ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

Bayesian Survival Analysis of Regression Model Using Weibull

Yasmin Khan[1] and Athar Ali Khan[2]

Research Scholar, Department of Statistics & Operations Research, Aligarh Muslim University, Aligarh, Uttar Pradesh, India [1]

Professor, Department of Statistics & Operations Research, Aligarh Muslim University, Aligarh, Uttar Pradesh, India [2]

Abstract: Weibull distribution is one of the most important and flexible distributions in survival analyses. In this paper, Bayesian regression analysis with censoring mechanism is carried out for a hypothetical survival data problem. Throughout the Bayesian approach is implemented using `R` and appropriate illustrations are made.

Keywords: Bayesian Inference, Right censoring, LaplaceApproximation, Survival function.

I. INTRODUCTION

Survival analysis is used when we wish to study the occurrence of some event in a population of subjects and the time until the event is of interest. This time is called survival time. In literature there are many different modeling approaches to survival analysis. In this paper modeling is done by Weibull distribution. There are many life time models but Weibull is used quite effectively to analyzed skewed data sets. Probability density and survival functions of Weibull model is respectively given as

$$f(y) = \frac{a}{b}\left(\frac{y}{b}\right)^{a-1} \exp\left[-\left(\frac{y}{b}\right)^{a}\right]$$

$$S(y) = \exp\left[-\left(\frac{y}{b}\right)^{a}\right]$$

A very important feature which create a special problem in survival analysis, called censoring mechanism. In this paper an attempt has been made to handle the complexity occurs due to censored observation under Bayesian paradigm and we discussed only about right censoring. The likelihood function for right censored data as,

$$L = \prod_{i=1}^{n} \Pr(y_i, \delta_i) = \prod_{i=1}^{n} [f(y_i)]^{\delta_i} [S(y_i)]^{1-\delta_i}$$

Where δ_i is an indicator variable which takes value 1 if observation is censored and 0 otherwise. Section III includes the hypothetical survival data of fifteen patients with censoring suffering from a disease and they were treated by two different treatments. Section IV and Section V consist of Bayesian regression analysis of treatment 1 and treatment 2 using `LaplacesDemon` Hall [1] package which is available in `R` R Development Core Team[2]. The goal of `LaplacesDemon` is to provide a complete and self-contained Bayesian environment within `R`. The main function of this package which is used in the paper is `LaplaceApproximation`. This function gives the approximated posterior estimates of the parameters in Bayesian framework. In order to deal with censoring mechanism we have developed a function which works well for the analysis of survival data. Comparison of survival curves for two treatments by using Weibull model is reported in section VI. Finally in the last a brief discussion and conclusion is given in section VII.

II. RELATED WORK

A significant amount of work has been done for estimating the linear regression model for censored data Miller[3], Buckley and James[4], Koul, Susarla, and Van Ryzin[5] and Collet[6]. Puja et al. [7] discuss the Bayesian survival analysis of head and neck cancer data and conclude that which therapy has give better performance for patients. In this paper all analyses and computation were undertaken using R software.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

III.DATA SET

For illustrative purpose a hypothetical data set given in Table I will be used. The event for this data set is the death of the patients, and so the censored data are those where the outcome is survived or unknown. Bayesian fitting of Weibull model for this data can be done in R by using laplace approximation. Here, fitting is done between survival time as response variable and treatment and age as regressors.

Table I: Survival time (in days), age for a group of patients diagnosed with a disease and receiving one of two treatments.

Time	Treatment	Age	Time	Treatment	Age
1	2	75	12	1	71
1	2	79	15+	1	73
4	2	85	22	2	66
5	2	76	25+	1	73
6+	2	66	37	1	68
8	1	75	55	1	59
9+	2	72	72+	1	61
9	2	70			

IV. BAYESIAN REGRESSION ANALYSIS FOR TREATMENT 1 OF THE DATA

The Weibull distribution is a parametric function widely used in survival analysis. Weibull distribution has two parameters, shape and scale.

$$y \sim weibull(a, b)$$

Where, a and b are shape and scale parameters and $\mu = X\beta$,

$$\beta \sim N(0,1000)$$

$$a \sim halfcauchy(25)$$

Thus, the log-likelihood is,

$$\log\ L(y \mid a, b) = \sum_{i=1}^{n} \log\left(\frac{a}{b}\left(\frac{y_i}{b}\right)^{a-1} \exp\left[-\left(\frac{y_i}{b}\right)^{a}\right]\right)$$

Prior,

$$p(a) = \frac{2b}{\pi\left(y^2 + b^2\right)}, \qquad y > 0$$

ISSN: 2319-8753

International Journal of Innovative Research in
Science, Engineering and Technology
(An ISO 3297: 2007 Certified Organization)
Vol. 2, Issue 12, December 2013

Thus, the posterior upto proportionality can be specified as,

$$p(a,b \mid y) \sim p(a)\log L(y \mid a,b)$$

Now, we consider Weibull regression model with two predictor treatment (treatment1 and treatment 2) and age. Thus, the regression model is,

$$Time = \beta_o + \beta_1 tretment_i + \beta_2 age_i + e_i$$

The fitting of Weibull model includes codes for creation of data and definition of model. R codes to fit Weibull model is being described as

```
y<-c(1,1,4,5,6,9,9,22)
censor<-c(1,1,1,1,0,0,1,1)
age<-c(75,79,85,76,66,72,70,66)
trt<-rep(2,8)
N<-8
J<-3
X<-cbind(1,age,trt)
mon.names<-c("LP","shape")
parm.names<-as.parm.names(list(beta=rep(0,J),log.shape=0))
MyData<-list(J=J,X=X,mon.names=mon.names,parm.names=parm.names,y=y)
Initial.Values <- c(rep(0,J), log(1))
Model<-function(parm,Data)
{
beta<-parm[1:Data$J]
shape<-exp(parm[Data$J+1])
beta.prior<-sum(dnorm(beta,0,1000,log=T))
shape.prior<-dhalfcauchy(shape,25,log=T)
mu<-tcrossprod(beta,Data$X)
scale<-exp(mu)
LL<-sum(censor*dweibull(Data$y,shape,scale,log=T)+
(1-censor)*pweibull(Data$y,shape,scale,log.p=T,lower.tail=F))
LP<-LL+beta.prior+shape.prior
Modelout<-list(LP=LP,Dev=2*LL,Monitor=c(LP,shape),
yhat=mu,parm=parm)
return(Modelout)
}

M3<-LaplaceApproximation(Model,Initial.Values,Data=MyData,
Sample=10000,Iterations=10000)

Print(M3)
```

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

Table II: Approximated posterior summary of treatment 1 using LaplaceApproximation function with posterior mode, posterior sd and their quantiles.

	Mode	SD	2.5%	50%	97.5%
Intercept	4.208	0.45	3.31	4.28	5.17
Treatment 1	4.368	0.62	3.35	4.26	5.22
Age	4.293	0.54	3.34	4.25	5.21
Shape	0.977	0.33	0.56	0.99	1.67

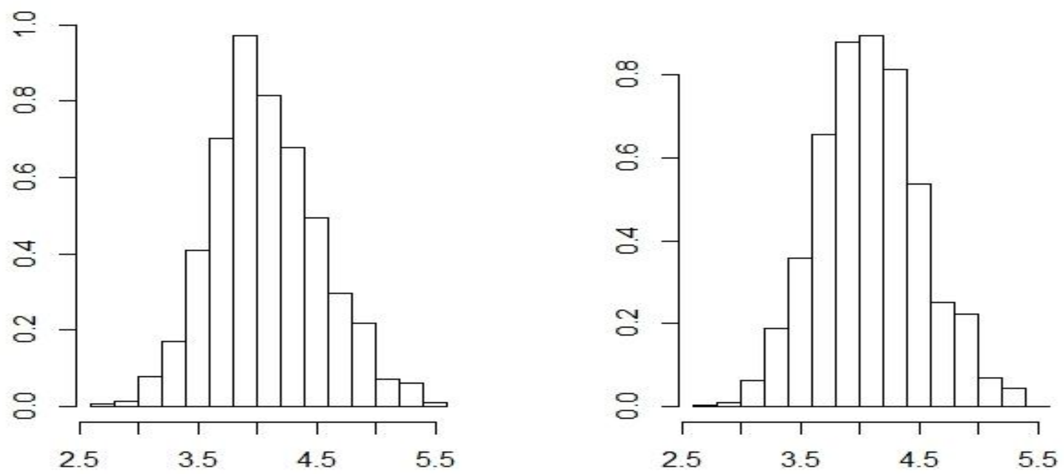

Fig. 1. Histogram showing the posterior density estimates of treatment1 (left) and age (right).

V. BAYESIAN REGRESSION ANALYSIS FOR TREATMENT 2 OF THE DATA

The R codes for the analysis of treatment 2 is almost same as given in section IV and are not shown here just to save the space. The approximated posterior summary is reported in Table III. Table III gives the posterior mode and posterior sd for treatment 2 also given with respective quantiles.

Table III: Approximated posterior summary of treatment2 using `LaplaceApproximation` function with posterior mode, posterior sd and their quantiles.

	Mode	SD	2.5%	50%	97.5%
Intercept	2.383	0.47	1.56	2.34	3.45
Treatment 2	2.356	0.46	1.49	2.32	3.38
Age	2.294	0.47	1.34	2.28	3.26
Shape	0.923	0.18	0.61	0.92	1.32

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

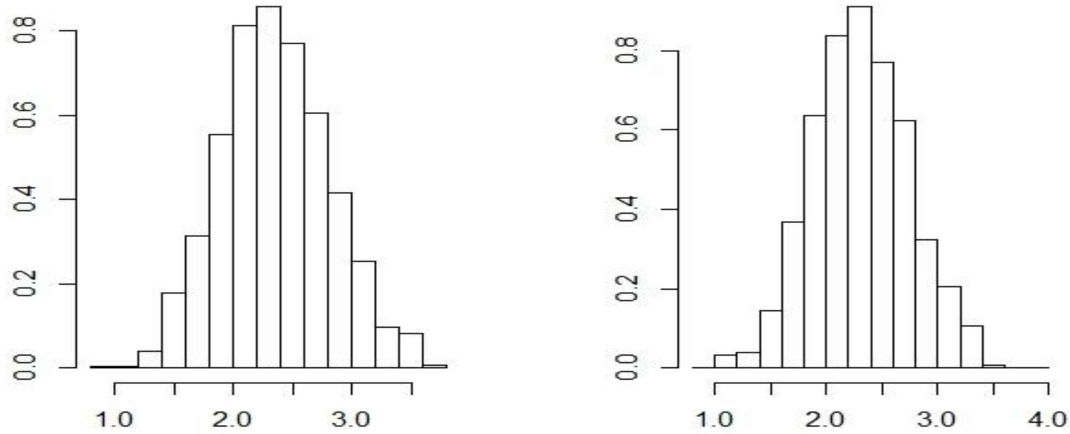

Fig. 2. Histogram showing the posterior density estimates of treatment2 (left) and age (right).

VI. COMPARISON OF THE TWO TREATMENTS BASED ON ESTIMATED SURVIVAL FUNCTION

Survival curves are used to study times required to reach any well-defined endpoint. In this section we will compare the Bayes estimate of survival function based on Weibull model in order to made conclusion that which treatment is better for the patients. It could be seen that the patients receiving treatment1 have higher survival probabilities than those receiving treatment2.

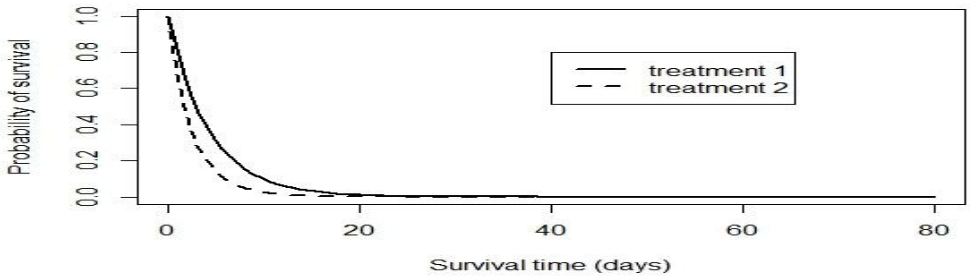

Fig. 3. Estimated survival curves of treatment1 and treatment2 using Weibull model of the data given in table I.

VII. CONCLUSION

Survival analysis and Bayesian methods are the two most active areas in the statistical literature. Survival problems become more complicated because of the presence of censoring. This paper successfully handles and manages censoring mechanism. Laplace approximation has made a great contribution for Bayesian estimation. Numerical and graphical illustrations are made for a hypothetical data. In this data patients suffering from the disease are observed of their survival times when treated with two different treatments.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 12, December 2013

Table II and table III summarizes the posterior mode and sd and the associated covariance matrix. Assuming normality, these outputs characterizes the marginal posterior distributions. This analysis is done under Weibull model because survival data are generally not symmetric and are positively skewed and Weibull model is also skewed in nature. The posterior density estimates of treatment 1 and age and treatment 2 and age of the data is shown in the form of histogram in Fig 1 and Fig 2 respectively. Comparing survival functions for two treatments in section V has always been an important task for medical data analyst. On the basis of this comparison it could be concluded that treatment 1 offers a better performance for patients than treatment 2 using Weibull model.

ACKNOWLEDGEMENT

The author would like to thank University Grants Commission (UGC), New Delhi for financial assistance.

REFERENCES

[1] B. Hall, “LaplacesDemon: Software for Bayesian Inference”, R package version 11.12.05., URL http://cran.r-project.org/web/packages/LaplacesDemon/index. Html 2011.

[2] R Development Core Team ,” R: A Language and Environment for Statistical Computing”, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org, 2011

[3] R. G. Miller, “Least Square Regression with Censored Data”, Biometrika, 63, 449-464, 1976.

[4] J. J. Buckley, and I. R. James, *“*Linear Regression with Censored Data”, Biometrika, 66, 429-436, 1979.

[5] H. Koul, V. Susarla, and J. Van Ryzin, “ Regression Analysis with Randomly Right-Censored Data”, The Annals of Statistics, 9, 1276-1288,1981.

[6] D. Collet, “Modelling Survival Data in Medical Research”, London: Chapman & Hall, 1994.

[7] M. Puja, K. S. Puneet, R. S. Singh, and S. K. Upadhyay, “Bayesian Survival Analysis of Head and Neck Cancer Data Using Lognormal model”, Communications in Statistics-Simulation and computation, 2013.