SciencePG
Science Publishing Group

On Bayesian Estimation of Dirichlet Process Lognormal Mixture Models and Comparison of Treatments in Censoring

Henry Ondicho Nyambega[1, *], George Otieno Orwa[2]

[1]Department of Mathematics, Kisii University, Kisii, Kenya

[2]Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:
hondicho@kisiiuniversity.ac.ke (H. O. Nyambega), gorwa@jkuat.ac.ke (G. O. Orwa)
[*]Corresponding author

Abstract: One current interest in medical research is the comparison of treatments in the analysis of survival times of patients. This is particularly problematic, especially for censored data, and when these data consists of several groups, where each group has distinct properties and characteristics but belong to the same distribution. There are various modeling schemes that have been contemplated to overcome these complexities inherent in the data. One such possibility is the Bayesian approach which integrates prior knowledge in analysis. In this paper, we focus on the use of Bayesian lognormal mixture model (MLNM) with related Dirichlet process (DP) prior distribution for estimating patient survival. The advances in the Bayesian paradigm have considerably bolstered the development and application of mixture modelling methodology in the field of survival analysis. The proposed MLN model is compared with the conventional parametric lognormal and the nonparametric Kaplan Meier (K-M) models used to estimate survival to establish model robustness. A simulation study that investigates the impact of censoring on these models is also described. Real data from past research is used to show the resulting Dirichlet process mixture model's robustness in the comparison of censored treatment. The results indicate that the proposed lognormal mixtures provide a better fit to complex data. Further, the MLN models are able to estimate various survival distributions and therefore appropriate to compare treatments. Clinicians will find these models useful especially when confronted with the obstacle of choosing a suitable therapy for a disease.

Keywords: Bayesian, Lognormal, Mixture Models, Treatment Comparison, Win BUGS

1. Introduction

Most medical data are censored and/or arise from several homogenous subgroups relating to one or several characteristics, for example when different treatments are administered to patients. There is therefore an increasing need for efficient estimation of patient survival through comparison of treatments. Several modeling procedures have been postulated in literature. One such scheme is the Bayesian framework that incorporates prior information regarding the data without compromising the accuracy of estimates [2]. And as reported by [5], models based on Bayesian nonparametrics present more flexibility leading to ease of computation. Bayesian nonparametric models however suffer from the possible impediments of inference, particularly due to the challenges associated with prior choice. The Dirichlet Process (DP) is the prior of choice for Bayesian non-parametric models. However the DP is inadequate in these settings since the posterior is not DP but a mixture of DP. Mixture DP distributions offer the options of discreteness and flexibility especially since they consider data as represented by weighted sum of distributions, with each distribution characterized by a unique parameter set representing a subspace of the population. Given the

complexities associated with these MDP models, posterior inference is feasible through tractable MCMC technique algorithms. [13], postulated that the Dirichlet Process (DP) prior for mixing portions can be handled by both a Bayesian framework through an MCMC algorithm. [9] use Bayesian techniques, with prior drawn from specific estimation points to analyze survival times data while [1] consider Dirichlet process and its extensions in boistatistics.

The main aim of this paper is to develop and apply the Bayesian Dirichlet process lognormal mixture approach to model patients' survival in cases of censoring. This aim is addressed through the estimation of a two component lognormal mixture model in a simulation study and an application to a leukemia remission dataset from [4]. A Bayesian Markov Chain Monte Carlo (MCMC) approach through Gibbs sampling and Metropolis-Hasting algorithm, assuming known number of mixing components was implemented.

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 MCMC sampling algorithms through the Win BUGS [14] software. The overall aim of MCMC sampling is to simulate from a 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.

The rest of this paper is organized as follows. In Section 2, we define the Dirichlet process mixture model and Bayesian computational approach for parameter estimation. We also provide the method of simulation and discuss the issue of model evaluation. In Section 3, we illustrate the model using simulated datasets and the leukemia dataset. The results are

discussed further in Section 4.

2. Material and Method

2.1. Model Formulation

In this section, we define the Dirichlet process lognormal mixture model for analyzing survival data. Let $t_1,...t_n$ be a random sample taken from a censored and possibly heterogeneous population. The two parameter lognormal distribution given by

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

where $\mu > 0$ is the scale parameter and $s^2 > 0$ is the shape parameter [15].

A DPLN mixture model for t can be written as

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

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

$$x_{ij} = \begin{cases} 1, if \quad ith \quad unit\ is\ drawn\ from\ the\ jth\ mixture\ component \\ 0, \quad elsewhere \end{cases} \quad (3)$$

then $\omega_j = p(x_{ij}) = 1$. The likelihood of a single t_i is given by

$$f(t_i \mid \omega, \mu, s^2) = \sum_{j=1}^{K} \omega_j f(t_j \mid \mu_j, s_j^2), i = 1,...,n \quad (4)$$

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

$$f(\mathbf{t} \mid \omega, \mu, s^2) = \prod_{i=1}^{n} \sum_{j=1}^{K} \omega_j f(t_j \mid \mu_j, s_j^2), i = 1,...,n \quad (5)$$

Thus the joint Likelihood is

$$f(\mathbf{t} \mid x, \omega, \mu, s^2) = \prod_{i=1}^{n} \sum_{j=1}^{K} \left(\omega_j(x_{ij}) f(t_j \mid \mu_j, s_j^2) \right)^{x_{ij}}, i = 1,...,n \quad (6)$$

The joint prior can be expressed as

$$f(\mu, s^2, \omega \mid \theta, \sigma^2, \alpha, \beta, \nu) = f(\omega \mid \nu) f(\mu \mid s^2, \sigma^2) f(s^2 \mid \alpha, \beta) \quad (7)$$

Now according to [10] posterior distribution is calculated through Bayes Theorem as

$$f(\theta \mid t) = \frac{f(t \mid \theta) f(\theta)}{\int f(\theta \mid t) d\theta} \propto f(t \mid \theta) f(\theta) \quad (8)$$

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

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

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

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

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

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

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

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

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

$$\times \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{K}(\mu_j - \theta)^2 \right\} \left[\sqrt{\frac{n}{2\pi}} \frac{\left(\sum_{i=1}^{n}\left(\log(t_i)\right)^2 - \frac{\left(\sum_{i=1}^{n}\left(\log(t_i)\right)^2\right)}{n} \right)^{\frac{n-3}{2}}}{2^{\frac{n-3}{2}} \Gamma\left(\frac{n-3}{2}\right)\left(s^2\right)^{\frac{n}{2}}} \right] \exp\left\{ \frac{\left(\sum_{i=1}^{n}\left(\log(t_i)\right)\right) - \sum_{i=1}^{n}\left(\log(t_i)\right)^2}{2s^2} \right\} \times \qquad (9)$$

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

which is Dirichlet Process mixture model. For each observation t_i, we define an indicator variable as

$$\delta_i = \begin{cases} 1, \text{ if } t_i \text{ is an uncensored failure time,} \\ 0, \text{ if } t_i \text{ is a censoring (right) time} \end{cases} \qquad (10)$$

If t_i is an uncensored failure time, that is, $i = 0$, the full conditional mixture DP model is as given by equation (9). For a rightly censored observation t_i, $i = 1$, then the posterior is given as

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

$$= \frac{\Gamma(\nu_1 + \dots + \nu_K)}{\Gamma(\nu_1)\dots\Gamma(\nu_K)} \omega^{\nu_1 - 1} \dots \omega^{\nu_K - i} \left[\prod_{j=1}^{K} \frac{1}{st\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2}(\mu_j - \theta)^2 \right\} \right] \times \quad (11)$$

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

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

$$\exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{K}(\mu_j - \theta)^2 \right\} \left[\sqrt{\frac{n}{2\pi}} \frac{\sum_{i=1}^{n}(\log(t_i))^2 - \dfrac{\left(\sum_{i=1}^{n}(\log(t_i))^2\right)^{\frac{n-3}{2}}}{n}}{2^{\frac{n-3}{2}}\Gamma\left(\frac{n-3}{2}\right)(s^2)^{\frac{n}{2}}} \right] (s_j^2)^{-((\alpha+n_j)/2+1)} \exp\left\{ \frac{\beta + 0.5\left(\sum_{i,j}(\log(t_i)) - \mu\right)^2}{s^2} \right\} \times \quad (12)$$

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

where n_j are the number of uncensored failure times in the jth cluster.

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 $f(t_j \mid \mu_j, s_j^2)$, defined by component specific parameters, μ_j, s_j^2. Since the components come from the lognormal parametric family, $f(t_j \mid \mu_j, s_j^2)$, then the parametric mixture model is

$$\sum_{j=1}^{K} \omega_j f(t_j \mid \mu_j, s^2) \quad (13)$$

In the Bayesian analysis of the model, we assume that $\omega \mid K \sim Dir(\nu_1, \dots, \nu_n)$, where $\nu_i; i = 1, \dots, n$ are fixed constants. Also, the component parameters μ_j, s_j^2 are assumed a priori independent, conditionally on K and, possibly, a vector of hyperparameters, $\alpha, \beta, \theta, \sigma^2$ so that the posterior is

$$f(\mu, s^2, \omega \mid K, \alpha, \beta, \theta, \sigma^2) = \prod_{j=1}^{K} f(\mu_j, s_j^2 \mid \alpha_j, \beta_j, \theta_j, \sigma_j^2) \quad (14)$$

If prior distributions are specified, such that

$$f(\mu, s^2, \omega \mid \theta, \sigma^2, \alpha, \beta, \nu) = f(\omega \mid \nu) f(\mu \mid s^s, \sigma^2) f(s^2 \mid \alpha, \beta) \quad (15)$$

then a sample from the joint posterior of can be obtained by means of Markov chain Monte Carlo methods [3]. In the Bayesian framework, a DP prior is assigned to the mixture model with a kernel distribution, to form a mixture DP model (MDPM) written as

$$F(t, G) = \int f(t \mid \mu, s^2) G(d\mu, s^2) \quad (16)$$

If we set G as a DP prior, then $G \sim DP(\nu, G_0)$ denotes 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 which makes the MDPM a nonparametric method and $\nu > 0$ is the mass parameter.

For this lognormal distribution mixture we model flexibility by conveniently assigning the following independent prior distributions [8] for the unknown parameters, and accordingly, express the MDPLN model hierarchically as

$$t_j \mid \mu_j, s_j^2 \sim f(t_j \mid \mu_j, s_j^2)$$
$$(\mu_j, s_j^2) \mid G \sim G$$
$$G \mid \nu, \beta, \theta, \sigma^2 \sim DP(\nu G_0)$$
$$\nu \sim Gamma(\alpha_\nu, \beta_\nu)$$
$$G_0 \sim inverse-Gamma(s^2 \mid \alpha, \beta) \cdot Normal(\mu \mid \theta, \sigma^2) \quad (17)$$
$$\theta \sim Normal(\mu_\theta, \sigma_\theta^2)$$
$$\sigma^2 \sim Inverse-Gamma(\alpha_\sigma, \beta_\alpha)$$
$$\beta \sim Gamma(\alpha_\beta, \beta_\beta)$$

In WinBUGS [12], censored data are modelled using a missing data approach through the command I(.,) as follows

$$t[i] \sim d\ln orm(mu[i]; s^2[i]) I(ces.time[i];) \quad (18)$$

where cens:time [i] is either zero for uncensored outcome or the ith recorded survival time for censored outcomes. Hence, censored survival times are considered drawn from a truncated lognormal distribution. We place the prior distributions on the parameters and hyper parameters as shown in equation (16), and choose small positive values for $\alpha_\nu, \alpha_\beta, \alpha_\sigma, \beta_\nu, \beta_\beta, \beta_\sigma, \mu_\theta, \sigma_\theta^2$ to express vague prior knowledge about these parameters and setting $\nu = 1$ [3].

2.2. Computation Method

The model described in Section 2.1 can be fitted using MCMC sampling with latent values Z_i to indicate component membership of the ith observation. Since

$$\omega_j = P(Z_i = j) \quad (19)$$

We can write

$$Z_i \sim Mult(\omega_1, \cdots, \omega_K) \quad (20)$$

In this approach, Z_i is sampled by computing posterior probabilities of membership, and the other parameters are sampled from their full posterior distributions, conditional on the latent indicators. This was implemented in the WinBUGS software package. The algorithm is set by selecting the permutation of components at each iteration minimizing the vector dot product, from the posterior distribution. The MCMC output was then aligned according to each picked permutation.

2.3. Simulation Algorithm

This section describes the sampling algorithm [6] for simulating from the complex posterior distribution. This is implemented by creating a Markov chain with the posterior as its stationary distribution through direct successive simulations from the component conditional distributions. We note that the Gibbs sampler and its various adaptations has been the most commonly used approach in Bayesian mixture estimation.

Our interest in this study was to estimate the parameter of Bayesian Dirichlet process lognormal mixture. We used the simulation algorithm for analyses. For this study, data were simulated from two component lognormal mixture models with the following parameter configurations:

$$f(t \mid K, \omega, \mu, s^2) = 0.4LN(4, 0.16) + 0.6LN(5, 0.09) \quad (21)$$

The censoring levels of 30% was applied to model and a sample size of N = 100 for n=100 individuals was used for all experiments. The following steps were applied to carry out the simulations.

1. Generate t_i from the respective model, for i=1, 2,…, n.
2. Generate censoring times by assuming that the largest C% survival times are right censored.
3. Fit the model based on the data, with 10000 iterations, 5000 burn-in.
4. Record posterior estimates of the model parameters, namely median and standard deviation.

3. Results

3.1. Simulated Data

Based on the nature of the survival data, a mixture of two Lognormal [13] distributions is considered. We generated N = 100 data sets, each with n = 100 individuals. We generated a right-censored data set by first generate a variable T for true survival time from the mixture distribution. Next, we generated another variable for censoring time, C from a uniform distribution on the interval (0, 4.784) so that each individual has approximately a 30% chance of having a right-censored survival time. This was to obtain approximately 30% censoring for the data set.

Since we know very little about the true values of these parameters, we used vague Gamma priors [3] as follows

$$\nu \sim Gamma(1, 0.001)$$
$$\theta \sim Normal(0, 10^6)$$
$$\sigma^2 \sim IG(2, 0.001) \quad (22)$$
$$\beta \sim Gamma(1, 0.009976)$$

These non-informative prior distributions were deployed to generate lifetime data sets resembling the nature of complex models [13], 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.

We carried out a convergence diagnostic test to ensure convergence of the Markov Chains by estimating the length of the burn-in period, before taking a sample from the converged chain. The plot in Figure 1 illustrates the trace history for μ and s^2 while Table 1 confirms the accuracy of the parameter estimates in the 30% of censoring.
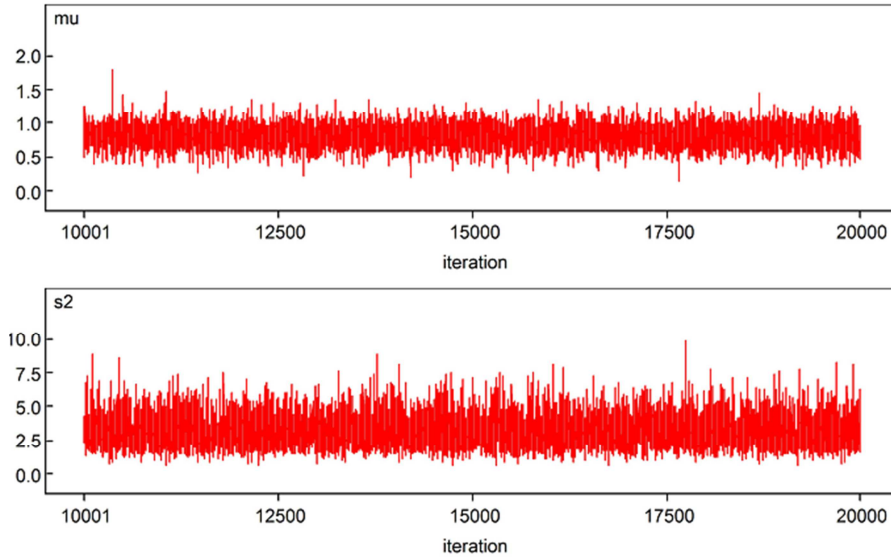
Figure 1. *Trace history for μ and s^2.*

Table 1. *Posterior estimates of the model parameters from 10000 iterations for a burn-in of 5000 with 30% censoring.*

Parameter	Posterior Mean	Posterior SD	95% CI
μ	0.908	0.1496	(0.872, 1.109)
s^2	2.718	1.095	(2.677, 3.255)
θ	0.8037	0.1168	(0.7146, 0.9927)
σ^2	2.935	1.1768	(2.865, 3.376)
β	2.542	0.3801	(2.498, 2.812)

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. The simulated data was used to illustrate the performance of the proposed MDPLN model with competing models in literature. We employed both graphical and quantitative methods to compare the parametric lognormal model, the non-parametric Kaplan Meier (K-M) and the proposed MDPL 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 (Equation 21). Figure 2 shows the survival curves (plots) obtained.

From the plots in Figure 2 we see 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 survival time data. However, the MDPLN model fits the data better than both parametric lognormal and the nonparametric K-M. To facilitate a quantitative comparison, the Kolmogorov-Smirnov (KS) test, a nonparametric test for goodness-of-fit [7], 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 2 shows the results from the comparison.
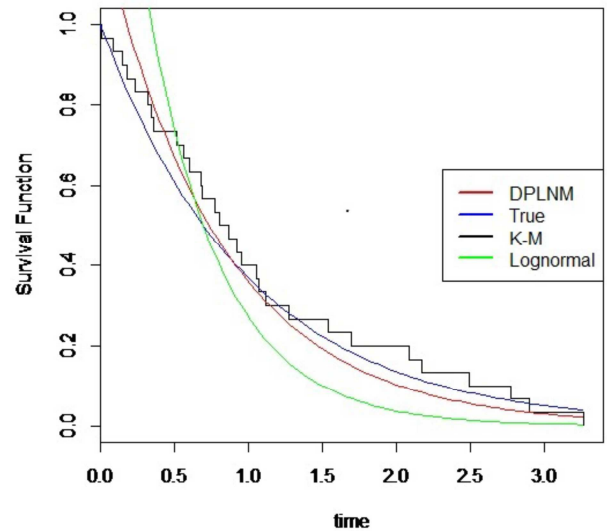

Figure 2. *Comparison of survival functions for lognormal, K-M and MDPLNM Functions.*

Table 2. *Kolmogorov-Smirnov goodness-of-fit test of failure time cumulative density and survival function estimation.*

Model	F(t)		S(t)	
	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

The results in Table 2 show that the estimated CDF for the mixture model using MDPLNM 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 MDPLN model offers the best estimate and therefore most appropriate to estimate survival.

3.2. Application

Here we analyzed data from remission times of 21 pairs of 42 acute leukemia patients [4] in a clinical trial designed to test the potency of 6-Mercaptopurine (6-MP) to lengthen remission in patients randomly assigned to maintenance therapy of either 6-MP or a placebo. As in the simulated example, we used the same prior distributions and the Sampling MCMC algorithm through Win BUGS with 10000 iterations (5000 to burn-in) to fit the data. In Figure 4 we illustrate and predict the survivor functions. The Survivor functions have also been compared to the Kaplan Meier estimator. From the figure there appears to be a good correspondence between the K-M and the plots for each set of treatment observation.
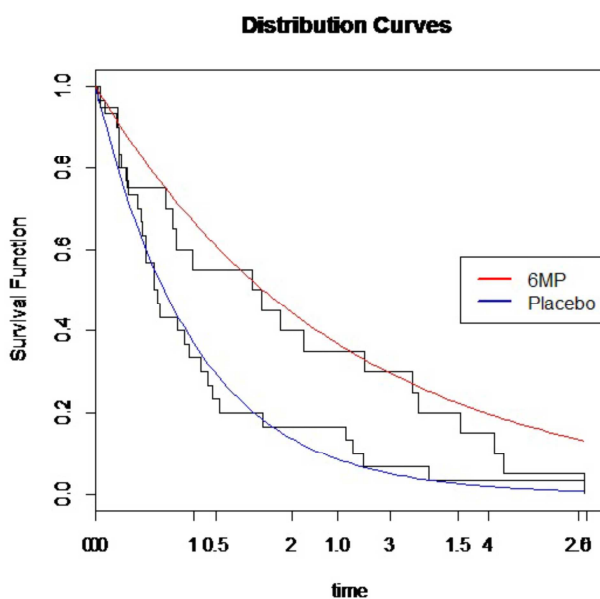

Figure 3. *Fitted survival curves and Kaplan Meier estimator for 6MP treatment and Placebo in Leukemia data.*

From figure 3, we conclude that patients who receive the 6-MP treatment have a longer survival rate than the patients in the placebo group. This justifies the effectiveness of 6-MP treatment in prolonging duration of remission for the patients.

In Table 3 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.

Table 3. *Comparison of treatments using Kolmogorov-Smirnov goodness-of-fit test.*

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

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 [4].

4. Conclusions and Further Developments

In this paper, 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. We show from simulation results and those from real data that the mixture Dirichlet Process LNM model is capable of capturing different shapes of complex survival time distributions. This indicates that the proposed model is more robust than the competing parametric lognormal and nonparametric Kaplan Meier models, and therefore shows promising potential to be applied to compare different treatments. Some extensions and modifications of the model would be extended to cases where we have unknown number of components K as data grows in complexity.

References

[1] Dunson, D. B. (2009). Nonparametric Bayes applications in biostatistics. In: Hjort N, Holmes C, Müller P, Walker S (eds) Bayesian nonparametric in practice. CUP.

[2] Eves, W. X. and Walker, S. G. (2015). Bayesian Information for sensors. Quality of reliability engineering international, 31 (8): 1717-1724. Computational Statistics and Data Analysis, 54: 416–428.

[3] 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.

[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] Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. Nature, 521: 452-459.

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

[7] 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.

[8] Hanley, W. and Millea, M. (2018). Maximum entropy priors with derived parameters in a specified distribution. arXiv preprint arXiv: 1804.08143.

[9] Kulasekera, K. B. and Zhao, M. (2008). A pointwise bayes-estimator of the survival probability with censored data. Stat Probab Lett 3, 78: 2597–2603.

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

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

[12] Ntzoufras, I. (2009). Bayesian modeling using WinBUGS. John Wiley and sons, Inc., New Jersey.

[13] Samwel Gershman,, J. and Blei David, M. (2012). A tutorial on Bayesian non-parametric model. Journal of Mathematical psychology, 56: 1-12.

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

[15] Yee Whye The. (2010). Dirichlet processes. In encyclopedia of machine learning, pages 280-287.