IJASRM
International Journal of Advanced Scientific Research and Management, Volume 4 Issue 9, Sep 2019
www.ijasrm.com
ISSN 2455-6378

Bayesian Regression Modelling For Censored Survival Data Applying Lomax Distributions

N.Sundaram
Assistant Professor, Department of Statistics
Presidency College (Autonomous, Affiliated to University of Madras)
Chennai-600005, India.

Abstract
This article on Regression and Bayesian regression methods describes Lomax and Exponential Lomax models. The proposed models have been based on censored survival data. The parameters are estimated using maximum likelihood method. The model performance and behavior of the explanatory variables are studied with the help of standard error (SE), Monte Carlo Standard Errors (MCSE), Markov Chain Convergence, Autocorrelation, Geweke Diagnostics, Posterior densities, Akaike Information Criterion (AIC) and Deviance Information Criterion (DIC). Exponential Lomax (EL), Exponential Lomax Regression (ELR) and Exponential Lomax Bayesian regression (ELBR) models are performed well than the Lomax (L), Lomax Regression (LR) and Lomax Bayesian Regression models (LBR). Also the ELBR model is found to be better and more suitable in describing the malignant melanoma data than the other models.
Keywords: *Lomax, Exponential Lomax, Geweke, MCSE, Markov Chain, Autocorrelation, Posterior densities, AIC, DIC.*

1. Introduction

The Lomax or Pareto II distribution was proposed by Lomax (1954). This distribution has found wide applications in areas such as analysis of the business failure, life time data, income and wealth inequality, size of cities, actuarial science, medical and biological sciences, engineering, lifetime and reliability modeling. Some properties and moments for the Lomax distribution have been discussed by Ahsanullah (1991). Recurrence relations for moments of lower generalized order statistics from Exponentiated Lomax distribution and its characterization were studied by Abdul-Moniem (2012). The Exponential Lomax Distribution was introduced and studied by El-Bassiouny et al., (2015). Ieren and Kuhe (2018) derived and studied the properties and applications of Lomax-Exponential distribution.

This article presents the Probability density function, Survivor function and Likelihood function of the Lomax and Exponential Lomax models in section 2. Regression concepts applied in Lomax and Exponential Lomax models for censored data is presented in section 3. In section 4, the Bayesian regression methods were applied to Lomax and Exponential Lomax models for censored survival data. The database is provided in section 5. An application of the proposed models has been explained in section 6. Discussions and Conclusions are presented in section 7.

2. Parametric Models

The study of modeling censored survival data has focused on predicting the probability of survival, or median lifetime, by comparing the survival distributions of human patients and the identification of risk or prognostic factors. In this section two models have been used widely to describe survival time. The probability density function, survivor function and likelihood function for Lomax and Exponential Lomax models are presented below.

A random variable $T \sim L(\lambda, \alpha)$ is used to indicate the Lomax model with scale parameter λ and shape parameter α. It has the probability density function

$$f(t)_L = \frac{\lambda\alpha}{(1+\lambda t)^{\alpha+1}} \; ; \; t > 0 \; and \; \lambda, \alpha > 0. \qquad (1)$$

The Survivor function is

$$S(t)_L = (1+\lambda t)^{-\alpha}. \qquad (2)$$

The Exponential Lomax model (El-Bassiouny et al., 2015) is defined by

$$f(t)_{LE} = \left(\frac{\lambda\alpha}{\beta}\right)\left(\frac{\beta}{t+\beta}\right)^{-\alpha+1} e^{-\lambda\left(\frac{\beta}{t+\beta}\right)^{-\alpha}} ;$$

$$t \geq -\beta \; and \; \lambda, \alpha, \beta > 0. \qquad (3)$$

The Survivor function is

$$S(t)_{LE} = e^{-\lambda\left(\frac{\beta}{t+\beta}\right)^{-\alpha}} \qquad (4)$$

The n pairs of observations for i^{th} individual is $(t_i, \delta_i ; i = 1,2,\dots,n)$, where δ_i is an indicator variable

$$\delta_i = \begin{cases} 1 \; if \; t_i \; is \; uncensored \\ 0 \; if \; t_i \; is \; censored. \end{cases}$$

IJASRM International Journal of Advanced Scientific Research and Management, Volume 4 Issue 9, Sep 2019
www.ijasrm.com
ISSN 2455-6378

The total likelihood function (Lee and Wang, 2003 and Collett, 1994) for each model can be written in the following way
For Lomax model (using (1) and (2)):

$$L_L = \prod_{i=1}^{n}\{f(t_i)_L\}^{\delta_i} \ \{S(t_i)_L\}^{1-\delta_i}, \quad (5)$$

For Exponential Lomax model (using (3) and (4)):

$$L_{EL} = \prod_{i=1}^{n}\{f(t_i)_{EL}\}^{\delta_i} \ \{S(t_i)_{EL}\}^{1-\delta_i}, \quad (6)$$

These models assume that the censoring times and survival times are independent. Parameters were estimated by maximizing expression (5) and (6).

3. Parametric Regression Models

A larger number of possible prognostic factors may be associated with the outcomes. One way to reduce the number of factors is to use a multivariate technique such as Non-Linear Mixed model. An attempt is made to examine the relationship between each individual factor and the dependent variable (survival time). To develop effective models significant factors are isolated and are used for the prediction survival times. Very often, a variable of significant prognostic value in one study is unimportant in another. Therefore, confirmation in a later study is very important in identifying prognostic factors.

Let $X = (x_1, x_2, \dots, x_p)$ denote a p-covariate vector of explanatory variables. The relationship of logarithm of survival time T and the covariates is linear and can be written as

$$logT = u + \sigma\varepsilon \quad (7)$$

where $u = \beta_0 + \sum_{j=1}^{p}\beta_j x_j, j = 1,2,\dots,p$, x_j are the covariates and β_j are the coefficients, $\sigma\ (>0)$ is an unknown scale parameter, and ε the error term is

$$\varepsilon = \frac{logt-u}{\sigma} \quad (8)$$

The Lomax regression model with scale parameter λ and shape parameter α is defined as follows.

$$\lambda = \exp\{-u/\sigma\}, \quad (9)$$

$$\alpha = 1/\sigma^{16} \quad (10)$$

The Exponential Lomax regression model with scale parameterλ, shape parameters α and β is defined as follows.

$$\lambda = \exp(u), \quad (11)$$

$$\beta = 1/\mu^{16}, \quad (12)$$

$$\alpha = \frac{\mu}{\sigma} \quad (13)$$

4. Parametric Bayesian Regression Models

For Lomax Bayesian regression model:
In the absence of prior information about the parameters in this model, we use diffuse Normal priors for the regression coefficients β_j and gamma distribution for scale parameter σ are as follows.

$$\beta_j \sim N(0, var)$$

and

$$\sigma \sim Gamma(shape, scale)$$

The Lomax Bayesian regression model with scale parameter λ and shape parameter α is defined as follows.

$$\lambda = \exp\{-u/\sigma\}, \quad (14)$$

$$\alpha = 1/\sigma^{16} \quad (15)$$

For Exponential Lomax Bayesian regression model:
Like in the Lomax Bayesian regression models, the absence of prior information about the parameters in this model, we use diffuse Normal priors for the regression coefficient β_j , Inverse-Gamma Distribution for shape parameter μ and gamma distribution for scale parameter σ are used as follows.

$$\beta_j \sim N(0, var),$$

$$\mu \sim Inverse - Gamma(shape, scale)$$

and

$$\sigma \sim Gamma(shape, scale)$$

The Exponential Lomax Bayesian regression model with scale parameter λ, shape parameters α and β is defined as follows:

$$\lambda = \exp(u), \quad (16)$$

$$\beta = 1/\mu^{16} \quad (17)$$

and

$$\alpha = \frac{\mu}{\sigma} \quad (18)$$

5. Database (Andersen et al., 1993)

The data consist of measurements made on 205 patients with malignant melanoma. This data contains survival time in days, patients status at the end of the study, patients sex, age in years at the time of the operation, year of operation, tumor thickness in mm. and indicator of ulceration. Each patient had their tumor removed by surgery. The surgery consisted of complete removal of the tumor together with about 2.5 cm. of the surrounding skin. Among the measurements taken were the thickness of the tumor and whether it was ulcerated or not. These are thought to be important prognostic variables in that patients with a thick and/or ulcerated tumor have an increased chance of death from melanoma.

6. Data analysis

In this section malignant melanoma data is used for the comparison of parametric models. The parameters, SE, MCSE, AIC and DIC values are estimated using maximum likelihood method and Bayesian approach and the values are presented in Table 1.

The L model is specified completely by the two parameters λ and α, and EL model is specified completely by three parameters λ, α and β. When we compare the parametric model L and EL, the EL model has the smaller AIC value (Table 1). EL model provides a better fit than the L model.

From Table 1, in the LR model, two explanatory variables tumor thickness and indicator of ulceration

are significant at 5% and 1% levels respectively. Similarly, three explanatory variables age in years at the time of the operation, tumor thickness and indicator of ulceration are significant at 5%, 1% and 0.1% levels in the ELR model. Moreover, the SE is reduced when fitting these two models. Further, ELR model has the smaller AIC value. Infer that the ELR model better fit than the LR model.

Three explanatory variables patients sex, age in years at the time of the operation and tumor thickness are significant at 5%, 1% and 0.1% in LBR model. Trace plots, autocorrelation plots, and posterior density plots for b1, b2, b3 and b4 are shown in Fig.1. The mixing of b1, b2, and b3 appear to be good at Markov chain Monte Carlo (MCMC) sample of 100000. It reveals that the sampling has gone well with no particular concerns about the convergence or mixing of the chains in the Lomax Bayesian regression model.

Similarly, ELBR model produced all the explanatory variables (patients sex, age in years at the time of the operation, tumor thickness in mm. and indicator of ulceration) are significant at 5%, 0.1%, 5% and 0.1% respectively. Trace plots, autocorrelation plots, and posterior density plots for b1, b2, b3 and b4 are shown in Fig.2. The mixing of b2 and b3 are appeared to be good at MCMC sample of 100000. It reveals that the sampling has gone well with no particular concerns about the convergence or mixing of the chains in the Exponential Lomax Bayesian regression model.

Moreover MCSE is reduced from LBR model to ELBR model. Exponential Lomax Bayesian Regression (ELBR) model has the smallest DIC value. Because of this, ELBR model fits better than the LBR model.

7. Discussions and Conclusions

The L and EL models play an important role in describing many probability distributions. This article is focused on Survival regression models with Bayesian approach i.e., L, EL, LR, ELR, LBR and ELBR models to analyze censored survival data of malignant melanoma patients. Among these EL, ELR and ELBR models are in good agreement based on AIC/DIC, SE, MCSE values and Posterior Plots. Clearly, the ELBR model provides a closer fit to the empirical survival function of malignant melanoma patients at MCMC sample of 100000. Also infer that among the explanatory variables, age at the time of the operation, the thickness of the tumor and ulcerate are important prognostic variables in that patients have an increased chance of risk from melanoma.

Table 1: Optimization Results for malignant melanoma data

Model	Parameter	Estimate (SE/MCSE)	AIC/ DIC
L	λ	0.0376 (0.0582)	466
	α	1.4209 (1.9666)	
EL	λ	0.00008 (0.0327)	**462**
	α	1.1104 (0.1212)	
	β	0.000001(0.0005)	
LR	λ	0.07216 (0.1369)	433
	α	3.0488 (5.1230)	
	Intercept	5.5751 (1.1678)	
	Sex	-0.4368(0.3013)	
	Age	-0.0113(0.0082)	
	Thickness	-0.0999(0.0457)*	
	Ulcer	-1.1523(0.3741)**	
ELR	λ	0.00008 (0.0012)	
	α	1.2246 (0.1413)	**432**
	β	0.00000004 (0.000013)	
	Intercept	-9.4902 (23.9343)	
	Sex	-0.3333(0.2216)	
	Age	-0.0134(0.0065)*	
	Thickness	-0.0848(0.0320)**	
	Ulcer	-0.9760(0.2722)***	
LBR (Number of MCMC iterations-100000)	λ	0.0044 (0.0590)	
	α	2.3981 (0.0050)	432
	Intercept	5.4325 (0.0692)	
	Sex	-0.4793(0.0067)*	
	Age	-0.0110(0.0004)**	
	Thickness	-0.1056(0.0015)***	
	Ulcer	-1.2482(0.0186)	
ELBR (Number of MCMC iterations-100000)	λ	0.0912 (0.0121)	**428**
	α	1.1683 (0.0178)	
	β	0.0025 (0.0162)	
	Intercept	-2.3952 (0.1304)	
	Sex	-0.3423(0.0195)*	
	Age	-0.0149(0.0003)***	
	Thickness	-0.0868(0.0014)*	
	Ulcer	-1.0569(0.0254)***	
* Significant at 5% ** Significant at 1% *** Significant at 0.1%			

ISSN 2455-6378

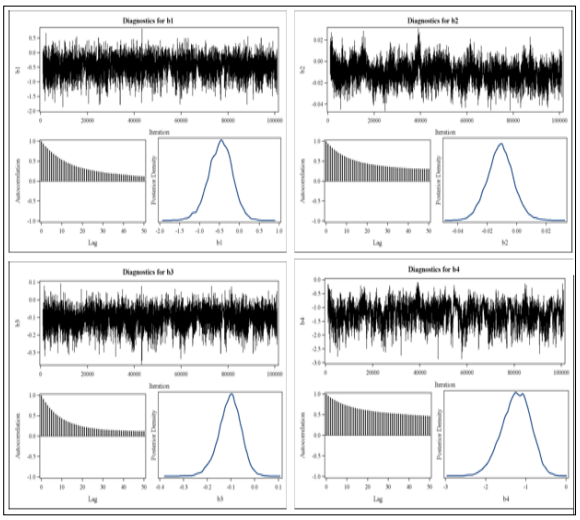

Fig.1 Posterior Plots for sex (b1), age (b2), thickness (b3) and ulcer (b4) in the Lomax Bayesian regression model.

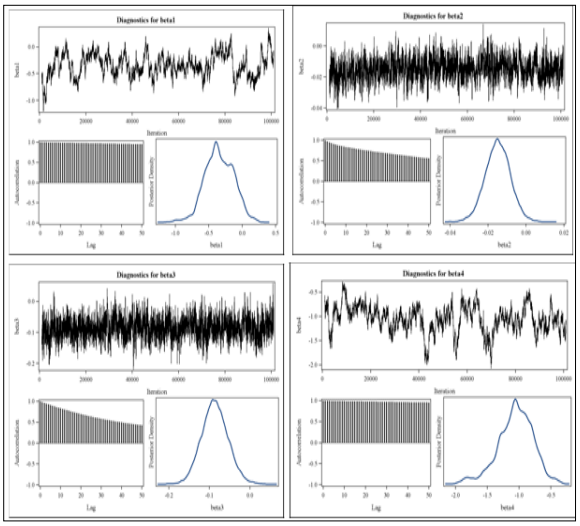

Fig.2 Posterior Plots for sex (b1), age (b2), thickness (b3) and ulcer (b4) in the Exponential Lomax Bayesian regression model.

References

[1] Abdul-Moniem I B, Recurrence relations for moments of lower generalized order statistics from exponentiated Lomax distribution and its characterization. International Journal of Mathematical Archive, 3: 2144–2150, (2012).

[2] Ahsanullah M, Recoed values of the Lomax distribution. Statistica Neerlandica, 45: 21–29, (1991).

[3] Andersen P K Borgan O Gill R D and Keiding N, *Statistical Models Based on Counting Processes*. Springer-Verlag, (1993).

[4] Collett D, Modelling survival data in medical research. Chapman & Hall, London, (1994).

[5] El-Bassiouny A H Abdo N F and Shahen H S, Exponential Lomax Distribution. International Journal of Computer Applications, 121(13): 24-29, (2015).

[6] Ieren T G and Kuhe D A, On the Properties and Applications of Lomax-Exponential Distribution. Asian Journal of Probability and Statistics, 1(4): 1-13, (2018).

[7] Lee E T and Wang J W, Statistical methods for survival data analysis. Third edition, John Wiley & Sons, Inc., publication, (2003).

[8] Lomax K S, Business failures: Another example of the analysis of failure data. Journal of the American Statistical Association, 49: 847–852, (1954).