Sule Lamido University Journal of Science & Technology Vol. 9 No. 1&2 [May, 2024], pp. 66-73
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

A Bayesian via Laplace Approximation of Power Lomax Model with Censored Data

Ogunde A. A.[1,*] , Phillips S. A.[2], Ajayi B.[3], and Oboh I. C.[4]

[1]Department of Statistics, University of Ibadan, Oyo State.
[2,4]Department of Statistics, Federal School of Statistics, Ibadan, Oyo State.
[3]Department of Mathematics and Statistics, Federal Polytechnic Ado-Ekiti, Ekiti State.

Corresponding author: debiz95@yahoo.com; Phone Number: 08067385628

Abstract

Power Lomax distribution is an extension of Lomax distribution that is more flexible, versatile and provides a better fit for skewed and censored data. In this work we introduce a solution to the closed form expression of the survival function of the model, which shows the model's adaptability and flexibility for modelling real lifetime data. Alternatively, Bayesian estimation by MCMC simulation via Laplace approximation was employed. Comparison using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) shows that the Power Lomax model is a good choice for fitting the survival models and Information Criterion (AIC) simulations by Markov Chain Monte Carlo Methods. Findings show that this procedure and method are better options for modelling Bayesian regression and survival/reliability analysis. However, the results of the censored data have been clarified by the simulation results.

Keywords: *Laplace Approximation, Markov Chain Monte Carlo Methods, Power Lomax model, Survival models.*

1. Introduction

In recent decade, Bayesian technique is now commonly used to model censored survival data. Its use is practically gaining dominance in the last few decades because it enables a more robust statistical tools to analyze complex data structures and parameters (Lindley & Smith, 1994). In this study, the Bayesian approach is applied to Inverse Power Lomax (IPL) model to obtain the posterior summaries of the density parameters using "Laplace's Demon" package in R software.

2. Motivation

We derived our motivation for the use of Bayesian technique based on the following:

(i) Bayesian methods allow an analyst to incorporate prior information into a data analysis/modeling problem to supplement limited data, often providing important improvements in precision (or cost savings).
(ii) Bayesian methods can handle, with relative ease, complicated data-model combinations for which no maximum likelihood (ML) software exists or for which implementing ML would be difficult. For example, available software for doing Bayesian computations can handle combinations of nonlinear relationships, random effects, and censored data that cannot be handled easily by available commercial software.
(iii) When using Bayesian method, it is very easy to obtain estimates and credible intervals for complicated functions of the model parameters such as the probability of failure or quantiles of a lifetime distribution

Sule Lamido University Journal of Science & Technology Vol. 9 No. 1&2 [May, 2024], pp. 66-73
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

According to Haddy et al. (2019), a random variable X follows the IPL distribution if its cumulative density function (CDF) is given by

$$F(t;a,b,c) = \left(1 + \frac{t^{-a}}{c}\right)^{-b}, \tag{1}$$

The corresponding probability density function (PDF) is

$$f(t;a,b,c) = \frac{ab}{c} t^{-\alpha-1}\left(1 + \frac{t^{a}}{c}\right)^{-b-1}, \tag{2}$$

The reliability and hazard rate functions of the IPL distribution are

$$R(t;a,b,c) = 1 - \left(1 + \frac{t^{-a}}{c}\right)^{-b}, \tag{3}$$

and

$$h(t;a,b,c) = \frac{\alpha\rho x^{-\alpha-1}\left(1 + \frac{t^{-a}}{c}\right)^{-b-1}}{c\left\{1 - \left(1 + \frac{t^{-a}}{c}\right)^{-b}\right\}}, \tag{4}$$

where b and c are positive shape parameters and a is a positive scale parameter. The graphs of the distribution, and density functions are given below in Figures 1 and 2.

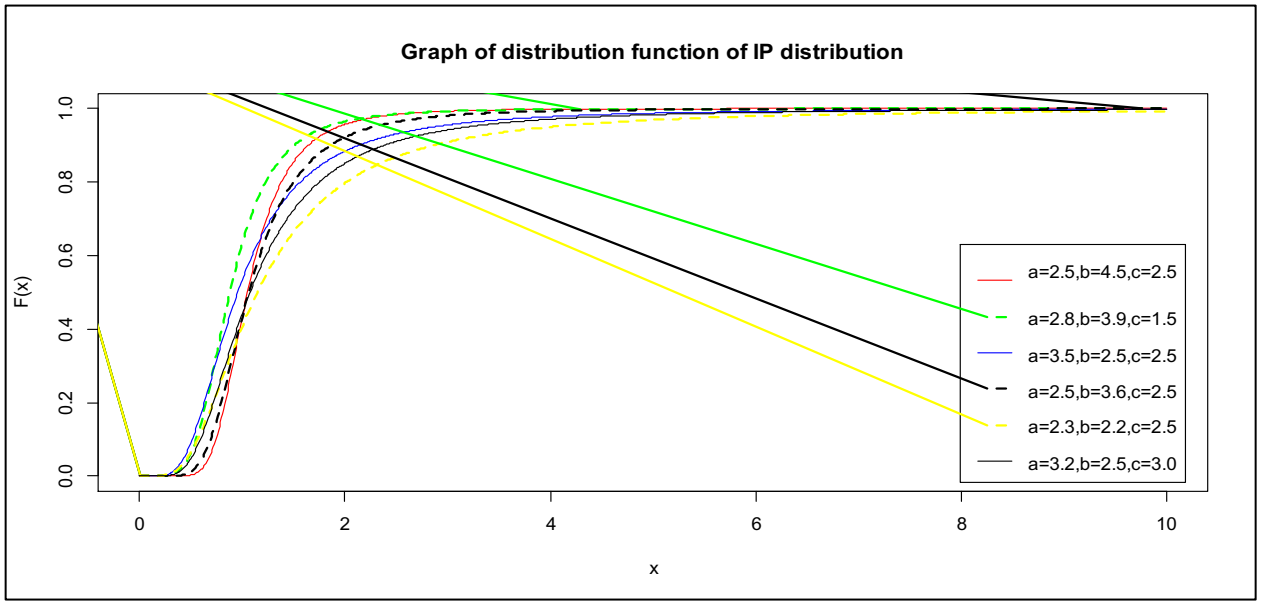

Figure 1: Graph of the distribution function of *IPL* distribution

Sule Lamido University Journal of Science & Technology Vol. 9 No. 1&2 [May, 2024], pp. 66-73
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

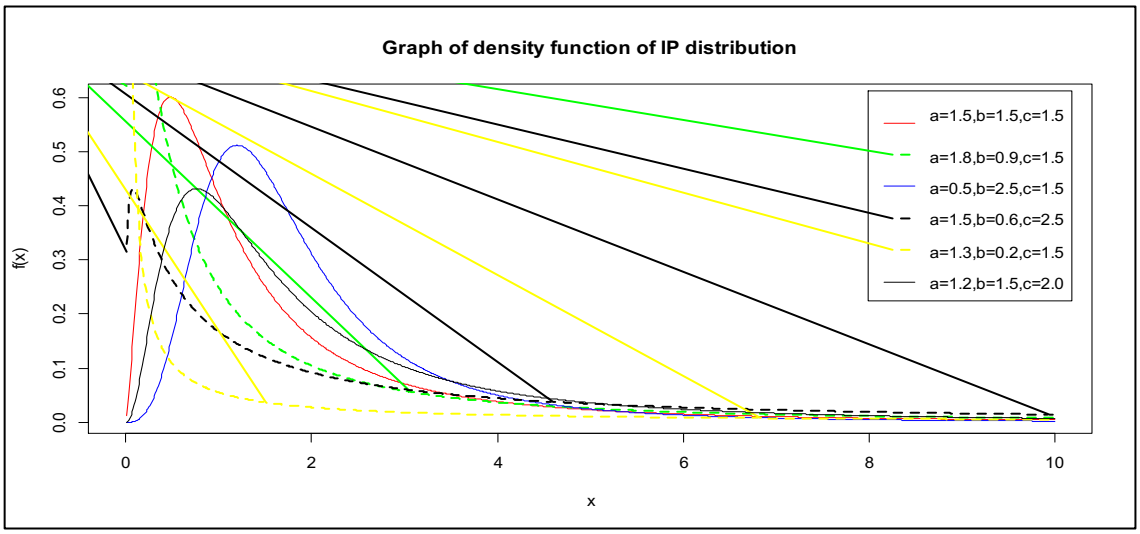

Figure 2: Graph of the distribution function of *IPL* distribution

The main objective of this study is to demonstrate the fitting of Inverse Power Lomax model in R which will be used to approximate the posterior probabilities with Laplace method and simulation in R. package for coding and solving high dimensional Bayesian analysis for modeling lifetime censored data.

3.0 Related Literature

Akhtar and Khan (2014) proposed an R package for coding and solving high dimensional Bayesian analysis models for modeling real life and censored survival data. They also established the fact that Laplace approximation is an attractive numerical approximation and will continues to develop, they suggested new methods. Khan and Khan (2013) proposed a procedure and LaplacesDemon code for solving a Laplace approximation based on Gamma model. A random censoring scheme was carried out by Liang (2004); Friesl and Hurt (2007); Abu-Taleb et al. (2007); and Saleem and Raza (2011). Liang (2004) considered the empirical Bayes estimation assuming exponentially distributed survival and censoring times with known parameter of censoring distribution. Abu-Taleb et al. (2007) derived the Bayes estimates of unknown parameters when both survival and censoring times are exponentially distributed using conjugate priors. Friesl and Hurt (2007) dealt with Bayesian estimation in exponential distribution under random censorship model and investigated the asymptotic properties of different estimators with particular stress on the Bayesian risk.

4.0 Bayesian Paradigm

In Bayesian Analysis, the three rudiments of Bayesian analysis are:
(i) Prior,
(ii) Likelihood, and
(iii) Posterior

Sule Lamido University Journal of Science & Technology Vol. 9 No. 1&2 [May, 2024], pp. 66-73
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

It should be noted that for the purpose of this study a non-informative prior work will be assumed for the purpose of estimation. Here, we assumed the data is fixed and parameters are random. Using a method developed by Tablemann and Kim (2004), a model distribution using the hazard function was employed. Assuming the hazard function at time t for an individual has the form

$$h(t/x) = h_0 e^{X^T \eta}. \tag{5}$$

Therefore, for IPL distribution taking the hazard given in equation (5) as the baseline hazard h_0, we have

$$h(t; a, b, c) = \frac{\alpha\rho x^{-\alpha-1}\left(1+\frac{t^{-a}}{c}\right)^{-b-1}}{c\left\{1-\left(1+\frac{t^{-a}}{c}\right)^{-b}\right\}} * e^{X^T\eta}, \tag{6}$$

where, $\eta = [\eta_1, \eta_2, \ldots, \eta_p]$ is a vector of regression parameters. The function $h_0(t)$ is known as the baseline hazard. It is defined as the value of the hazard function when the covariate vector $x = 0$ *or* $\eta = 0$. Equation (5) implies that the covariate act multiplicatively on the hazard rate.

The survival function of T is given by

$$S(t/x) = exp[-h(t/x)], \tag{7}$$

Plugging equation (6) in (7), we have

$$S(t/x) = exp\left[-\frac{\alpha\rho x^{-\alpha-1}\left(1+\frac{t^{-a}}{c}\right)^{-b-1}}{c\left\{1-\left(1+\frac{t^{-a}}{c}\right)^{-b}\right\}}\right] * e^{X^T\eta}, \tag{8}$$

Thus, the PDF of T given x is

$$f(t/x) = h(t/x).S(t/x). \tag{9}$$

Therefore, plugging equation (7) and (8) into (9), we have an expression for the PDF of *IPL* under Bayesian paradigm given as

$$f(t/x) = \frac{\alpha\rho x^{-\alpha-1}\left(1+\frac{t^{-a}}{c}\right)^{-b-1}}{c\left\{1-\left(1+\frac{t^{-a}}{c}\right)^{-b}\right\}} * e^{X^T\eta}.exp\left[-\frac{\alpha\rho x^{-\alpha-1}\left(1+\frac{t^{-a}}{c}\right)^{-b-1}}{c\left\{1-\left(1+\frac{t^{-a}}{c}\right)^{-b}\right\}}\right] * e^{X^T\eta}. \tag{10}$$

4.1 Likelihood function of IPL regression model with Censoring

Suppose that there are n subjects under study, and that t_i *and* t_{ci} are respectively the associated survival time and censoring time respectively. The $t's$ are assumed to be independent and identically distributed with density (t) and survival time (t). The exact survival time t_i of an individual will be observed only if, $t_i \leq t_{ci}$. The framework for the set of data for n pair of random variables (y_i, δ_i), where,

$$y_i = \min(t_i, t_{ci}) \tag{11}$$

and

SLU JST
Sule Lamido University Journal of Science & Technology Vol. 9 No. 1&2 [May, 2024], pp. 66-73
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

$$\delta_i = \begin{cases} 1 \; if \; t_i \leq t_{ci} \\ 0 \; if \; t_i > t_{ci} \end{cases} . \tag{12}$$

Then the likelihood function for $(\beta, h_0(t))$ for a set of right censored data on n subjects is given by

$$L \propto \prod_{i=1}^{n} f(y_i/x_i)^{\delta_i} . S(t_{ci}/x_i)^{1-\delta_i} . \tag{13}$$

The expression in (13) is the corresponding likelihood function when censoring is observed, right censoring and $\delta_i$is a censoring indicator variable which takes value 1 if its observed and 0 if not. To estimate the characteristics of posterior summaries of the model is actually an intricate case (Koul, et al., 1981; Lawless, 2003.). In this study we generate a solution to closed form expression for the survival function of IPL distribution and demonstrate suitability of its flexibility by Bayesian estimation of the MCMC simulation using Random-Walk. Metropolis algorithm was applied.

Putting Equations (8) and (10) in (13), we have

$$L \propto \prod_{i=1}^{n} M^{\delta_i} . \left(exp\left[- \frac{\alpha\rho x^{-\alpha-1}\left(1+\frac{t^{-a}}{c}\right)^{-b-1}}{c\left\{1-\left(1+\frac{t^{-a}}{c}\right)^{-b}\right\}} \right] * e^{X^T\eta} \right)^{1-\delta_i} , \tag{14}$$

where

$$M^{\delta_i} = \frac{\alpha\rho x^{-\alpha-1}\left(1+\frac{t^{-a}}{c}\right)^{-b-1}}{c\left\{1-\left(1+\frac{t^{-a}}{c}\right)^{-b}\right\}} * e^{X^T\eta} . exp\left[- \frac{\alpha\rho x^{-\alpha-1}\left(1+\frac{t^{-a}}{c}\right)^{-b-1}}{c\left\{1-\left(1+\frac{t^{-a}}{c}\right)^{-b}\right\}} \right] * e^{X^T\eta} .$$

4.2 Prior

Here, we set the prior to the IPL proportional hazard model whose likelihood function is given in equation (12). If $Y \sim IPL(a, b, c)$. Prior probabilities are specified for a, b, and c: a is half Cauchy (γ), b is half Cauchy $(\gamma)$and c is normal (w)

$$P(a/\gamma) = \frac{2\gamma}{\pi(a^2+\gamma^2)}, \quad a > 0 \tag{15}$$

$$P(b/\gamma) = \frac{2\gamma}{\pi(b^2+\gamma^2)}, \quad b > 0 \tag{16}$$

The half Cauchy distribution with scale parameter γ = 25 is used as non-informative prior distribution for shape parameter. As Gelman and Hill (2007) recommend that, the uniform or if more information is necessary the half-Cauchy distribution is almost flat.

4.3 Posterior

Since, c> 0 and η can take any value on the real line, hence we consider the log link function

$$logc = X^T\eta, \tag{17}$$

Then the joint posterior distribution is given by

SLUJST
Sule Lamido University Journal of Science & Technology Vol. 9 No. 1&2 [May, 2024], pp. 66-73
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

$$p(a,\ b,\ c,k/y,X) = p(y/a,b,c).p(a).p(\ b).p(\ c). \tag{18}$$

$$P(a,b,c/y,X)\sim \prod_{i=1}^{n} f(y_i/x_i)^{\delta_i}.S(t_{ci}/x_i)^{1-\delta_i}.p(a).p(\ b).p(\ c). \tag{19}$$

$$P(a,\ b,\ c/y,X)\sim \prod_{i=1}^{n} M^{\delta_i}.\left(exp\left[-\frac{\alpha\rho x^{-\alpha-1}\left(1+\frac{t^{-a}}{c}\right)^{-b-1}}{c\left\{1-\left(1+\frac{t^{-a}}{c}\right)^{-b}\right\}} \right] * e^{X^T\eta} \right)^{1-\delta_i} \times \frac{2\gamma}{\pi(a^2+\gamma^2)}.\frac{2\gamma}{\pi(b^2+\gamma^2)}.\prod_{j=1}^{p}\left\{ \frac{1}{\sqrt{2\pi}1000} e^{-\frac{\eta_j^2}{2*1000^2}} \right\}. \tag{20}$$

4.4 Laplace's Approximation

Statistic and Laplace (2003) explain that Bayesian methods depend on non-0informative prior models which provide same output with the non-Bayesian procedures. The extent to which non-informative prior is valid depends on the availability and type of data provided, sample size, and prior distribution (Tierney et al., 1989).

5.0 Results and Discussion

Multiple myeloma (MM) can be described as a neoplasm of plasma cells which affects 1 to 5 per 100,000 individuals each year globally, which is more pronounced in the West. It was found to be the second most common hematologic malignancy in the United States. The data were obtained from Krall et al. (1975) obtained from a study of 48 patients aged between 50 and 80 years. This data set have been discussed by Collet (1994; 2003) and Khan et al. (2015). As of the end of the study some patients are still alive, and so these individuals contribute right-censored survival time. The variable in the data sets is given in Table 1, as follows**:**

Table 1. Variables of interest for parameters estimation

Variable	Variable name	Object names	Coding
1	Age	x_1	Year (Integer)
2	Sex	x_2	0=female,1=male
3	Bun	x_3	Blood Urea Nitrogen
4	Ca	x_4	Serum calcium
5	Hb	x_5	Serum Haemoglobin
6	Pcells	x_6	Percentage of plasma cells
7	Protein	x_7	Ben-Jones protein (0=absent, 1=Present)

SLU JST
Sule Lamido University Journal of Science & Technology Vol. 9 No. 1&2 [May, 2024], pp. 66-73
https://doi.org/10.56471/
Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

Table 2. Posterior mean, posterior standard deviation (SD) and 95% credible region based on Laplace Approximation

Variables	Mean	SD	95% credible region
Intercept	1.214	0.2301	(0.242,1.021)
Age	0.035	0.001	(-0.024,0.012)
Sex	0.041	0.023	(-0.142,0.023)
Bun	0.011	0.013	(-0.121,-0.041)
Ca	-0.064	0.071	(-0.044,0.062)
Hb	0.172	0.019	(0.082,0.212)
Pcells	-0.043	0.020	(-0.015,0.021)
Protein	0.515	0.201	(0.112,0.841)
Shape 1	0.524	0.075	(0.132,0.615)
Shape 2	1.428	0.601	(0.402,0.772)

Table 2 shows that only three out of the seven regressor variables of multiple myeloma patients are significant. The covariates are Bur, Hb and Protein with credible regions (-0.121, -0.041), (0.082,0.212), and (0.112,0.841) respectively, and because they do not include zero, they can be taken as the appropriate regressor variable for modeling survival data. The credible region for age is (-0.024,0.012) which includes zero in it; therefore, it is not significant regressor variable for modeling. The same for sex, Ca and Pcells also are insignificant covariates.

5.1 Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)

Akaike (1974) developed and introduced a unique and vast criterion for goodness of fit measurement with certain assumptions, which includes:

(i) A parametric distribution contains a true model with proper probability density function.
(ii) The value of the estimate obtained using maximum likelihood method with the least values becomes the best model for consideration.

Akaike (1974), is given by;

$$AIC = 2p - 2ln(L) \tag{21}$$

Shwarz (1978) also introduced the Bayesian Information Criterion (BIC), given by

$$BIC = -2lnL + kln(n) \tag{22}$$

where $L =$ the likelihood function of the estimated model, $x =$ the observed data set, $n =$ the number of samples and $k =$ the number of parameters to be estimated.

Table 3. Measures of goodness of fits for the models considered.

Model	AIC	BIC
IPL	40.60	40.20
Inverse Lomax	49.10	48.10

Table 3 shows the result of comparison between inverse power Lomax model and the sub model the inverse Lomax model which indicates the inverse power Lomax model is having the smaller value which clearly indicates that it provides a better fit that the Inverse Lomax model based on the data used.

Online ISSN: 2736-0903
Print ISSN: 2736-089X
www.slujst.edu.ng

6. Conclusion

In this work, Bayesian approach has been employed via Laplace approximation to model the censored multiple myeloma survival data using the Inverse Power Lomax distribution as the baseline model. These distributions have been used as a Bayesian model to fit the survival data. This paper includes the derivation of joint and marginal posterior densities of the distribution. Asymptotic approximation and simulated posterior summary of regression model of *IPL* distribution were obtained using the Laplace Approximation algorithm. It was discovered that only three out of the seven regressor variables of multiple myeloma patients are significant. The covariates are Bur, Hb and Protein with credible regions (-0.121, -0.041), (0.082, 0.212), and (0.112, 0.841) respectively, and because they do not include zero, they can be taken as the appropriate regressor variable for modeling survival data. The credible region for age is (-0.024,0.012) which includes zero in it; therefore, it is not significant regressor variable for modeling. The same for sex, Ca and Pcells also are insignificant covariates.

References

Abu-Taleb, A. A., Smadi, M. M., & Alawneh, A. J. (2007). Bayes estimation of the lifetime parameters for the exponential distribution. *Journal of Mathematics and Statistics*, 3 (3), 106-108.

Akaike, H. (1974). A new look at the statistical model identification. Automatic Control, *IEEE Transactions on*, 19(6), 716-723.

Akhtar, M. T., & Khan, A. A. (2014). Bayesian analysis of generalized log-Burr family with R. *Springer Plus*, 3,185.

Collet, D. (1994). Modelling Survival Data in Medical Research. Chapman & Hall, London.

Collet, D. (2003). Modelling Survival Data in Medical Research, second edition. Chapman & Hall, London.

Gelman, A. and Hill, J. (2007). Data Analysis Using Regression and Multi-level/Hierarchical Models, Cambridge University Press, New York.

Friesl, M., & Hurt, J. (2007). On Bayesian estimation in an exponential distribution under random censorship. *Kybernetika*, 43, 45-60.

Khan, Y., & Khan, A. (2013). Bayesian Survival Analysis of Regression Model Using Weibull. International *Journal of Innovative Research in Science, Engineering and Technology*, 2(12), 7199-7204

Krall, J. M., Uthoff, V. A., and Harley, J. B. (1975). A setup procedure for selecting variables associated with survival. *Biometrics.* 31, 49-57.

Koul, H., Susarla, V., & Van, R. J. (1981). Regression Analysis with Randomly Right-Censored Data. *The Annals of Statistics*, 9, 1276-1288.

Liang, T. (2004). Empirical Bayes estimation with random right censoring. *International Journal of Information and Management Sciences*, 15, 1-12.

Tableman, M., and Kim, J. S. (2004). Survival Analysis using S: Analysis of Time-to-Event Data. Chapman & Hall/CRC, Boca Raton.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464. http://dx.doi.org/10.1214/aos/1176344136

Statisticat, L. L. C., & Laplace's D. (2013). Complete environment for Bayesian Inference. R Package Version, 13.03.04.

Tierney, L., Kass, R. E., & Kadane, J. B. (1989). Fully exponential Laplace approximations to expectations and variances of non-positive functions. *Journal of the American Statistical Association*, 84(407), 710-716.