Prior Specification in Bayesian Estimation when Fitting Dirichlet Process Lognormal Mixture Models

Henry Ondicho Nyambega[1], Dr. George O. Orwa[2], Dr. Joseph K. Mung'atu[3] & 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

In this paper Prior specification on Bayesian inference when fitting a Dirichlet process mixture model with a lognormal kernel (DPLNMM) in the presence of censoring is considered. We study the effects of prior choices on posterior inference by varying prior dispersion. Simulation and an application to leukemia data study was carried out with different priors. The Bayesian approach via the Gibbs Sampler Markov Chain Monte Carlo algorithm designed to fit the model was used in analysis. We also show that the DPLNM model is relatively robust to the form of prior used from the MCMC output.

Keywords: Lognormal, Prior, Dirichlet Mixture models, MCMC

1. Introduction

Bayesian framework has been considered as attractive methods in survival modelling (Hjort, 2003 and Muller & Quintana, 2004). The 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 traditional likelihood, along with a given set of survival data to obtain the posterior distribution of the parameter of interest on which the statistical inference is based (Singpurwalla, 2006. The prior information could come from past data or from previous comparable experiments (Gelman et al., 2004). Then, by using appropriate MCMC sampling technique, random samples from the posterior distribution can be generated and with these samples, inferences on model parameters and their functions made.

For Dirichlet process mixture models, Doss and Huffer, 2003 suggest using Markov chain Monte Carlo (MCMC) simulation techniques to approximate the posterior distribution. In particular, an MCMC method called Gibbs sampling is used to generate random samples from the complex posterior distribution through direct successive simulations from the component conditional distributions.

The Bayesian methods have thus gained popularity because of the incorporation of external information about the parameters of interest into the inference process, thus greatly improving inference. Further, the use of Bayesian methods is easily extended to more complex models, allowing one to not only express uncertainty in the many parameters involved, but also relationship between them. Bayesian data analysis is also valuable in survival analysis in cases of censoring, where one usually obtains limited information directly from the data, and in cases where the observed data consist of several groups, where each group has different properties and characteristics of the one family but uses the same distribution (Singpurwalla, 2006).

One important issue in Bayesian estimation when fitting mixture models is prior specification. In this paper, we provide a simple recommendation for the choice of these priors through prior

sensitivity. A prior distribution of a parameter is the probability distribution that represents the parameter's uncertainty before data are observed, and therefore priors are subjective. Within the Bayesian paradigm, objective results are obtained by using prior distributions that have a minimal impact on the posterior distribution. Additionally, Congdon, 2010 argue that in the Bayesian approach hierarchical models can be introduced in the analysis adding more flexibility in the model and some improvements in the analysis. Hierarchical models are those in which one or more parameters of the model are in turn dependent on a series of other parameters (called hyper-parameters) based on certain probability density functions (hyper-priors). In this case, hyper-parameters follow a particular prior distribution too. So that different levels of hierarchy can be set up in the analysis. A variety of different prior distributions have been proposed in literature to model the survival quantities. A good overview of prior distributions can be found in Ferguson, 1973; Doksum, 1974; Mazzuchi & Singpurwalla, 1985; Hjort, 1990 and Arjas & Gasbarra, 1994.

As long as we maintain positivity, continuity, and consistency, there are essentially three alternative approaches to specifying a prior. The first is to fully specify a subjective prior, the second is to choose a convenient family of conditionally conjugate priors, and the third is to use some default objective prior.

The first technique is considered ideal. Though these priors often present computing challenges, are difficult to elicit, and are elicited for each model, the method provides a full probability analysis which represents the researcher's beliefs. The second method offers a good approximation to a full probability analysis while presenting fewer challenges in computation and elicitation. Finally, the third option is advantageous when there are many models to consider or there is only weak prior knowledge. In these situations, since the full probability model is not specified one may define priors that correspond to integrable (proper) posteriors for some minimal sample size. Despite these shortcomings, Berger, 2006 suggest that these priors often provide unbiased posterior distributions, and are preferred because they have weak prior knowledge.

A well accepted criterion for the choice of a DPM model prior is that the prior has a large or full topological support. Intuitively, such a prior can reach every corner of the parameter space and thus can be expected to have consistent posterior. More flexible models have higher complexity and hence the process of prior elicitation becomes more complex.

Priors are usually constructed from the consideration of mathematical tractability, feasibility of computation, and good large sample behavior. The form of the prior is chosen according to some default mechanism while the key hyper-parameters are chosen to reflect any prior beliefs. Raiffa & Schlaifer, 1961 summarize these characteristics as follows

1. The class should be analytically tractable. Therefore, the posterior distribution should be easily computed, either analytically or through simulation.

2. The class should be rich, in the sense of having a large enough support.

3. The hyperparameters defining the prior should be easily interpreted.

Although it is not always possible to completely satisfy all of these requirements, in this paper, we emphasizes the importance of these features when constructing priors on spaces of distributions.

The objective of this paper is to study the influence of prior specifications for ν, in hierarchical finite DPLNM models considering Gamma priors with varying dispersion on the estimation of the model parameters and credible intervals of estimates. This aim is addressed through the estimation of a two component lognormal mixture model in a simulation study and an application to a leukemia dataset.

The rest of this paper is organized as follows. Section 2 motivates the modelling approach and presents the mixture of lognormal model. Section 3 considers a strategy for posterior inference using a Gibbs sampling MCMC algorithm through WinBUGS software and discuss prior specification. In Section 4, we demonstrate the performance of the model, using simulated data generated from a mixture of two lognormal distributions and real dataset. Finally, in Section 5 we summarize our results and outline areas of future research.

2. The Dirichlet Process Lognormal Mixture Model

In this section, we define the lognormal mixture model for analyzing survival data in the presence of censoring. We also discuss Gibbs Sampling MCMC algorithm through the WinBUGS (WinBUGS, 2001)software.

Let T be a non negative random variable representing patient's survival time and t be a realization of the random variable T. Assume that we observe survival time t on patients possibly from a heterogeneous population. The two parameter lognormal density function for survival time is given by

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

where $\mu > 0$ a scale parameter which equals the logarithm of the median time to failure and $s^2 > 0$ is a shape parameter (Ibrahim et al., 2001b).

A mixture of K lognormal densities (Marin et al., 2005a) is then defined by

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

where ω_j are mixing weights satisfying, $\omega_j > 0$ with $\sum_j^K \omega_j = 1$ and $f(t|\mu_j, s_j^2)$, $j = 1, \cdots, K$ is a kernel density of the lognormal distribution. Mixing on both the shape and scale parameters of the lognormal kernel results in a flexible mixture that can model a wide range of distributional shapes.

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

$$x_{ij} = \begin{cases} 1 & \text{if ith unit is drawn from the jth mixture component,} \\ 0 & \text{elsewhere.} \end{cases} \tag{3}$$

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

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

$$f(t_i \mid \omega, \theta) = \sum_{j=1}^{K} \omega_j f(t_i \mid \theta_j), \quad i = 1, \cdots, n \tag{4}$$

and for a vector of observations $\boldsymbol{t} = (t_1, \ldots, t_n)$,

$$f(\mathbf{t_i} \mid \omega, \theta) = \prod_{i=1}^{n} \sum_{j=1}^{K} \omega_j f(t_i \mid \theta_j), \quad i = 1, \cdots, n \tag{5}$$

Thus the joint Likelihood becomes

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

We place the following prior distributions on the parameters to give the DPLNM model hierarchically as

$$\begin{aligned}
&t_j \mid \mu_j, s_j^2 \sim f(t \mid \mu_j, s_j^2) \\
&(\mu, s^2) \mid G \sim G \\
&G \mid \nu, \beta, \theta, \sigma^2 \sim \mathrm{DP}(\nu G_0) \\
&\nu \sim \mathrm{Gamma}(\alpha_\nu, \beta_\nu) \\
&G_0 \sim \text{Inverse-Gamma}(s^2 \mid \alpha, \beta) \cdot \mathrm{Normal}(\mu \mid \theta, \sigma^2) \\
&\theta \sim \mathrm{Normal}(\mu_\theta, \sigma_\theta^2) \\
&\sigma^2 \sim \text{Inverse-Gamma}(\alpha_\sigma, \beta_\sigma) \\
&\beta \sim \mathrm{Gamma}(\alpha_\beta, \beta_\beta)
\end{aligned} \tag{7}$$

where G_0 is a parametric distribution function, the center or base distribution of the process, ν is a positive scalar precision parameter and $G \sim DP(\nu G_0)$ denotes that a DP prior is placed on the distribution function G.

The parameter ν of the DP prior $\mathrm{DP}(\nu G_0)$ controls how close a realization of the process is to the base distribution G_0. The larger the value of ν the closer a realization of the process is to G_0 (Ferguson, 1973). In the DP mixture model (7), ν controls the number of distinct components of the mixture (Antoniak, 1974, and Escobar & West, 1995) and therefore, prior information about the number of components can be incorporated through the prior for ν. Since we do not have strong prior information in this direction, we choose values roughly equal to the sample median and interquartile range as prior guesses for the population median and interquartile range, respectively. That is, we choose α_ν and β_ν yielding Gamma priors for ν that place mass both on small and large values.

As suggested in Kottas, 2006 there is posterior learning for ν when sample sizes are moderate to large (e.g., $n > 50$), while with small sample sizes, it appears to be difficult for the data to

inform about ν. The base distribution G_0 of a DP mixture model can be considered as our prior guess on μ and s^2 (Antoniak, 1974).

Therefore, a convenient choice of G_0 for the DPLNM model would be

$$G_0(\mu, s^2 \mid \theta, \sigma^2, \beta) = \text{Normal}(\mu \mid \beta, \sigma^2) \cdot \text{inverse-Gamma}(s^2 \mid \alpha, \beta) \tag{8}$$

where the inverse-Gamma distribution has the probability density function is

$$f(s^2 \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\left(s^2\right)^{-(\alpha+1)} \exp\left(-\frac{\beta}{s^2}\right) \tag{9}$$

so that prior to posterior analysis is efficient, the model is flexible and prior information can be incorporated through its parameters. This enables the Gibbs sampler (West et al., 1994, Bush & MacEachern, 1996) to efficiently and easier implemented to fit the model.

The joint prior can then be expressed as

$$f(\mu, s^2, \omega \mid \theta, \sigma^2, \alpha, \beta, \nu) = f(\omega \mid \nu) f(\mu \mid s^2, \sigma^2) f(s^2 \mid \alpha, \beta) \tag{10}$$

For each observation t_i, we define an indicator variable as

$$\delta_i = \begin{cases} 0 & \text{if } t_i \text{ is an uncensored failure time,} \\ 1 & \text{if } t_i \text{ is a censoring (right) time.} \end{cases} \tag{11}$$

Then for a rightly censored observation t_i, $i = 1$, and the full conditional posterior distribution of the model is given as

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

$$\exp\left\{-\beta \sum_{j=1}^{k} \frac{1}{s_j^2}\right\} \exp\left\{-\frac{1}{2\sigma^2} \sum_{j=1}^{k} (\mu_j - \theta)^2\right\} \left[\sqrt{\frac{n}{2\pi}} \frac{\left(\sum_{i=1}^{n} \left(\log t_i\right)^2 - \frac{\left(\sum_{i=1}^{n} \log t_i\right)}{n}\right)^{\frac{n-3}{2}}}{2^{\frac{n-3}{2}} \Gamma\left(\frac{n-3}{2}\right) \left(s^2\right)^{\frac{n}{2}}}\right] \times \tag{12}$$

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

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

3. Posterior Inference

To apply model (7), we choose values for the parameters of the priors for ν, β, θ and σ^2. Choosing the base distribution given by (8) offers both the computational convenience of using Gibbs Sampling and the flexibility of incorporating prior information. This is achieved if we

assume θ and σ^2 to be random and set $\alpha = 2$ yielding an inverse Gamma distribution with infinite variance, with all the parameters of the priors for ν, β, θ and σ^2 fixed. Flexibility is added by taking β random. We specify prior distributions for the hyper-parameters θ, σ^2, and β, especially, we assume Normal$(\mu_\theta, \sigma^2_\theta)$, inverse-Gamma$(\alpha_\sigma, \beta_\sigma)$ and Gamma$(\alpha_\beta, \beta_\beta)$ prior distributions to θ, σ^2, and β, respectively. Finally, we place a Gamma prior on ν which also facilitates the implementation of the Gibbs sampler (Escobar & West, 1995).

We run WinBUGS using a missing data approach through the $I(\cdot,)$ command (Ntzoufras, 2009)

$$t[i] \sim dlnorm(mu[i], s2[i])I(cens.time[i],) \tag{13}$$

where $cens.time[i]$ is either zero for uncensored outcome or the i^{th} recorded survival time for censored outcomes. Hence, censored survival times are assumed to be drawn from a truncated lognormal distribution.

We choose small positive values for $\alpha_\sigma, \beta_\sigma, \alpha_\beta, \beta_\beta$ to express vague prior knowledge about these parameters setting $\nu = 1$ (Marin et al., 2005a). We carried out posterior inference, using a Gibbs sampling scheme as suggested by Diebolt & Robert, 1994 and Robert & Casella, 2000, by the introducing of an indicator variables $Z_i, i = 1, \cdots, n$, which define from which element of the mixture the i^{th} observation has been generated. Thus,

$$p(Z_i = j \mid K, \boldsymbol{\omega} = \omega_j) \tag{14}$$

As in Lunn et al., 2000 and Spiegelhalter et al., 2002, here, we sample the Z_i's by computing posterior probabilities of membership, and the other parameters are sampled from their full posterior distributions, conditional on the latent indicators through the WinBUGS software package. Thus,

$$\begin{aligned}
f(\mu_j \mid \mu_{-j}, s^2, \omega, t, x, \theta, \sigma^2, \alpha, \beta, \nu) &\sim \text{Normal}\left(\frac{s_j^2\theta + \sigma^2 \sum\limits_{t_i \in t_j} \log(t_i)^2}{s_j^2 + \sigma^2 n_j}, \frac{\sigma^2 s_j^2}{s_j^2 + \sigma^2 n_j}\right) \\
f(s_j^2 \mid \mu_j, \mu_{-j}, \nu, \theta, \sigma^2, \alpha, \beta, t) &\sim \text{inverse-Gamma}\left(\alpha + \frac{n_j}{2}, \beta + \frac{1}{2}\sum_{t_i \in t_j}(\log(t_i - \mu_j)^2\right) \\
f(\sigma^2 \mid s_j^2, \mu_j, \mu_{-j}, \nu, \theta, \alpha, \beta, t) &\sim \text{Inverse-Gamma}\left(\alpha_\sigma + \frac{n}{2}, \beta_\sigma + \sum_{j=1}^{n}(\mu_j - \theta)^2\right) \\
f(\beta \mid \alpha_\beta, \beta_\beta, s^2) &\sim \text{Gamma}\left(\alpha_\beta + \alpha n, \beta_\beta + \sum_{j=1}^{n}\frac{1}{s_j^2}\right) \\
f(\theta \mid \sigma^2, s_j^2, \mu_j, \mu_{-j}, \nu, \theta, \alpha, \beta, t) &\sim \text{Normal}\left(\frac{\sigma^2\mu_\theta + \sigma_\theta^2 \sum\limits_{j=1}^{n} \mu_j}{\sigma^2 + \sigma_\theta^2 n}, \frac{\sigma_\theta^2\sigma^2}{\sigma^2 + \sigma_\theta^2 n}\right)
\end{aligned} \tag{15}$$

For ν, an auxiliary variable u is introduced and a Beta distribution prior is assigned to ν (Escobar & West, 1995). Then ν sampled from a mixed Gamma posterior distribution

$$f(u \mid \nu) \sim \text{Beta}(\nu + 1, n)$$
$$f(\nu \mid u) = c\text{Gamma}(\alpha_\nu + n, \beta_\nu - \log(u)) + (1 - c)Gamma(\alpha_\nu + n - 1, \beta_\nu - \log(u)), \tag{16}$$

where

$$c = \frac{\alpha_\nu + n - 1}{n(\beta_\nu - \log(u)) + \alpha_\nu + n - 1} \tag{17}$$

Finally, the conditional posterior of the mixing weight ω, $f(\omega|\mu_j, s_j^2, \mu_{-j}, \theta, \alpha, \sigma^2, \nu, \beta, t)$ is drawn by first drawing z_j independently from

$$\text{Beta}\,(1, \nu + n) \tag{18}$$

Kottas, 2006 introduces a procedure is based on the constructive definition of the Dirichlet process which computes $\omega_j = z_j \prod_{i=1}^{j-1} z_i$.

4. Results

4.1 Simulated Data

Based on the nature of the survival data, a mixture of two Lognormal (LN) distributions is considered. This mixture has a long tail (Singpurwalla, 2006) 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. This true mixture model (Mclachlan & Peel, 2000) is given as

$$0.4\text{LN}(4, 0.16) + 0.6\text{LN}(5, 0.09) \tag{19}$$

Initially, we simulated a sample of size $n = 100$ from the two component mixture with 10% of the sampled data being right censored and the remaining 88% completely observed.

We run the Bayesian MCMC in WinBUGS for 10000 observations (5000 to burn-in) and investigated the distribution of $f(t|\mu_j, s_j^2)$, treating θ, σ^2 and β as random parameters in the DPLNM model. Since very little is known about the true values of these parameters, we used vague Gamma priors, the non-informative distributions, as discussed in section 3, so as to generate survival data sets resembling mixture models (Kottas, 2006), as follows

$$\begin{aligned} \nu &\sim \text{Gamma}(1, 0.1) \\ \theta &\sim \text{Normal}(0, 10^6) \\ \sigma^2 &\sim IG(2, 0.001) \\ \beta &\sim \text{Gamma}(1, 0.009976) \end{aligned} \tag{20}$$

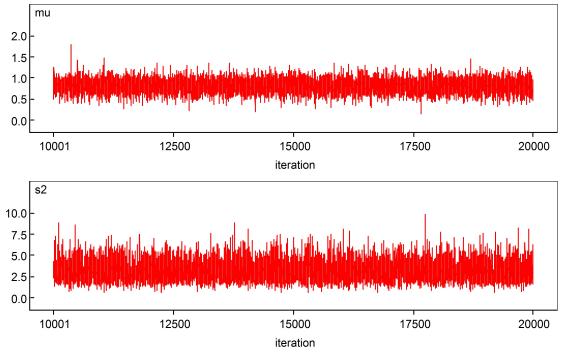

Figure 1. Trace History for μ and s^2.

These priors each have a variance of 10^6, so should not influence the posterior distribution much. A large prior variance is indicative of a vague distribution and therefore reflects our relative ignorance about the true parameters.

We initialized the model, running the chains to check for convergence. Figure 1 shows the trace history of the value of μ and s^2 from a burn-in period of 5000 at 10,000 iterations in the DPLNM model. The plot illustrates that the mixing of the algorithm seems to be quite good, while Convergence of the MCMC algorithm, assessed through multiple chains, was fast. Mixing of the chains was also satisfactory considering the large number of latent variables involved. This is a good indication that the chains have reached the equilibrium distribution. In Figure 2 we illustrates the posterior distributions of these variables, and in Table 1 we show the summary statistics results.

Table 1. Posterior estimates of the DPLNM model parameters from 10000 iterations after a burn-in of 5000 for Gamma(1, 0.1).

Parameter	Posterior Mean	Posterior Std Dev	95% Bayesian CI
μ	0.808	0.149	(0.67, 1.11)
s^2	3.17	1.09	(2.88, 5.66)
θ	0.204	0.117	(0.115, 0.493)
σ^2	2.94	1.177	(3.07, 5.47)
β	1.18	0.380	(0.999, 1.61)

We can see that these values that the parameters posterior means for lies comfortably within the 95% Bayesian CI. Within each data set the point estimates of the parameters are broadly similar. However, the credible intervals vary and potentially could lead to different inferences.

We can see in Figure 2 that the posterior densities of the parameters do look like Gamma distributions.

4.2 Prior Sensitivity

In this section we investigate the sensitivity of the DPLNM model to the form of prior used. An important criticism of the Bayesian method is the subjectivity inherent in the choice of prior (Bernardo & Smith, 1994). Bayesian modeling requires the specification of priors for model parameters.

We consider three priors for ν, Gamma$(2, 0.9)$, Gamma$(2, 0.1)$ and Gamma$(3, 0.05)$ distributions, yielding increasing values for the prior mean and variance of ν. These distributions

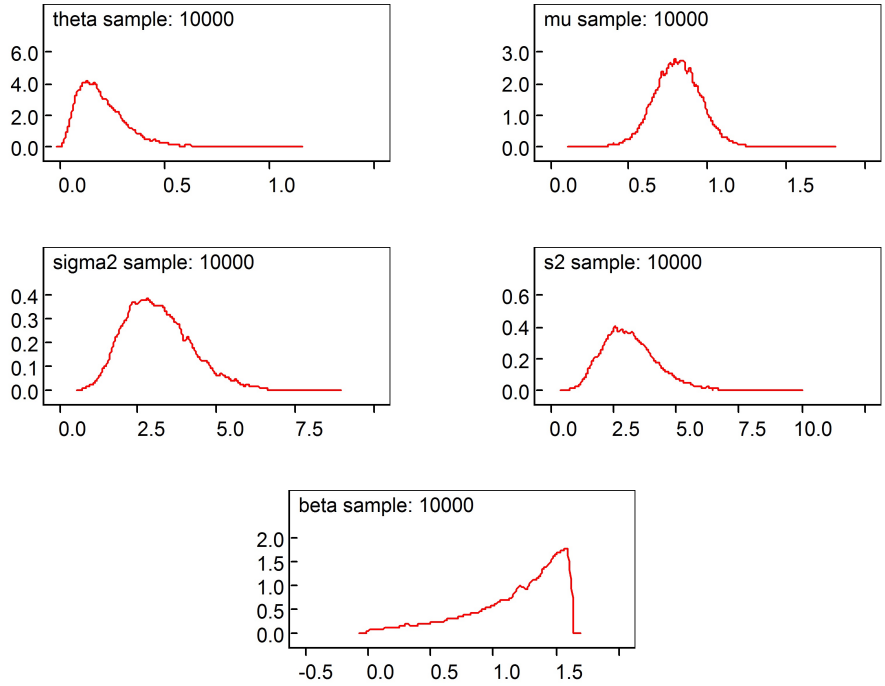

Figure 2. Posterior densities for β, θ, μ, σ^2 and s^2 for $\nu \sim$ Gamma$(1, 0.1)$.

all have large variances, so are suitably vague, but the distributions themselves have different shapes. In all cases, there is learning for ν from the data, although, under the more dispersed priors, the tail of the posterior is affected by the prior. However, the posterior changes are certainly less dramatic than the changes in the prior. We report the estimation results in terms of the mean and the 95% credible interval in Table 2.

Table 2. Posterior estimates and Credible Intervals of DPLNM model parameters for Different ν Gamma Priors.

	Gamma(2, 0.9)			Gamma(3, 0.05)		
Parameter	Mean	SD	CI	Mean	SD	CI
μ	0.835	0.209	(0.755, 1.45)	0.893	0.269	(0.748, 2.07)
s^2	3.46	1.29	(3.68, 4.70)	3.96	1.59	(3.69, 5.83)

The figure show that posterior point estimates under the three different priors for ν all give relatively very similar distributions, suggesting that the model is not very sensitive to changes in the prior distribution. However in some instances a Gamma$(3, 0.05)$ prior distribution consistently gives wider credible intervals. In all cases, there is learning for ν from the data, although, under the more dispersed priors, the tail of the posterior is affected by the prior. However, the posterior changes are certainly less dramatic than the changes in the prior. This demonstrated that a relatively small prior presence is adequate on posterior inference for DPLNM model implementation.

4.3 Real Dataset

The case study employed in this paper consists data on remission times, for leukemia patients taken from Freireich et al., 1963. The study involves two treatments, 6-Mercaptopurine (6-MP) and a Placebo each with 21 patients. The values of the 6-MP treatment were heavily censored while for the placebo they were completely observed. As in the simulated example, we used the same prior distributions and a MCMC algorithm with 10000 iterations (5000 to burn-in) to fit the data. Freireich et al., 1963 had shown that patients received the 6-MP treatment have a longer survival rate than the patients in the placebo group. We see from the resulting parameter estimates, under three prior choices, shown in Table 3 for censored values of 6-MP treatment, that there is a correspondence between the result obtained by Freireich et al., 1963 and ours. Thus posterior inference is robust with respect to the choice of prior on ν.

Table 3. Posterior estimates and Credible Intervals of DPLNM model parameters for Different ν Gamma Priors for 6-Mercaptopurine (6-MP).

	Gamma(1, 0.1)			Gamma(2, 0.9)			Gamma(3, 0.05)		
Par	Mean	SD	CI	Mean	SD	CI	Mean	SD	CI
μ	0.826	0.209	(0.78, 1.21)	0.835	0.219	(0.755, 1.45)	0.893	0.269	(0.748, 2.07)
s^2	3.46	1.29	(3.68, 4.70)	3.76	1.41	(3.99, 5.35)	3.96	1.59	(3.69, 5.93)

Of more interest is the effect of these prior choices on posterior inference for survival quantities. Since the DPLNM model is sensitive to the choice of prior survival quantities such as the survival function, mean survival time of a patient or the expected value of the hazard function can be estimated.

5. Conclusions and future Work

We have investigated the effect of prior choices on posterior inference for DPLNM model implementation and demonstrated that a relatively small prior presence is adequate. The priors we used were all vague, to reflect our lack of knowledge about the parameters, but of different forms. Since the results showed relatively similar expected values for the parameters, we concluded that the DPLNM model is relatively robust to the form of prior used.

While mixture mixtures models seem plausible, the complexity of adding more parameters to be estimated can result to poor estimation. The decision on what assumption to impose is arbitrary hence can result in slow convergence of chains. Thus these complications may not warrant mixtures as a better alternative. For these reasons, this may present possibilities for future work.

Finally, as an extension we may also test the fit by use of Compound distribution model where we combine distributions in such a way as to obtain an approximate aggregate distribution for estimation, and then consider various censoring mechanisms.

References

Antoniak, C.E. (1974). *Mixtures of Dirichlet processes with applications to nonparametric problems*. Ann. Statist. 2, 1152-1174.

Arjas, E. and Gasbarra, D. (1994). *Nonparametric bayesian inference from right censored*

survival data, using the gibbs sampler. Stat Sin, 4:505524.

Berger, J. (2006). *The case for objective bayesian analysis*. Bayesian Analysis, 1(3):385402.

Bernardo, J.M. & Smith, A.F.M. (1994). *Bayesian Theory*. Wiley, Chichester.

Bush, C.A. & MacEachern, S.N. (1996). *A semiparametric Bayesian model for randomised block designs*. Biometrika 83, 275-285.

Congdon, P.D. (2010). *Applied Bayesian hierarchical methods*. Chapman & Hall/CRC. Taylor & Francis Group. Boca Raton, Florida.

Diebolt, J and Robert, CP.(1994). *Estimation of finite mixture distributions through Bayesian sampling*. Journal of the Royal Statistical Society: Series B, 56, 363375.

Doksum, K. (1974). *Tailfree and neutral random probabilities and their posterior distributions*. Annals of Probability, 2:183201.

Doss, H. and Huffer, F. (2003). *Monte Carlo methods for Bayesian analysis of survival data using mixtures of Dirichlet process priors*. Journal of Computational and Graphical Statistics, 12(2):282307.

Escobar, M.D. & West, M. (1995). *Bayesian density estimation and inference using mixtures*. J. Amer. Statist. Assoc. 90, 577-588.

Ferguson, T. (1973). *A Bayesian analysis of some nonparametric problems*. Annals of Statistics, 1(2):209230.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman and Hall, New York.

Hjort, N. (2003). *Topics in non-parametric Bayesian statistics. In: Highly structured stochastic systems*. Oxford Univ. Press, Oxford.

Hjort, N. (1990). *Nonparametric Bayes estimator based on beta processes in models for life history data*. Annals of Statistics, 18:12591294.

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

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

Lunn, DJ., Thomas, A., Best, N and Spiegelhalter, D. (2000). *WinBUGS-A Bayesian modelling framework: Concepts, structure, and extensibility*. Statistics and Computing, 10, 325337.

Marin, JM., Mengersen, K and Robert, CP. (2005b). *Bayesian Modelling and Inference on Mixtures of Distributions*. Handbook of Statistics 25, D. Dey and C.R. Rao (eds). Elsevier-Sciences.

Mazzuchi, T. and Singpurwalla, N. (1985). *A bayesian approach to inference for montone failure rates*. Stat Probab Lett 3, pages 135141.

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

Muller, P. and Quintana, F. (2004). *Nonparametric bayesian data analysis*. Statistical Sciences, 19:95110.

Ntzoufras, I. (2009). *Bayesian Modelling using WinBUGS*. Wiley, New Jersey.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. MIT Press, Cambridge.

Robert, CP and Casella, G. (2000). *Monte Carlo Statistical Methods*. Springer, New York.

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

Spiegelhalter, N., Best, N., Carlin, B and vanderLinde, A. (2002). *Bayesian measures of model complexity and fit*. Journal of the Royal Statistical Society:Series B, 64 (4), 583639.

Stephens, M. (1997). *Bayesian Methods for Mixtures of Normal Distributions. PhD thesis*, The University of Oxford.

West, M., Muller, P. & Escobar, M.D. (1994). *Hierarchical priors and mixture models, with application in regression and density estimation*. In Aspects of uncertainty: Atribute to D.V. Lindley (eds A.F.M. Smith & P. Freeman). New York: Wiley.

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