International Journal of Mathematics and
Computer Science, **17**(2022), no. 3, 1351–1362

R package for the length-biased power Garima distribution to analyze lifetime data

Kittipong Klinjan, Sirinapa Aryuyuen

Department of Mathematics and Computer Science
Faculty of Science and Technolog
Rajamangala University of Technology Thanyaburi
Pathum Thani, Thailand

email: kittipong_k@rmutt.ac.th, sirinapa_a@rmutt.ac.th

(Received March 3, 2022, Accepted April 26, 2022)

Abstract

Recently, a lifetime distribution, the length-biased power Garima (LBPG) distribution, has been suggested. In this article, the application of the LBPG distribution along with using the `LBPG` package in the R language program to be analyzed the lifetime data are proposed. The model parameters are estimated using the maximum likelihood (ML) and Anderson-Darling (AD) methods. The performances of the proposed estimators are compared based on numerical calculations for various values of the distribution parameters, and sample sizes in terms of the root mean squared error and estimated values. Simulation results show that the AD method is better than the ML method. However, we found that the ML method has an efficiency close to the AD method for a large sample size. To illustrate the applicability of the LBPG distribution, we analyze three applications of various real data sets consisting of right-skewed data, left-skewed data, and symmetric data. It turns out that the LBPG distribution is better than the length-biased Garima distribution considered in this study.

Key words and phrases: Length-biased power Garima distribution, R package, `LBPG` package, power Garima distribution, lifetime data.
AMS (MOS) Subject Classifications: 62-04.
ISSN 1814-0432, 2022, http://ijmcs.future-in-tech.net

1 Introduction

Statistical models play a role in industry, commerce, medicine, science, social sciences, economics, among other fields. A statistical model is needed to describe events. The statistical model is explained under the probability distribution, and a function used to study the probability of an event of interest within the studied scope. Probability distributions and an excellent statistical model must be covered and be flexible; that is one of the fundamental issues to the development or quality improvement. Because such issues can reflect the effectiveness of developments in that matter very well, which is constantly evolving to reduce the existing restrictions or to create a new approach to be more flexible or more comprehensive and relevant to different contexts as with the issue of statistical methods, it is an important and indispensable tool in the data analysis process for all research studies. Therefore, statistical methods are constantly being developed to be comprehensive and flexible to align with the changing research problem conditions. When the recorded observation from an event cannot randomly sample from the actual distribution, this happens when the original observation is damaged, and an event occurs in a non-observable manner. In these inappropriate situations, values are reduced, and units or events do not have the same occurrences as if they follow the exact distribution. The weighted distributions are applied in various research areas related to bio-medicine, reliability, ecology, etc [1]. In addition, the weighted distribution reduces to the length-biased distribution when the weight function considers only the length of the units.

The concept of length-biased distribution is found in various applications in biomedical areas, such as family history and disease, survival analysis, intermediate events, and latency period of AIDS due to blood transfusion [2]. Recently, Klinjan and Aryuyueyn [3] proposed the length-biased distribution so-called the length-biased power Garima (LBPG) distribution. The LBPG distribution is flexible for lifetime data analysis because it has various shapes of the probability density function; i.e., left-skewed shape, right-skewed shape, and close to symmetric shape. The LBPG distribution is appropriate modeling in terms of lifetime data in many filed; i.e., engineering, science, etc. Tools to assist in data management, including data analysis, is another part that helps the operation be faster. As technology has advanced dramatically, it is necessary to develop statistical tools that are convenient and quick to use and to the point. In the development process, the choice of tools and language for development must be flexible and responsive to the theory that has been developed. The researcher will use the R language program, the most popular data science language. There are many packages and libraries provided for doing different tasks. For example, the R package for distribution, Phaphan [4] introduced the R package for the two-parameters Crack distribution.

In this paper, we show the capabilities and features of a new R package for the LBPG distribution to apply data analysis in engineering and science easily. We propose the application of the LBPG distribution along with the use of the `LBPG` package in the R language program to do an analysis of the lifetime data. We compare two methods of parameter estimation of the LBPG distribution and illustrate our work with a simulation.

2 A Brief Review of the LBPG Distribution

In this section, we give the probability functions of the LBPG distribution. Let X be a random variable distributed as the LBPG distribution with parameters λ and β, denoted by $X \sim$ LBPG (λ, β). Its probability density function (pdf) is

$$f(x;\lambda,\beta) = \frac{\lambda\beta^{1+1/\lambda}(1+\beta+\beta x^{\lambda})x^{\lambda}\exp(-\beta x^{\lambda})}{(2+1/\lambda+\beta)\,\Gamma\,(1+1/\lambda)}, x > 0, \tag{2.1}$$

where $\lambda > 0$, $\beta > 0$, and $\Gamma(\mathrm{t}) = \int_0^\infty \mathrm{s}^{\mathrm{t}-1}\exp(-\mathrm{s})d\mathrm{s}$ is the gamma function.

Its corresponding cumulative density function (cdf) is

$$F(x;\lambda,\beta) = 1 - \frac{(1+\beta)\gamma\left(1+1/\lambda, \beta x^{\lambda}\right) + \gamma\left(2+1/\lambda, \beta x^{\lambda}\right)}{(2+1/\lambda+\beta)\,\Gamma\,(1+1/\lambda)}, \tag{2.2}$$

where $\gamma\,(t,x) = \int_0^x s^{t-1}\exp(-s)ds$ is the lower incomplete gamma function, which is the result of $\gamma\,(t,x)$ by using `gamma_inc(t,x)` function based on `gsl` package in R [6]. For $\lambda = 1$, the LBPG reduces to the length biased Garima (LBG) distribution (see Klinjan and Aryuyuen [3]).

From the cdf in (2.2), $F(x;\lambda,\beta)$ of a random variable X, the quantile function Q returns a threshold value x below whose random draws from the given cdf would fall p percent of the time. In terms of the distribution function F, the quantile function Q returns the value x such that $F(x;\lambda,\beta) = p$ for $0 < p < 1$. The quantile function of the LBPG distribution is

$$Q(p;\lambda,\beta) = F^{-1}(p;\lambda,\beta), 0 \le p \le 1. \tag{2.3}$$

From the quantile function in (2.3), a random variate generation of the LBPG distribution is obtained as

$$X_i = Q(u_i; \lambda, \beta); 0 \leq u_i \leq 1, i = 1, 2, 3, \ldots, n, \tag{2.4}$$

where u_i is a value of uniform random variable on (0,1).

3 R Package for the LBPG Distribution

In 2022, Klinjan et al. [5] provided a new contributed package for R that is the `LBPG` package. The `LBPG` package requires R version 4.1.3 or higher, and it is published only on https://CRAN.R-project.org/package=LBPG. There is a function that allows users to install it directly-R program from the `devtools` package. Therefore, the step to install the `LBPG` package on R [7] or RStudio [8] is as follows:

```
install.packages("LBPG")
library(LBPG)
install.packages("gsl")
library(gsl)
```

After installing the `LBPG` package, users can use the `help(LBPG)` command to retrieve an R package manual. This new package includes the two distributions (i.e., LBPG and LBG distributions) and the corresponding calling sequences to compute the probability density function, random numbers, and distribution function for some distributions. The `LBPG` package consists of nine functions:

(1) `dLBPG(x,lambda,beta)`, the `dLBPG` function gives the pdf of the LBPG distribution as in (2.1),

(2) `pLBPG(q,lambda,beta)`, the `pLBPG` function gives the cdf of the LBPG distribution as in (2.2),

(3) `qLBPG(p,lambda,beta)`, the `qLBPG` function gives the quantile function of the LBPG distribution as in (2.3),

(4) `rLBPG(n,lambda,beta)`, the `rLBPG` function generates random numbers of the LBPG distribution as in (2.4).

In the calling sequence for using the functions `x` must be a vector of data values; `q` must be a vector of quantiles; `p` must be a vector of probabilities; `n` must be a number of observations; `lambda` must be a value of the parameter λ; `beta` must be a value of the parameter β.

4 Parameter Estimation of the LBPG Distribution

In this section, two-parameter estimation methods, the maximum likelihood (ML) and the Anderson-Darling (AD) estimators are used to estimate the parameters of the LBPG distribution.

4.1 The ML estimation for the LBPG Distribution

Let $X_1, X_2, \ldots, X_n$ be independently and identically distributed (i.i.d.) sample of size n with pdf $f(x_i; \lambda, \beta)$ as (2.1), the likelihood function can be expressed as

$$L(\lambda, \beta) = \prod_{i=1}^{n} f(x_i; \lambda, \beta) = \frac{\lambda^n \beta^{n(1+1/\lambda)} \sum_{i=1}^{n} x_i^{\lambda} \left(1 + \beta + \beta x_i^{\lambda}\right) \exp\left(-\beta x_i^{\lambda}\right)}{(2 + 1/\lambda + \beta)^n [\Gamma\left(1 + 1/\lambda\right)]^n}.$$

The log-likelihood function of the LBPG distribution can be written as the following form

$$\begin{aligned} \ell(\lambda, \beta) &= n\left(1 + 1/\lambda\right) \log \beta - n \log\left(2 + 1/\lambda + \beta\right) - n \log \Gamma\left(1 + 1/\lambda\right) \\ &\quad + n \log \lambda + \sum_{i=1}^{n} \log\left[x_i^{\lambda}\left(1 + \beta + \beta x_i^{\lambda}\right) \exp\left(-\beta x_i^{\lambda}\right)\right]. \end{aligned} \tag{4.5}$$

The components of the unit score vector are obtained by taking the partial derivatives of the $\ell(\lambda, \beta)$ with respect to parameters λ and β, and the ML estimates of parameters can be obtained by setting the score functions equal to zero, i.e.,

$$\frac{\partial \ell(\lambda, \beta)}{\partial \lambda} = 0 \text{ and } \frac{\partial \ell(\lambda, \beta)}{\partial \beta} = 0. \tag{4.6}$$

In this work, we employ the `nlm` function in the R language program and the `dLBPG` function in the `LBPG` package [5] to obtain the ML estimates of λ and β as the following.

```
#=============== ML estimator for the LBPG distribution =================#
library(LBPG)
logLBPG<-function(x,t){
  lambda<-t[1];  beta<-t[2];
  loglike<--sum(log(dLBPG(x,lambda,beta)))
```

```
  return(loglike)
} t.start<-c(lambda0,beta0)
nlm(logLBPG,x=x,t.start)
```

4.2 The AD estimation for the LBPG Distribution

Let $X_1, X_2, \ldots, X_n$ be i.i.d. sample of size n, and $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ is ordered random sample with cdf $F(x_i; \lambda, \beta)$ as (2.2). The estimators of the parameters λ and β for the LBPG distribution is obtained by minimizing the function as follows:

$$A(\lambda, \beta) = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \left\{ \log \left[F\left(x_{(i)}; \lambda, \beta\right) \right] + \log \left[1 - F\left(x_{(n+1-i)}; \lambda, \beta\right) \right] \right\}. \tag{4.7}$$

The AD estimators are calculated by solving the non-linear equations as follows;

$$\frac{\partial A(\lambda, \beta)}{\partial \lambda} = 0 \text{ and } \frac{\partial A(\lambda, \beta)}{\partial \beta} = 0. \tag{4.8}$$

In this work, we employ the `nlm` function in the R language and the `pLBPG` function in the `LBPG` package [5] to obtain the AD estimates of λ and β as the following.

```
#=============== AD estimator for the LBPG distribution ================#
library(LBPG)
ADE_LBPG<-function(x,t){
  lambda=t[1]; beta=t[2]; sx<-sort(x); r<-length(sx); L<-numeric();
  for(v in 1:r){
   L[v]<-((2*v)-1)*(log(pLBPG(sx[v],lambda,beta))
         +log(1-pLBPG(sx[r+1-v],lambda,beta)))
  }  M<--r-(1/r)*sum(L);  return(M)
}  t.start<-c(lambda0,beta0)
nlm(ADE_LBPG,x=x,t.start)
```

5 Simulation Study for the LBPG Parameter Estimation

Simulation studies are computer experiments that involve creating data by pseudo-random sampling from known probability distributions. They are an

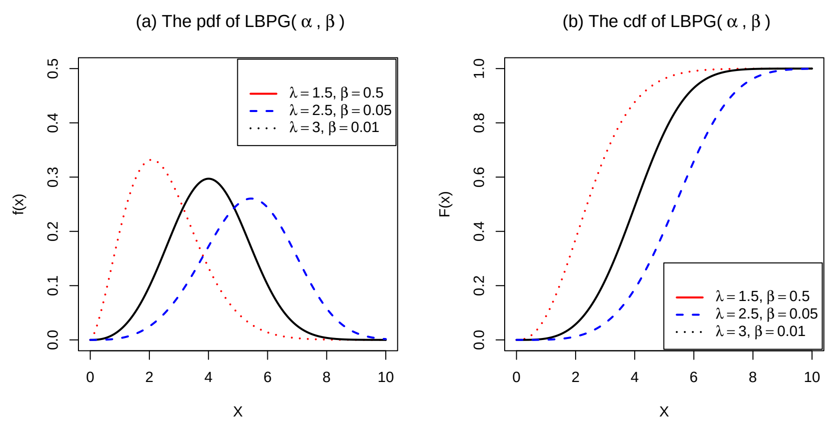

Figure 1: Plots of pdf and cdf of the LBPG distribution with the specified values of λ and β.

invaluable tool for statistical research, particularly for the evaluation of new methods and for the comparison of alternative methods [9].

In this section, we compare the performances of the proposed estimators (i.e., ML and AD methods) of the LBPG parameters λ and β. This comparison is carried out by taking random samples of different sizes (n = 20, 50, 100, 150 and 200) with various pairs of parameters values (λ, β) = (1.5,0.5), (2.5,0.05), (3,0.01), see Figure 1. Each case is considered by experiments with 1,000 replications. The estimators are compared in terms of their root mean square errors (RMSE) and the parameters' estimated values. The results are summarized in Table 1.

From Table 1, we see that:

(i) The RMSE values are decreasing as the sample sizes values are increasing for all cases considered in this section.

(ii) The estimated values of the suggested estimators decrease as the sample sizes increase and approach zero for all cases for large n, respectively.

(iii) It can be observed that the AD method gives the RMSE values less than the ML method for most cases. However, the ML method has the RMSE values close to the AD method for $n = 50, 100, 150, 200$.

6 Application on Real Data

In this section, we use data sets to illustrate the usefulness of the LBPG distribution. The first data set (Data I) is the brake pad lifetime of each car. The distribution of this data is heavily skewed to the right (skewness = 0.8655). This data set explains the lifetimes of 98 vehicles given by Lawless [10].

The second data (Data II) is the fracture toughness from the silicon nitride which is set of a left-skewed (skewness = −0.4167) data (see [11, 12]).

The third data (Data III) represents the symmetric behavior (skewness = 0.2085) of the tensile strength of about 100 observations of carbon fibers [13].

The summary of these data are provided in Table 2.

Table 1: Estimates and RMSE values of parameters of the LBPG distribution

Cases	n	Parameters	ML Estimates	ML RMSE	AD Estimates	AD RMSE
1	20	$\lambda = 1.5$	1.6248	0.3901	1.5293	0.3734
		$\beta = 0.5$	0.4862	0.2200	0.5435	0.2484
	50	$\lambda = 1.5$	1.5311	0.1424	1.5143	0.1482
		$\beta = 0.5$	0.4911	0.0957	0.5018	0.1013
	100	$\lambda = 1.5$	0.5087	0.1203	1.4992	0.1286
		$\beta = 0.5$	0.5025	0.0835	0.5090	0.0889
	150	$\lambda = 1.5$	1.5089	0.1201	1.4993	0.1284
		$\beta = 0.5$	0.5023	0.0832	0.5088	0.0884
	200	$\lambda = 1.5$	1.5089	0.0990	1.5020	0.1076
		$\beta = 0.5$	0.4999	0.0696	0.5046	0.0750
2	20	$\lambda = 2.5$	2.7508	0.6589	2.5975	0.6000
		$\beta = 0.05$	0.0512	0.0452	0.0643	0.0573
	50	$\lambda = 2.5$	2.5861	0.3701	2.5344	0.3556
		$\beta = 0.05$	0.0501	0.0265	0.0555	0.0313
	100	$\lambda = 2.5$	2.5377	0.2383	2.5086	0.2523
		$\beta = 0.05$	0.0511	0.0203	0.0540	0.0229
	150	$\lambda = 2.5$	2.5130	0.1959	2.4958	0.2089
		$\beta = 0.05$	0.0519	0.0171	0.0536	0.0190
	200	$\lambda = 2.5$	2.5202	0.1601	2.5057	0.1743
		$\beta = 0.05$	0.0502	0.0137	0.0517	0.0152
3	20	$\lambda = 3.0$	3.2797	0.7647	3.0922	0.6979
		$\beta = 0.01$	0.0112	0.0429	0.0241	0.0548
	50	$\lambda = 3.0$	3.1114	0.4220	3.0435	0.4210
		$\beta = 0.01$	0.0100	0.0273	0.0147	0.0308
	100	$\lambda = 3.0$	3.0130	0.2636	2.9847	0.2845
		$\beta = 0.01$	0.0125	0.0189	0.0150	0.0218
	150	$\lambda = 3.0$	3.0264	0.2225	3.0028	0.2349
		$\beta = 0.01$	0.0108	0.0156	0.0127	0.0174
	200	$\lambda = 3.0$	3.0078	0.1900	2.9939	0.2055
		$\beta = 0.01$	0.0113	0.0136	0.0125	0.0150

Table 2: Summary of real data sets

Data sets	n	Median	Mean	Variance	Skewness	Kurtosis
I	98	47.85	65.05	714.38	0.8655	4.1439
II	119	4.38	4.33	1.04	-0.4167	3.0935
III	87	2.74	2.63	0.98	0.2085	2.8325

Table 3: Estimates and KS values of the real data sets

		ML		AD	
Data	Statistics	LBG	LBPG	LBG	LBPG
I	$\hat{\lambda}$	-	1.8918	-	1.9758
	$\hat{\beta}$	0.0398	0.0007	0.0369	0.0005
	KS test	0.2075	0.0604	0.1730	0.0543
	(p-value)	(0.0004)	(0.8665)	(0.0057)	(0.9346)
II	$\hat{\lambda}$	-	3.8980	-	3.9838
	$\hat{\beta}$	0.5986	0.0046	0.5341	0.0040
	KS test	0.2985	0.0684	0.2643	0.0611
	(p-value)	(<0.0001)	(0.6332)	(<0.0001)	(0.7658)
III	$\hat{\lambda}$	-	2.0961	-	2.0844
	$\hat{\beta}$	0.9597	0.2295	0.8821	0.2312
	KS test	0.2024	0.0746	0.1803	0.0716
	(p-value)	(0.0016)	(0.7180)	(0.0070)	(0.7634)

The LBPG distribution is compared with the LBG distribution in Table 3. We note that the LBPG distribution gives the lowest values for the KS test and the largest p-value. Hence, the LBPG distribution could be a good alternative in explaining three data sets. For each distribution, the AD method is compared with the ML method. The AD method gives the lowest values for the KS test and the largest p-value. Hence, the AD method could be chosen as a good alternative to estimate the parameters of each distribution. However, the ML gives the values of the KS test close to the AD method. More visual comparisons where the estimated distribution functions for the four models of the two distributions and two methods of parameter estimation are provided in Figure 2.

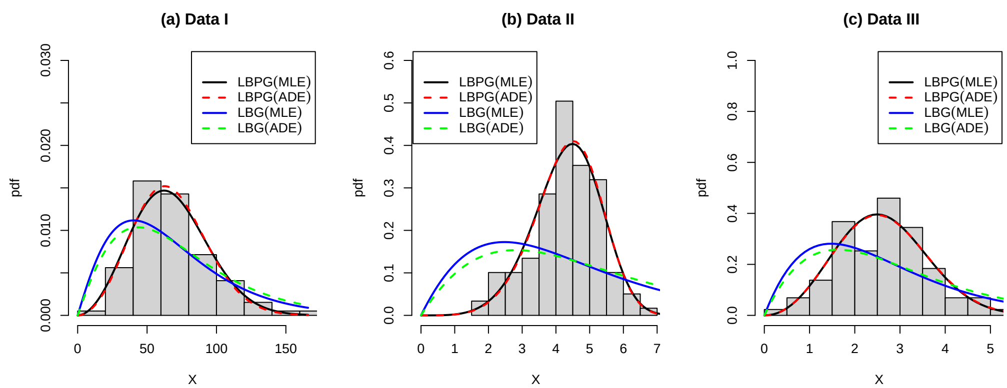

Figure 2: Plots of the estimated pdf distributions of the data sets.

7 Conclusions

In this article, we proposed the application of the LBPG distribution [3] along with using the `LBPG` package [5] in the R language program to do an analysis of the lifetime data. The `LBPG` package includes the probability density function, distribution function, quantile function, and random generation procedure for the LBPG distribution. Two methods of parameter estimation of the LBPG distribution, including the ML and AD methods, were compared. Simulation results showed that the AD method was better than the ML method. However, we found that the ML method is close to the AD method for a large sample size. Finally, a lifetime data analysis for three data sets consisting of right-skewed, left-skewed, and symmetric data was performed. As a result, the LBPG distribution was better than the LBG distribution and the AD method was more efficient than the ML method.

Acknowledgment. The author would like to thank the anonymous reviewers for their comments and suggestions.

References

[1] A. A. Rather, C. Subramanian, Length biased Sushila distribution, Universal Review, **7**, no. XII, (2018), 1010–1023.

[2] R. C. Gupta, H. O. Akman, On the reliability studies of a weighted inverse Gaussian model, J. Stat. Plan. Inference, **48**, no. 1, (1995), 69–

83.

[3] K. Klinjan, S. Aryuyuen, The length-biased power Garima distribution and its application to model lifetime data, Songklanakarin J. Sci. Tech., **43**, no. 3, (2021), 667–676.

[4] W. Phaphan, R package for the two-parameters Crack distribution, Int. J. Math. Comput. Sci., **16**, no. 4, (2021), 1455–1467.

[5] K. Klinjan, O. Thepdaeng, T. Sombunpen, A. Rattanalertnusorn, S. Aryuyuen, LBPG: The length-biased power Garima distribution, R package version 0.1.2, https://CRAN.R-project.org/package=LBPG, 2022.

[6] R. K. S. Hankin, Special functions in R: introducing the gsl package, R News, **6**, no. 4, (2006).

[7] R Core Team, R: a language and environment for statistical computing.R foundation for statistical computing, Vienna, Austria, https://R-project.org/, 2022.

[8] RStudio Team, RStudio: Integrated development for R, 2022-03-10, https://rstudio.com/, 2022.

[9] T. P. Morris, I. R. White, M. J. Crowther, Using simulation studies to evaluate statistical methods. Stat. Med., **38**, no. 11, (2019), 2074–2102.

[10] J. F. Lawless, Statistical models and methods for lifetime data 2nd Edition. John Wiley and Sons, Inc., Floboken, New Jersey, 2003.

[11] S. Nadarajah, S. Kotz, On the alternative to the Weibull function. Eng. Frac. Mech., **74**, no. 3, (2007), 451–456.

[12] M. Z. Arshad, M. Z. Iqbal, A. Al Mutairi, A comprehensive review of datasets for statistical research in probability and quality control, J. Math. Comput. Sci., **11**, no. 3, (2021), 3663–3728.

[13] M. D. Nichols, W. J. Padgett, A bootstrap control chart for Weibull percentiles, Qual. Reliab. Eng. Int., **22**, no. 2, (2006), 141–151.