ISSN XXXX XXXX © 2019 IJESC

Research Article Volume 9 Issue No.3

Bayesian Approach to Survival Modeling of Remission Duration for Acute Leukemia

Olawale B. Akanbi[1], Oladapo M. Oladoja[2], Christopher G. Udomboso[3]
Department of Statistics
University of Ibadan, Ibadan, Nigeria

Abstract:
The problem of analyzing time to event data arises in a number of applied fields like biology and medicine. This study constructs a survival model of remission duration from a clinical trial data using Bayesian approach. Two covariates; drug and remission status, were used to describe the variation in the remission duration using the Weibull proportional hazards model which forms the likelihood function of the regression vector. Using a uniform prior, the summary of the posterior distribution; Weibull regression model of four parameters ($\eta, \mu, \beta_1, \beta_2$) was obtained. With Laplace transform, initial estimates of the location and spread of the posterior density of the parameters were obtained. In this present study, data from children with acute leukemia was used. The information from the Laplace transform was used to find a density for the Metropolis random walk algorithm from Markov Chain Monte Carlos simulation to indicate the acceptance rate (24.55%).

Keywords: Clinical trial, covariates, Laplace transform, Metropolis random walk algorithm.

I. INTRODUCTION

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. The starting point of the study must be defined. In a survival analysis, we usually refer to the time variable as survival time**,** because it gives the time that an individual has "survived" over some follow up period [3]. Survival Modeling is constructing model for lifetimes in a study. Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations. Bayesian data analysis involves setting up a full probability model, conditioning on observed data and evaluating the fit of the model and the implication of the posterior distribution [2].

Leukemia is a cancer of blood-forming tissues, hindering the body's ability to fight infection. Fewer than 100 thousands per year in Nigeria. Common types includes: Acute Lymphotic Leukemia (ALL - common in children), Acute Myelogenous Leukemia (AML - common in children and adults), Chronic Lymphotic Leukemia (CLL - common in adults) and Chronic Myelogenous Leukemia (CML - common in adults). The various symptoms of Leukemia includes: excessive sweating, fatigue and weakness, unintentional weight loss, bone pain and tenderness, painless, swollen lymph nodes, liver enlargement, red spots on the skin, bleeding and bruising, fever or chills, frequent infections. This study is aimed at studying the remission duration after the effect of drug treatment on acute leukemia. The data for this study is the results of a clinical trial of a drug 6-mercaptopurine (6-MP) versus a placebo in 42 children with acute leukemia. The trial was conducted at 11 hospitals. Patients were selected who a complete or partial remission of their leukemia had induced by treatment with the drug prednisone. The trial was conducted by matching pairs of patients at a given hospital by remission status (complete or partial) and randomizing within the pair to either a 6-MP or placebo maintenance therapy. Patients were followed until their leukemia returned (relapse) or until the end of the study (in months). Bayesian inference was used because the difficulty in fitting survival models especially in the presence of complex censoring schemes. The introduction of a Bayesian analysis of the four-parameter generalized modified Weibull (GMW) distribution in presence of cure fraction, censored data and covariates was considered applicable to data from patients with gastric adenocarcinoma by obtaining inferences by using MCMC (Markov Chain Monte Carlo) method [4]. It was observed that survival of African American breast cancer patients follows the Weibull probability distribution by fitting the model to represent the survival ability of the general population and incorporated other patient specific covariate factors affecting the survival time to predict the life time of a patient [5]. Proposal of various statistical methods based on Map Reduce for selecting relevant feature by using Map Reduce-based k-nearest neighbor to classify microarray data. After comparative analysis, it was observed that the models consume much less execution time than models in processing big data [6].

II. METHODOLOGY

To construct a model for life times in a survival study for a set of n individuals, one observes the lifetimes

$$t_1, t_2, ..., t_n \qquad (1)$$

It is possible that some of the lifetimes are not observable since some individuals are still alive at the end of the study. In this case, we represent the response by the pair (t_i, δ_i) , where t_i is the observation and δ_i is the censoring indicator. If $\delta_i = 1$, the observation is not censored and t_i is the actual survival time. Otherwise, when $\delta_i = 0$, the observation t_i is the censored time. Describing the variation times in the survival times using p covariates $x_1, x_2, ..., x_p$, we describe

the relationship by using Weibull proportional hazards model. This model is expressed as the log linear

$$\log t_i = \mu + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \sigma u_i \quad (2)[1]$$

Where $x_{i1}, x_{i2}, ..., x_{ip}$ are the values of the p covariates for the ith individual and u_i is assumed to have a Gumbel distribution. There are $p+2$ unknown parameters in this model, the p regression coefficients, the constant term μ, and the scale parameter σ.

The density of the log time, $y_i = \log t_i$ is given by

$$f_i(y_i) = \frac{1}{\sigma}\exp(z_i - e^{z_i}), \quad (3)\ [1]$$

Where $z_i = (y_i - \mu - \beta_1 x_{i1} - ... - \beta_p x_{ip})/\sigma$.. Also, the survival function of the ith individual is given by

$$S_i(y_i) = \exp(-e^{z_i}). \quad (4)[1]$$

Then the likelihood function of the regression vector $\beta = (\beta_1, ..., \beta_p), \mu$ and σ is given by

$$L(\beta, \mu, \sigma) = \prod_{i=1}^{n} \{f_i(y_i)\}^{\delta_i} \{S_i(y_i)\}^{1-\delta_i}. \quad (5)[1]$$

Suppose we assign μ, β uniform priors and the scale parameter σ the usual non informative prior proportional to $1/\sigma$. Then the posterior density is given, up to a proportionality constant, by

$$g(\beta, \mu, \sigma \mid data) \alpha \frac{1}{\sigma} L(\beta, \mu, \sigma). \quad (6)[1]$$

III. RESULTS AND DISCUSSION

We calculate the Kaplan – Meier estimates of the overall survival according to the leukemia data and the relation between the patients treated with 6-Mercaptopurine and Placebo drugs. We confirmed that the subjects under the two treatments have different survival time patterns as shown in Figure 1 below.

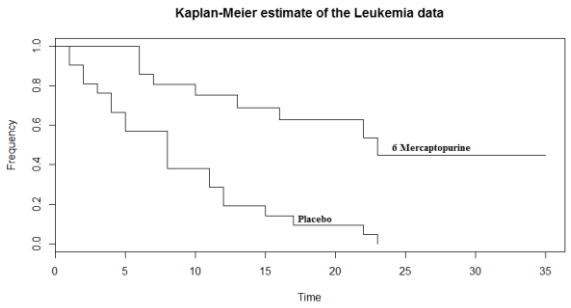

Figure.1. kaplan – meier estimate of the leukemia patients.

Following this, we fitted a log linear model where the response variable TIME was the survival time in days following randomization to one of the two drug treatments. Also, we recorded a censoring variable STATUS that indicates if TIME is an actual survival time (STATUS = 1) or censored at that time (STATUS = 0). The two covariates are TREAT, the treatment group, and REMSTA, the remission status of the patients. The log linear model is

$$\log TIME_i = \mu + \beta_1 TREAT_i + \beta_2 REMSTA + \sigma u_i \quad (7)$$

Which is best explained in equation (8) as

$$\log TIME_i = 2.3333 + 1.2695 TREAT_i - 0.0494 REMSTA + \sigma u_i \quad (8)$$

Unlike the normal regression model, the posterior distribution of the parameters of this survival model cannot be simulated by standard probability distributions. By making all parameters real valued by transforming the scale parameter σ to $\eta = \log \sigma$. We computed the joint posterior of $\theta = (\eta, \mu, \beta_1, \beta_2)$.

The posterior mode of

$$\theta = -0.3151, 2.3336, 1.2695, -0.0496.$$

The associated variance-covariance matrix V is

$$\begin{pmatrix} +0.0221 & -0.0156 & +0.0154 & +0.0052 \\ -0.0156 & +0.3029 & -0.0220 & -0.1599 \\ +0.0154 & -0.0220 & +0.0960 & -0.0047 \\ +0.0052 & -0.1599 & -0.0047 & +0.0928 \end{pmatrix} \quad (9)$$

We then find a proposal density (which is a multivariate normal density) for the Metropolis random walk chain with a satisfactory rate of 24.55%. Thereafter, we simulated draws from the marginal posterior densities of β_1, β_2 & σ. The simulated draws from the marginal posterior densities of β_1 (corresponding to TREAT) is displayed in Figure 2, while that of β_2 (corresponding to AGE) is displayed in Figure 3 and the scale parameter σ is displayed in Figure 4.

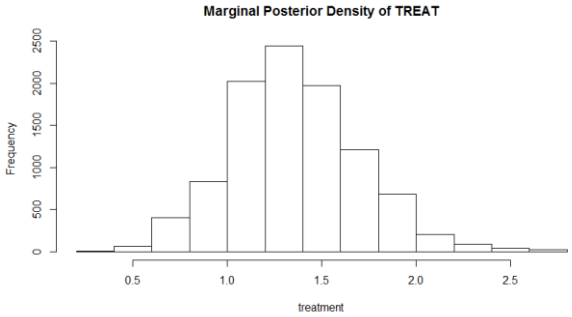

Figure.2. Histogram Showing the Posterior Probability of Treatment

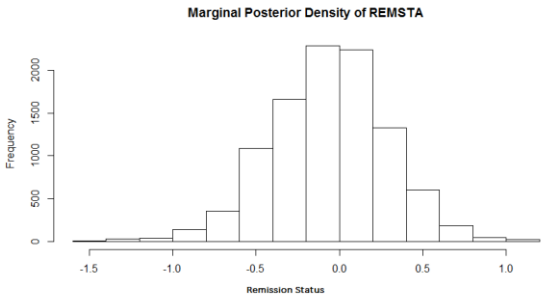

Figure.3. Histogram Showing the Posterior Probability Of Remission Status Of Patient

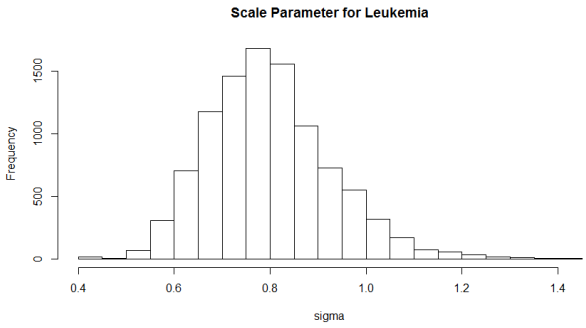

Figure.4. Histogram Showing the Posterior Probability Of The Scale Parameter

A simulated sample of draws for estimating the survival curve for an individual in the treatment group is summarized in Table 1 by the 5th, 50th and 95th percentiles. The graph is displayed in Figure 5. The green line corresponds to the 5th percentile, the blue lines corresponds to the 50th percentile and the red line corresponds to the 95th percentile of the posterior of survival for each time t. The procedure was repeated for a grid of t values between 0 and 25 days.

Table .1. Percentiles For The Simulated Draws

	η	μ	β_1	β_2
5%	0.6152	3.7706	2.2454	0.5374
50%	0.7882	10.2902	3.7907	0.9627
95%	1.0309	30.3738	6.9318	1.6552

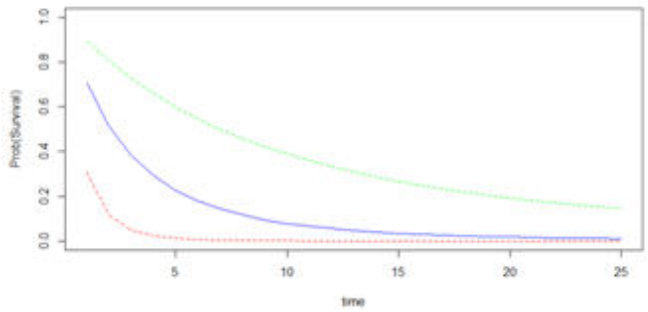

Figure.5. posterior distribution of probability of survival s(t) for leukemia patients.

IV. CONCLUSION

From our results in Figure 2, we see that the value TREAT = 1.25 is at the center of the posterior distribution and so, there is insufficient evidence to conclude from these data that TREAT ≠ 1.25. This means that there is insufficient evidence to conclude that the risk of death is higher (or lower) with drug treatment on leukemia patients. Since we are interested in the estimation of a patient's survival curve, we simulated values from the joint posterior distribution of the remission status and scale parameter using 10000 iterations. We set up a grid of values of t, simulate a sample of values of S(t) for each value of t on the grid and then summarize the posterior sample by computing the percentiles. Graphing these percentiles as a function of the time variable t, since there is little evidence that TREAT ≠ 1.25, this survival curve represents the risk for the leukemia patients under the two drug treatments.

V. REFERENCES

[1]. Albert,J. (2009). Bayesian Computation with R. *Springer Texts in Statistics.*

[2]. Gelman,A, Carlin,J.B., Stern,H.S., Dunson,D.B., Vehtari, A. and Rubin, D. B. (2014). Bayesian Data Analysis. *Taylor & Francis Group.*

[3]. Klein, J. P. and Moeschberger, M. L. (2003). Survival Analysis Techniques for Censored and Truncated Data. *Springer Texts in Statistics.*

[4]. Martinez, E. Z., Achar, J. A., Jacome, A. A. and Santos, J. S. (2013). Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. Computer Methods and Programs in Biomedicine. 343-355.

[5]. Minh, H. P. (2014). Survival Analysis - Breast Cancer. *Undergraduate Journal of Mathematical Modeling: One + Two* 6, Issue 1, Article 4.

[6]. Mukesh, K. and Santanu, K. R. (2016). Analysis of Micro array leukemia Data using an efficient Map Reduce-based k-nearest-neighbor classifier. *Journal of Biomedical Informatics* 60, 395-409.

[7]. Moslemi, A., Mahjub, H., Saidijam, M., Poorolajal, J. and Soltanian, A.R. (2016). Bayesian Survival Analysis of High-Dimensional Microarray Data for Mantle Cell Lymphoma Patients. *Asian Pacific Journal of Cancer Prevention*, 17(1), 95-100.

[8]. Omurlu,I. K.,Ozdama, K. Ture, M. (2009). Comparison of Bayesian Survival analysis and Cox regression analysis in simulated and breast cancer data sets. *Expert Systems with Applications*, 36, 11341-11346.