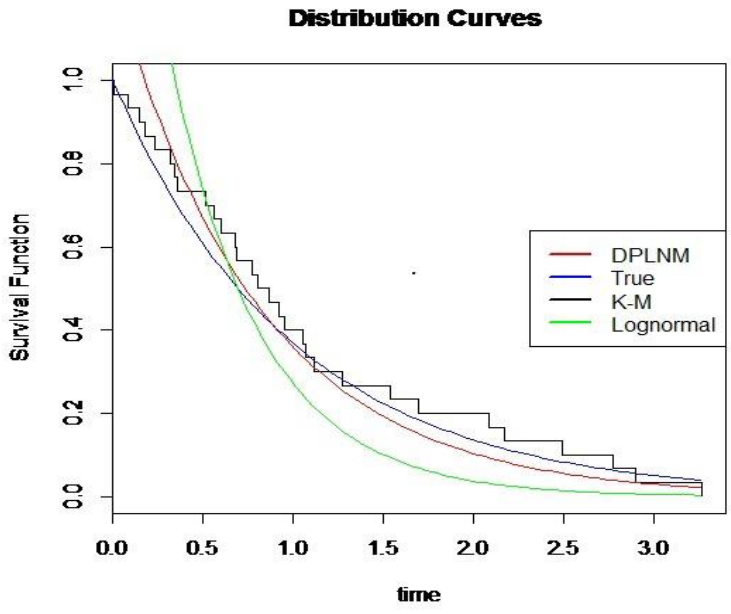

Figure 3: Comparison of lognormal, K-M and DPLNM Functions

From the comparison by observation from the plots, Figure 3 shows that the parametric lognormal is not capable of capturing the generated mixture distribution with long tail and thus is not a good choice for estimating the mixture lifetime. However, the DPLNM model fits the data better than the nonparametric KM. To facilitate a quantitative comparison, the Kolmogorov-Smirnov (KS) test (Silverman, 1986), a nonparametric test for goodness-of-fit (Gupta et al., 2008), was used to assess the appropriateness of the proposed models against the true mixture model. The KS test summarizes the discrepancy between observed values and the values expected under the models in question. Table 1 shows the results from the comparison.

	F(t)		S(t)	
Model	Test Stat	p-value	Test Stat	p-value
Lognormal	0.4785	0.0040	0.5215	0.001
KM	0.2963	0.2560	0.7037	0.008
DPLNMM	0.1476	0.8680	0.8524	0.667

Table 1: Kolmogorov-Smirnov goodness-of-fit test of failure time cumulative density and survival function estimation

The results in Table 1 show that the estimated CDF for the mixture model using DPLNMM has the smallest test statistics value of 0.1476 with a p-value of 0.8680>0.05. A smaller test statistics reflects a better model fit. We conclude that DPLNM model is the best estimate.

3.2 Real Data Problem

Here we analyze data from remission times of 21 pairs of 42 acute leukemia patients (Freireich et al., 1963) in a clinical trial designed to test the ability of 6-Mercaptopurine (6-MP) to prolong the duration of remission. Patients in remission were randomly assigned to maintenance therapy with either 6-MP treatment or a placebo. As in the simulated example, we used the same prior distributions and a Gibbs Sampling MCMC algorithm through Win BUGS with 20000 iterations (10000 to burn-in) to fit the data. In Figure 4 we illustrated and predicted the survivor functions. The Survivor functions have also been compared to the Kaplan Meier estimator where there appears to be a good correspondence between the two for each set of treatment observation.

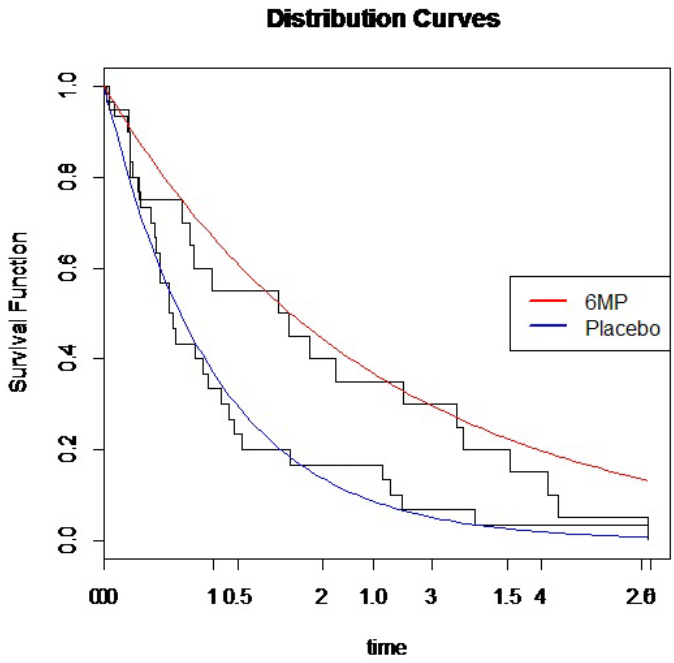

Figure 4: Fitted survival curves and Kaplan Meier estimator for 6MP treatment and Placebo in Leukemia data.

From the figure, we can conclude that patients who receive the 6-MP treatment have a longer survival rate than the patients in the placebo group.

In Table 2 we show a quantitative comparison using Kolmogorov-Smirnov test, a nonparametric test for goodness-of-fit, for testing statistical differences in survival between groups. The null hypothesis states that the

leukemia patient groups have the same survival distribution against the alternative that the survival distributions are different.

$S(t)_{6MP}$		$S(t)_{PLCB}$	
Test Stat	p-value	Test Stat	p-value
0.5946	0.000510	0.4173	0.060

Table 2: Comparison of treatments using Kolmogorov-Smirnov goodness-of-fit test

From these p-values for each test statistic, we conclude, at the 0.05 significance level, that patients who receive the 6-MP treatment have a longer survival rate than the patients in the placebo group. This result supports earlier findings by Freireich et al., 1963.

IV. Conclusions And Further Developments

In this article, we have illustrated how Bayesian methods can be used to fit a mixture of lognormal model with a known number of components to heterogeneous, censored survival data using MCMC algorithm through the Win BUGS software to estimate the survivor function. Some extensions and modifications are possible.

Firstly, this study only involved two candidate models for comparison. More models can be obviously included in the analysis.

Secondly, we have considered a DPLNM model for a heterogeneous population without covariates. 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, the model would be extended to cases where we have unknown number of components K as data grows in complexity.

References

[1]. Escobar, M. and West, M. (1995). Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association, 90:577–588.

[2]. Farcomeni, A. and Nardi, A. (2010). A two-component Weibull mixture to model early and late mortality in a Bayesian framework. Computational Statistics and Data Analysis, 54:416–428.

[3]. Ferguson, T. (1973). A bayesian analysis of some nonparametric problems. Annals of Statistics, 1(2):209–230.

[4]. Freireich, E. J., Gehan, E. A., and Frei, E. (1963). The effect of 6-mercaptopurine on the duration of steroid induced remissions in acute leukemia: A model for evaluation of other potentially useful therapy. Blood, 1:699–716.

[5]. Giudici, P., Givens, G. H., and Mallick, B. K. (2009). Bayesian modeling using WinBUGS. John Wiley and sons, Inc., New Jersey.

[6]. Gupta, A., Mukherjee, B., and Upadhyay, S. K. (2008). Weibull extensionmodel: A bayes study using markov chain monte carlo simulation. Reliability Engineering and System Safety, 93(10):1434–1443.

[7]. Hanson, T. E. (2006). Modeling censored lifetime data using a mixture of gammas baseline. Bayesian Analysis}, 1(3):575–594.

[8]. Ibrahim, J.G., Chen, M-H. & Sinha, D. (2001b). Bayesian survival analysis. New York: Springer.

[9]. Kottas, A. (2006). Nonparametric Bayesian survival analysis using mixtures of weibull distributions. Journal of Statistical Planning and Inference, 136 (3), 578-596.

[10]. Lindley, D. V. (1961). Introduction to probability and statistics from a Bayesian viewpoint: Part 2, Inference. Aberystwyth: University College of Wales.

[11]. Marin, J. M., Mengersen, K. and C. P. Robert. (2005b). Bayesian Modelling and Inference on

[12]. Mixtures of Distributions. Handbook of Statistics 25, D. Dey and C.R. Rao (eds). Elsevier-Sciences.

[13]. McAuliffe, J. D., Blei, D. M., and Jordan, M. (2006). Nonparametric empirical Bayes for the dirichlet process mixture model. Statistical and Computing, 16:5–14.

[14]. Mclachlan, G. and Peel, D. (2000). Finite Mixture Models. John Wiely, New York.

[15]. Nobile, A. and Fearnside, A. (2005). Bayesian mixtures with an unknown number of components: the allocation sampler. Department of Statistics, University of Glasgow. Technical Report 05-4.

[16]. Quiang, J. (1994). A Bayesian Weibull survival model. Unpublished Ph.D. Thesis, Institute of Statistical and Decision Sciences, Duke University: North Corolina.

[17]. Silverman, B. W. (1986). Density estimations for statistics and data analysis. Monographs on Statistics and Applied Probability, London: Chapman and Hall.

[18]. Singpurwalla, N. (2006). Reliability and risk: A Bayesian perspective, Wileys, England.

[19]. Tanner, M. Y. and Wong, W. H. (1987). The calculation of posterior distribution by data augmentation. Journal of the American Statistical Association 67: 702-708.

[20]. Tsionas, E. G. (2002). Bayesian analysis of finite mixtures of Weibull distributions. Commun. Stat. Theor. Math. 31:37–48.

[21]. WinBUGS (2001). WinBUGS User Manual:Version 1.4. UK: MRC Biostatistics Unit [computer program], Cambridge.