Research Article

Volume 1 Issue 4 - May 2017
DOI: 10.19080/BBOAJ.2017.01.555574

Biostat Biometrics Open Acc J

Copyright © All rights are reserved by Pedro Luiz Ramos

A New Long-Term Survival Distribution

Francisco Louzada and Pedro Luiz Ramos*

Institute of Mathematical Science and Computing, University of São Paulo, Brazil

Submission: March 18, 2017; **Published:** May 15, 2017

***Corresponding author:** Pedro Luiz Ramos, Institute of Mathematical Science and Computing, University of São Paulo, Brazil,
Email: pedrolramos@usp.br

Abstract

In this paper a new long-term survival distribution named long term weighted Lindley distribution is proposed. This new model allows us to fit data where a part of the population is not susceptible to the event of interest. This model can be used in clinical studies where a portion of the population can be cured during a treatment. An account of mathematical properties of the new distribution is presented such as moments and survival properties. The maximum likelihood estimators (MLEs) for the parameters are presented. A numerical simulation is carried out in order to verify the performance of the MLEs. Finally, an important application related to the leukemia free-survival times for transplant patients are discussed to illustrate our proposed distribution.

Keywords: Long-term survival distribution; new distribution; Survival model Weighted Lindley distribution

Introduction

Advances in computational methods and numerical simulations has made possible to incorporate efficient models that are capable to be used for describing real problems. Although generalizations of the standard exponential distribution are more common, many new distributions have been proposed based on the Lindley distribution, e.g., the generalized Lindley [1], extended Lindley [2], exponential Poisson Lindley [3] and Power Lindley [4] distribution, to list a few. Another important generalized was presented by Ghitany et al. [5], such distribution is known as weighted Lindley (WL) and its probability density function (p.d.f) is given by

$$f(t; \lambda, \phi) = \frac{\lambda^{\phi+1}}{(\lambda+\phi)\Gamma(\phi)} t^{\phi-1} (1+t)e^{-\lambda t}, \quad (1)$$

for all $t > 0$, $\phi > 0$ and $\lambda > 0$ and $\Gamma(\phi) = \int_0^\infty e^{-x} x^{\phi-1} dx$ is the gamma function. Many properties of the WL were discussed earlier. Ghitany et al. [5] presented basic properties of the p.d.f, hazard function and the mean residual life function. Mazucheli et al. [6] compared different estimation procedures of the parameters of the WL distribution using numerical simulations. Wang & Wang [7] presented bias-corrected MLEs expression in which outperforms the other estimation procedures. Ali [8] considered a Bayesian approach and also presented the Bonferroni and the Lorenz curves, many entropies and order statistics for this distribution.

Recently, Ramos & Louzada [9] presented a three parameters generalized weighted Lindley distribution based on mixture of generalized gamma distribution. Although this distribution provided much flexibility in the hazard function, it does not allow us to fit data where a part of the population is not susceptible to the event of interest. For instance, in clinical studies a portion of the population can be cured during a treatment. Models that consider such characteristic are usually called long term survival models. In this study, a new long-term survival distribution is proposed, hereafter we shall call it the long term weighted Lindley distribution or simplistically the LWL distribution. Some mathematical properties for the LWL distribution is provided such as moments, survival properties and hazard function. The maximum likelihood estimators of the parameters and its asymptotic properties are discussed. We performed a numerical simulation to examine the performance of the MLEs. Finally, our proposed methodology is illustrated in a real data set related to the leukemia free-survival times (in months) for the 50 allogeneic transplant patients.

The paper is organized as follows. Section 2 reviews the weighted Lindley distribution. Section 3 presents the new long-term survival distribution. Section 4 discusses the inferential procedure based on MLEs. Section 5 displays the results of a simulation study in order to verify the

performance of the MLEs. Section 6 illustrates the relevance of our proposed methodology in a real lifetime data. Finally, Section 7 summarizes the study.

Weighted Lindley Distribution

Finite mixture models have been received special attention in order to describe hetero-geneous survival data. The chosen model can be expressed as a mixture model given by

$$f(t|\lambda, \varphi) = pf_1(t|\lambda, \varphi) + (1-p)f_2(t|\lambda, \varphi),$$

where $p = \lambda/(\lambda + \varphi)$ and $f_j(t|\lambda, \varphi)$, $j=1, 2$ has p.d.f Gamma $(\varphi+j-1, \lambda)$ distribution. Many mathematical results can be derived using the proposed structure. For instance, the mean and variance of the WL distribution can be easily computed by

$$\mu = \frac{\phi(\lambda + \phi + 1)}{\lambda(\lambda + \phi)} \quad \text{and} \quad \sigma^2 = \frac{(\phi + 1)(\lambda + \phi)^2 - \lambda^2}{\lambda^2(\lambda + \phi)^2}$$

Using this property Ghitany et al. [5] derived the survival function of T given by

$$S(t|\lambda, \phi) = \frac{(\lambda + \phi)\Gamma(\phi, \lambda t) + (\lambda t)^\phi e^{-\lambda t}}{(\lambda + \phi)\Gamma(\phi)}$$

where $\Gamma(\phi, \lambda t)$ is the upper incomplete gamma. The hazard function of the WL distribution is given by

$$h(t|\lambda, \phi) = \frac{\lambda^{\phi+1} t^{\phi-1} (1+t) e^{-\lambda t}}{(\lambda + \phi)(\phi, \lambda t) + (\lambda t)^\phi e^{-\lambda t}}. \quad (2)$$

The authors proved that the hazard rate function (2) is bathtub (increasing) shaped if $0 < \varphi < 1$ ($\varphi \geq 1$) for all $\lambda > 0$. Figure 1 shows some examples from its shapes under different values of φ and λ .

Figure 1: Hazard function of the WL distribution for different values of φ and λ .

Cure Rate Model

An important aspect to be considered is the presence of long-term survivors where a portion of the population may be not susceptible to the event of interest [Maller and Zhou, [10]; or Perdoná & Louzada [11]]. In this case, we assume that the population is divided in two groups: a not susceptible to the event of interest with probability π and the other susceptible (in risk) to the event with probability $(1-\pi)$. The long-term survivor is expressed as

$$S(t) = \pi + (1-\pi)S_0(t), \quad (3)$$

where $\pi \in (0,1)$ and $S_0(t)$ is the survival function related to the susceptible group. The obtained survival function (not conditional) is improper and its limit corresponds to the individual proportion cure. From the survival function one can easily derive the p.d.f. given by

$$f(t) = -\frac{d}{dt} S_{pop}(t) = (1-\pi)f_0(t), \quad (4)$$

where $f_0(t)$ is the p.d.f. related to the susceptible group. Thus, from (3) and (4) the hazard function is given by

$$h(t) = \frac{(1-\pi)f_0(t)}{\pi + (1-\pi)S_0(t)}. \quad (5)$$

Considering that $f_0(t)$ follows a WL distribution, then the p.d.f. of the LWL distribution is given by

$$f(t|\theta) = \frac{(1-\pi)(\lambda + \phi + r)\Gamma(\phi + r)}{\lambda^r(\lambda + \phi)\Gamma(\phi)} \quad (6)$$

where $\theta = c(\lambda, \varphi, \pi)$ is the vector of parameters and $\lambda > 0$, $\varphi > 0$ and $\pi \in (0,1)$. The r th moments of T about the origin is

$$E(T^r) = \frac{(1-\pi)(\lambda + \phi + r)\Gamma(\phi + r)}{\lambda^r(\lambda + \phi)\Gamma(\phi)}$$

Therefore, the mean and variance of the LWL distribution can be easily computed

by

$$\mu = \frac{(1-\pi)\phi(\lambda + \phi + 1)}{\lambda(\lambda + \phi)}$$

$$\sigma^2 = \frac{\phi(1-\pi)(\lambda^2(\phi\pi + 1) + 2\lambda(\phi^2\pi + \phi(\pi + 1) + 1) + \phi((\phi^2 + 1)\pi + \phi(2\pi + 1) + 1))}{\lambda^2(\lambda + \phi)^2}$$

The survival function and the hazard function of the LWL distribution is given by

$$L(\theta; D) = \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(t_i; \theta)^{1-\delta_i}. \quad (7)$$

$$L(\theta; D) = \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(t_i; \theta)^{1-\delta_i}. \quad (8)$$

Figure 2 displays some examples from the shapes of the p.d.f. and the survival function of the LWL distribution

Figure 2: Left panel: probability density function of the LWL distribution. Right panel: survival functions of the LWL distribution.

Inference

Let us consider the case where the failure time of the i^{th} individual may not be observed and is subject to right censoring. Additionally, the random censoring times C_{is} are independent of T_{is} (non censored time) and their distribution does not depend on the parameters. In a sample of size n , the

data set is $D = (t_i, \delta_i)$, where $t_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$. This random censoring scheme has as special case type I and II censoring mechanism. The likelihood function is given by

$$L(\theta; D) = \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(t_i; \theta)^{1-\delta_i}.$$

Let $T_1, \dots, T_n$ be a random sample of LWL distribution, the likelihood function considering data with random censoring is given by

$$L(\theta; D) = \frac{(1-\pi)^d \lambda^{d(\phi+1)}}{(\lambda+\phi)^n \Gamma(\phi)^n} \prod_{i=1}^n (t_i^{\phi-1} (1+t_i) e^{-\lambda t_i})^{\delta_i} \times ((\pi(\lambda+\phi)\Gamma(\phi) + (1-\pi)((\lambda+\phi)\Gamma(\phi, \lambda t) + (\lambda t)^{\phi} e^{-\lambda t}))^{1-\delta_i}.$$

The log-likelihood function is given as

$$l(\theta; D) = d \log(1-\pi) + d(\phi+1) \log(\lambda) - n \log(\Gamma(\phi)) + \sum_{i=1}^n ((1-\delta_i) \log((\pi(\lambda+\phi)\Gamma(\phi) + (1-\pi)((\lambda+\phi)\Gamma(\phi, \lambda t) + (\lambda t)^{\phi} e^{-\lambda t}))) + \sum_{i=1}^n \delta_i \log(t_i^{\phi-1} (1+t_i) e^{-\lambda t_i})).$$

The maximum likelihood estimation method is widely used in statistical inference. It has many desirable properties, including consistency, asymptotic efficiency and in-variance. The maximum likelihood estimators are obtained from the maximization of the log-likelihood function (10). Here, we performed directly the maximization of log-likelihood function using existing statistical packages. Further information about the numerical procedures will be discussed in the next section.

Under mild conditions the obtained estimators are consistent and efficient with an asymptotically normal joint distribution given by

$$\hat{\theta} \sim N_3[\theta, I^{-1}(\theta)] \text{ as } n \rightarrow \infty,$$

where $I(\theta)$ is the Fisher information matrix, 3×3 and $I_{ij}(\theta)$, is the Fisher information of θ in i and j given by

$$I_{ij}(\theta) = E \left[-\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\theta; D) \right], i, j = 1, 2, 3.$$

Since it is not possible to compute the Fisher information matrix, an alternative is to consider the observed information matrix, where the terms is given by

$$H_{ij}(\theta) = -\frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\theta; D), i, j = 1, 2, 3.$$

For large samples, approximated confidence intervals can be constructed for the individual parameters θ_i $i = 1, \dots, 3$, assuming a confidence coefficient $100(1 - \gamma)\%$, the marginal distributions is given by

$$\hat{\theta}_i \sim N[\theta_i, H_{ii}^{-1}(\theta)] \text{ for } n \rightarrow \infty \quad i = 1, 2, 3.$$

Table 1: MREs, MSEs, C95% estimates for 10000 considering $n = (10, 25, 50, 100, 200)$ and 30% of censorship.

	$\phi = 0.5$			$\lambda = 2$			$\pi = 0.3$			0.3
n	MRE	MSE	C95%	MRE	MSE	C95%	MRE	MSE	C95%	Mp
10	1.518	0.34	0.981	1.698	7.777	0.988	1.011	0.018	0.87	0.306
25	1.158	0.049	0.962	1.228	1.306	0.963	1.001	0.008	0.943	0.303
50	1.07	0.015	0.958	1.093	0.372	0.955	1.002	0.004	0.938	0.303
100	1.032	0.006	0.955	1.041	0.144	0.956	1	0.002	0.943	0.302
200	1.014	0.003	0.953	1.02	0.063	0.955	1.001	0.001	0.945	0.303

Simulation Study

In this section a simulation study is presented to compare the efficiency of the maximum likelihood method. This procedure is conducted by computing the mean relative errors (MRE) and the mean square errors (MSE) given by

$$MRE_i = \frac{1}{N} \sum_{j=1}^N \frac{\hat{\theta}_{i,j}}{\theta_i}, \quad MSE_i = \frac{1}{N} \sum_{j=1}^N (\hat{\theta}_{i,j} - \theta_i)^2, \quad \text{for } i=1, 2, 3.$$

where N is the number of estimates obtained through the MLE approach. The 95% coverage probabilities of the asymptotic confidence intervals are also evaluated. From our approach, it is expected that the MLEs returns the MREs closer to one with smaller RMSEs. Additionally, considering a 95% confidence level, the frequencies of intervals that covered the true values of θ should be closer to 95%.

The simulation study is performed considering the values: $\theta = ((0.5, 2, 0.3), (0.5, 2, 0.6), (3, 2, 0.3), (3, 2, 0.6))$, $N = 10,000$ and $n = (10, 25, 50, 100, 200)$. We have considered two scenarios where we obtained approximately 0.3 and 0.6 proportions of censored data, i.e., 30% and 60% of censorship. Since the censoring is random, we may observe different proportions of censoring for each data set. However, considering this approach we expected that the mean of the proportions of censored observations M_p return the nominal level assumed. It is worth mentioning that, the results of this simulation study were similar for different choices of θ . The maxLik package available [12] in R was considered to find the maximization of the log-likelihood function [10]. This estimation procedure is well-behaved as we have not faced numerical problems, such as evidence of failure of convergence or end on multiple maxima. The programs can be obtained, upon request. Tables 1-4 display the MREs, MSEs, and the coverage probability with a 95% confidence level (C95%) of the estimates obtained through the MLE. From Tables 1-4, we observed that the estimates of λ , ϕ and π are asymptotically unbiased, i.e., the MREs tend to one when n increases and the MSEs decrease to zero for n large. Additionally, the coverage probabilities tend to 0.95, i.e., the MLEs has good coverage properties for the parameter estimators. Therefore, this estimation procedure can be used in practical applications as will be shown in the next section.

Table 2: MREs, MSEs, C95% estimates for 10000 considering n = (10,25,50,100,200) and 60% of censorship.

	$\phi = 0.5$			$\lambda = 2$			$\pi = 0.6$			0.6
n	MRE	MSE	C95%	MRE	MSE	C95%	MRE	MSE	C95%	Mp
10	1.922	1.002	0.986	2.063	14.19	0.994	0.935	0.021	0.931	0.563
25	1.342	0.168	0.974	1.454	3.913	0.978	0.997	0.009	0.942	0.599
50	1.134	0.038	0.959	1.19	0.997	0.964	1.000	0.005	0.938	0.602
100	1.059	0.013	0.96	1.079	0.307	0.956	1.000	0.002	0.947	0.602
200	1.03	0.005	0.952	1.04	0.127	0.953	1.000	0.001	0.95	0.601

Table 3: MREs, MSEs, C95% estimates for 10000 considering n = (10,25,50,100,200) and 30% of censorship.

	$\phi = 0.5$			$\lambda = 2$			$\pi = 0.3$			0.3
n	MRE	MSE	C95%	MRE	MSE	C95%	MRE	MSE	C95%	Mp
10	1.49	6.235	0.987	1.571	2.86	0.987	0.977	0.017	0.872	0.299
25	1.295	2.53	0.985	1.319	1.008	0.987	0.998	0.008	0.947	0.305
50	1.165	1.033	0.974	1.178	0.386	0.978	1.000	0.004	0.933	0.306
100	1.079	0.369	0.97	1.086	0.132	0.981	1.000	0.002	0.946	0.306
200	1.037	0.158	0.964	1.04	0.055	0.974	0.999	0.001	0.951	0.306

Table 4: MREs, MSEs, C95% estimates for 10000 considering n = (10,25,50,100,200) and 60% of censorship.

	$\phi = 0.5$			$\lambda = 2$			$\pi = 0.6$			0.6
n	MRE	MSE	C95%	MRE	MSE	C95%	MRE	MSE	C95%	Mp
10	1.509	7.187	0.983	1.675	4.046	0.984	0.906	0.022	0.916	0.547
25	1.373	3.815	0.984	1.439	1.727	0.985	0.98	0.009	0.947	0.591
50	1.258	1.962	0.982	1.286	0.79	0.984	0.995	0.005	0.946	0.6
100	1.143	0.786	0.977	1.159	0.301	0.981	0.999	0.002	0.949	0.603
200	1.076	0.319	0.968	1.085	0.117	0.976	0.997	0.001	0.948	0.602

Cancer Data Application

In this section, we considered the data set presented by Ghitany & Al-Awadhi [13]. The data describes the leukemia

free-survival times (in months) for the 50 allogeneic transplant patients. Table 5 shows the full data set (+ indicates censored observations).

Table 5: Leukemia free-survival times (in months) for the 50 allogeneic transplant patients (+ indicates censored observations).

0.03	0.493	0.855	1.184	1.1283	1.48	1.776	2.138
2.5	2.763	2.993	3.324	3.421	4.178	4.441+	5.691
5.855+	6.941	6.941+	7.993+	8.882	8.882	9.145+	11.48
11.513	12.105+	12.796	12.933+	13.849+	16.612+	17.138+	20.066
20.329+	22.368+	26.766+	28.717+	28.717+	32.928+	33.783+	34.221+
34.770+	39.539+	41.118+	45.033+	46.033+	46.941+	48.289+	57.041+
58.322+	60.625+						

The main authors considered the Burr XII distribution to fit the data under random censoring scheme. Although the maximum likelihood estimators were presented success-fully, the data clearly indicates the presence of a not susceptible group to the event of interest. Therefore, long-term survival distributions are more suitable to describe the proposed data set. Having discovered that, we considered the proposed LWL distribution to fit the data. However, firstly we would like to verify the behavior of the empirical hazard function. This can be achieved considering the TTT-plot (total time on test) proposed by Barlow and Campo [3]. The TTT-plot is obtained

through the consecutive plot of the values $[r/n, G(r/n)]$ where

$$G(r/n) = \left(\sum_{i=1}^r t_i + (n-r)t_{(r)} \right) / \sum_{i=1}^n t_i, \quad r = 1, \dots, n, \quad i = 1, \dots, n$$

and t_i is the order statistics. In case of the curve is concave (convex), the hazard function is increasing (decreasing), on the other hand if it's starts convex and then concave (concave and then convex) the hazard function has bathtub (inverse bathtub) shape.

Figure 3 presents the TTT-plot [14], the empirical survival function adjusted by the Kaplan-Meier estimator and the LWL survival distribution. Based on the TTT-plot there is an

indication that the hazard function has bathtub failure rate. Comparing the empirical survival function with the adjusted by the LWL distribution we observed a goodness of the fit for our proposed model. The MLEs were obtained using the same procedure described in Section 5. Table 6 displays the MLEs, the standard error (SE) and the confidence intervals under 95% confidence level for ϕ , λ . Therefore, through our proposed methodology the data related to the leukemia free-survival times (in months) for the 50 allogeneic transplant patients can be described by the LWL distribution.

Figure 3: TTT-plot and survival function adjusted by the empirical survival function (Kaplan-Meier estimator) and the LWL distribution.

Table 6: MLE, SE, and confidence intervals under 95% confidence level for ϕ , λ and π .

θ	MLE	SE	CI95%(θ)
ϕ	0.6066	0.2142	(0.1868; 1.0263)
λ	0.2317	0.065	(0.1041; 0.3593)
π	0.5305	0.0749	(0.3837; 0.6773)

Discussion

In this paper, we have proposed a new long-term survival distribution, the so called long term weighted Lindley distribution and its mathematical properties were studied. The maximum likelihood estimators of the parameters and their asymptotic properties were discussed. The simulation study showed that the MLEs present efficient estimators for three unknown parameters even for small sample sizes. The practical importance of the LWL distribution was reported using an important data set related to the leukemia free-survival times (in months) for the 50 allogeneic transplant patients. Many extensions of the present work can be considered for instance, the inclusion of covariates x in the long-term survival term using the logistic link function given by $\pi(x) = \exp(x\beta) / [1 + \exp(x\beta)]$, where β is the vector of parameters related to the covariates. Other approach that

can be considered is the presence of covariates under the assumption of Cox model, i.e., the proportional hazards. Finally, our analysis can be considered using the Bayesian approach.

Acknowledgement

The authors are thankful to the Editorial Board and to the reviewers for their valuable comments and suggestions. The research was partially supported by CNPq, FAPESP and CAPES of Brazil.

References

1. Zakerzadeh H, Dolati A (2009) Generalized lindley distribution. Journal of Mathematical extension 3: 13-25.
2. Bakouch HS, Al-Zahrani BM, Al-Shomrani AA, Marchi VA, Louzada F (2012) An extended lindley distribution. Journal of the Korean Statistical Society 41: 75-85.
3. Barreto-Souza W, Bakouch HS (2013) A new lifetime model with decreasing failure rate. Statistics 47(2): 465-476.
4. Ghitany M, Al-Mutairi D, Balakrishnan N, Al-Enezi L (2013) Power lindley distribution and associated inference, Computational Statistics & Data Analysis 64: 20-33.
5. Ghitany M, Alqallaf F, Al-Mutairi D, Husain H (2011) A two-parameter weighted lindley distribution and its applications to survival data. Mathematics and Computers in Simulation 81: 1190-1201.
6. Mazucheli J, Louzada F, Ghitany M (2013) Comparison of estimation methods for the parameters of the weighted lindley distribution. Applied Mathematics and Computation 220: 463-471.
7. Wang M, Wang W (2017) Bias-corrected maximum likelihood estimation of the parameters of the weighted lindley distribution. Communications in Statistics - Simulation and Computation 46(1): 530-545.
8. Ali S (2015) On the bayesian estimation of the weighted lindley distribution. Journal of Statistical Computation and Simulation 85(5): 855-880.
9. Ramos PL, Louzada F (2016) The generalized weighted lindley distribution: Properties, estimation and applications. Cogent Mathematics 3(1): 1256022.
10. Maller RA, Zhou S (1995) Testing for the presence of immune or cured individuals in censored survival data. Biometrics pp. 51(4): 1197-1205.
11. Perdoná GC, Louzada-Neto F (2011) A general hazard model for lifetime data in the presence of cure rate. Journal of Applied Statistics 38(7): 1395-1405.
12. Henningsen A, Toomet O (2011) maxlik: A package for maximum likelihood estimation in r. Computational Statistics 26(3): 443-458.
13. Ghitany M, Al-Awadhi S (2002) Maximum likelihood estimation of burr xii distribution parameters under random censoring. Journal of Applied Statistics 29(7): 955-965.
14. Barlow RE, Campo RA (1975) Total time on test processes and applications to failure data analysis. Tech Rep DTIC Document.

This work is licensed under Creative Commons Attribution 4.0 License

Your next submission with Juniper Publishers

will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission

<https://juniperpublishers.com/online-submission.php>