

Chi-squared type test for Gamma-Lindley distribution

Abderrahmane Bensid, Halim Zeghdoudi and Azzedine Benchettah

Abstract. The Gamma Lindley distribution [19, 20], of which the Lindley distribution is a particular case, it proposed because the Lindley distribution has only one parameter, which is not flexible for analyzing and modeling different types of lifetime data and survival analysis. In this article, we present a modified Chi-squared goodness of fit for the Gamma Lindley distribution, for censored data. Based on Nikuline-Rao-Robson statistics Y^2 , proposed by Bagdanovicius and Nikuline (2011) for censored data. An application of this distribution in survival analysis is given to discussing the potential of the presented test. **keywords**Lindley distribution;Gamma Lindley distribution; Censored data; Maximum likelihood estimation; NRR statistic

M.S.C. 2010: 62P05, 62E15, 62F03, 62G05.

Key words: statistical analysis; gamma-Lindley distribution; chi-squared test; maximum likelihood estimators.

1 Introduction

The statistical analysis and modeling of lifetime data are essential in almost all applied sciences, a number of one parameter continuous distributions for modeling lifetime data has been introduced, one of them called the Lindley distribution. It was introduced by Lindley (1958), Sankaran (1970) used the Lindley distribution as mixing distribution of Poisson parameter, Asgharzadeh et al. (2013), Ghitany et al.(2008 a) and (2008 b) rediscovered and studied the new distribution bounded to the Lindley distribution. In this article we focus on the Gamma Lindley distribution, introduced by Zeghdoudi and Nedjar (2016a,2016b), of which the Lindley distribution is a particular case, it is a mixture of Lindley and gamma distributions. In addition, in the recent research of Nikulin et al. ([2], [29], [30]) considered these applications of this model in the accelerated lifetimes (AFT) models and redundant systems. Nowadays, the BS distribution has known as cumulative damage distributions and it is a very useful in fatigue, reliability and survival analysis.

The chi-square test (Snedecor and Cochran, 1989) is used to test if a sample of data came from a population with a specific distribution. An attractive feature of

the chi-square goodness-of-fit test is that it can be applied to any univariate distribution for which you can calculate the cumulative distribution function. The chi-square goodness-of-fit test is applied to binned data (i.e., data put into classes). This is actually not a restriction since for non-binned data you can simply calculate a histogram or frequency table before generating the chi-square test. However, the value of the chi-square test statistic is dependent on how the data is binned. Another disadvantage of the chi-square test is that it requires a sufficient sample size in order for the chi-square approximation to be valid.

According to Bagdonavičius and Nikulin (2011), Bagdonavičius et al. (2011), Bagdonavičius et al. (2010) and Nikulin et al. (2011), we describe a chi-squared test for testing composite parametric hypotheses when data are right censored. This problem arises naturally in reliability and survival analysis. Here, it will be the first time that someone considers the test for the gamma Lindley distribution or any other Lindley distribution.

The paper is organized as follows: Section 2 is devoted to introduce and give some properties of the gamma-Lindley distribution. In section 3 we give the maximum likelihood estimators of this distribution in the case of censored data. In section 4 we give the construction of modified chi-squared goodness of fit test in both of general case and gamma Lindley distribution. Finally, we use the approach of Nikulin-Rao-Robson (NRR), based on maximum likelihood estimators, and give an illustrative example for a show the applicability of this test.

2 Gamma Lindley distribution (GaLD)

In this section, we give the density, the cumulative distribution, the survival, the hazard, and the cumulative hazard functions of the new distribution named Gamma Lindley distribution introduced by Zeghdoudi and Nedjar (2016a, 2016b) .

Suppose that T is random variable taking values in $]0, \infty)$, where $T \rightsquigarrow \text{GaLD}$. The density function of T is given by

$$(2.1) \quad f(t) = \frac{\theta^2 ((\beta + \theta\beta - \theta)t + 1) \exp(-\theta t)}{\beta(1 + \theta)}; t > 0, \beta > \frac{\theta}{1 + \theta}, \theta > 0$$

Hence, the cumulative distribution function of the *GaLD* distribution with parameters θ and β , is given by

$$(2.2) \quad F(t) = 1 - \frac{((\beta + \theta\beta - \theta)(\theta t + 1) + \theta) \exp(-\theta t)}{\beta(1 + \theta)}; t > 0, \beta > \frac{\theta}{1 + \theta}, \theta > 0.$$

Then, the corresponding survival, hazard, and cumulative hazard functions are

$$(2.3) \quad S(t) = \frac{((\beta + \theta\beta - \theta)(\theta t + 1) + \theta) \exp(-\theta t)}{\beta(1 + \theta)},$$

$$(2.4) \quad \lambda(t) = -\frac{S'(t)}{S(t)} = \frac{((\beta + \theta\beta - \theta)t + 1)\theta^2}{\theta(\beta + \theta\beta - \theta)t + \beta + \theta\beta},$$

$$(2.5) \quad \Lambda(t) = -\ln S(t) = \theta t + \ln(\beta(1 + \theta)) - \ln((\beta + \theta\beta - \theta)(\theta t + 1) + \theta),$$

respectively.

3 Maximum likelihood estimation with censored data

Let $X_i, i = 1, \dots, n$, be independent and identically distributed (*i.i.d*) random variables with common continuous distribution function F , and let $C_i, i = 1, \dots, n$, be independent and identically distributed sequence with continuous distribution function G .

suppose, instead of the sample X_1, \dots, X_n , we observe the right censored simple

$$(T_1, \delta_1), \dots, (T_n, \delta_n).$$

where

$$T_i = X_i \wedge C_i, \delta_i = 1_{\{X_i \leq C_i\}}, i = 1, \dots, n.$$

Let consider the distribution of the random vector (T_i, δ_i) in the case of random censoring with absolutely continuous censoring times C_i with density function $g_i(x)$. In this case, we have the likelihood function as

$$L(t, \theta) = \prod_{i=1}^n f^{\delta_i}(t_i, \theta) S^{1-\delta_i}(t_i, \theta) \bar{G}^{\delta_i}(t_i) g^{1-\delta_i}(t_i).$$

where and \bar{G}_i is the survival function of the censoring time C_i .

Now we suppose that censoring is non-informative, So the members with \bar{G}_i and g_i do not contain θ , so they can be rejected. The likelihood function is

$$L(t, \theta) = \prod_{i=1}^n f^{\delta_i}(t_i, \theta) S^{1-\delta_i}(t_i, \theta) = \prod_{i=1}^n \lambda^{\delta_i}(t_i, \theta) S(t_i, \theta), \delta_i = 1_{\{T_i \leq C_i\}}$$

In our case, we have

$$(3.1) \quad L(t, \theta) = \prod_{i=1}^n \left(\left(\frac{(\beta + \theta\beta - \theta)t_i + 1}{\theta(\beta + \theta\beta - \theta)t_i + \beta + \theta\beta} \theta^2 \right)^{\delta_i} \frac{(\beta + \theta\beta - \theta)(\theta t_i + 1) + \theta}{\beta(1 + \theta)} \right) \exp\left(-\theta \sum_{i=1}^n t_i\right).$$

The ln-likelihood function is

$$(3.2) \quad \begin{aligned} \ln L(t, \theta) &= \sum_{i=1}^n (\delta_i \ln((\beta + \theta\beta - \theta)t_i + 1) + 2 \ln \theta \delta_i - \delta_i \ln((\beta + \theta\beta - \theta)t_i + \beta + \theta\beta)) \\ &\quad + \sum_{i=1}^n (\ln((\beta + \theta\beta - \theta)(\theta t_i + 1) + \theta) - \theta t_i - n \ln \beta (1 + \theta)). \end{aligned}$$

The score functions for the parameters θ and β are given by

$$(3.3) \quad \begin{aligned} \frac{\partial l(t_i; \theta, \beta)}{\partial \theta} &= \sum_{i=1}^n \left(\delta_i \frac{(\beta - 1)t_i}{(\beta + \theta\beta - \theta)t_i + 1} + \frac{2}{\theta} \delta_i - \frac{(\beta + \theta\beta - 1)t_i + \beta}{\theta(\beta + \theta\beta - \theta)t_i + \beta + \theta\beta} \right) \\ &\quad - \sum_{i=1}^n \left(t_i + \frac{(\beta - 1)(\theta t_i + 1) + (\beta + \theta\beta - \theta)t_i + 1}{(\beta + \theta\beta - \theta)(\theta t_i + 1)} - \frac{1}{1 + \theta} \right), \end{aligned}$$

$$(3.4) \quad \frac{\partial l(t_i; \theta, \beta)}{\partial \beta} = \sum_{i=1}^n \left(\delta_i \frac{(1 + \theta)t_i}{(\beta + \theta\beta - \theta)t_i + 1} - \delta_i \frac{\theta((1 + \theta)t_i + 1) + 1}{\theta(\beta + \theta\beta - \theta)t_i + \beta + \theta\beta} + \delta_i \frac{(\theta t_i + 1)(1 + \theta)}{(\beta + \theta\beta - \theta)(\theta t_i + 1) + \theta} - \frac{1}{\beta} \right).$$

4 Chi-squared type test for right censored data

Let be consider next the hypothesis

$$H_0 : F(x) \in \mathcal{F}_0 = \{F_0(x; \theta), \theta \in \Theta \subseteq \mathbb{R}^d\},$$

where $\theta = (\theta_1, \theta_2, \dots, \theta_d)^T$ are unknown d dimensional parameters and F_0 is a known distribution function. The idea is to divide the time interval $[0, \tau]$ into $k > d$ smaller intervals $I_j = (a_{j-1}, a_j]$ with equal expected numbers of failures e_j , where $a_0 = 0$ and $a_k = \tau$.

Here, the estimated a_j is given by

$$\hat{a}_j = \Lambda^{-1} \left(\frac{E_j - \sum_{l=1}^{i-1} \Lambda(X_{(l)}, \hat{\theta})}{n - i + 1}, \hat{\theta} \right), \hat{a}_k = X_{(n)}, j = 1, \dots, k,$$

where $\hat{\theta}$ is the maximum likelihood estimator of the parameter θ , Λ^{-1} is the inverse of the cumulative hazard function Λ , $X_{(i)}$ is the i th element in the ordered statistics $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ and $E_j = (n - i + 1) \Lambda(\hat{a}, \hat{\theta}) + \sum_{l=1}^{i-1} \Lambda(X_{(l)}, \hat{\theta})$ is random data functions.

A chi-squared test which was proposed by Bagdanovicius and Nikulin (2009), based on the vector

$$(4.1) \quad Z = (Z_1, Z_2, \dots, Z_k)^T, Z_j = \frac{1}{\sqrt{n}} (U_j - e_j), j = 1, 2, \dots, k.$$

where U_j represent the numbers of observed failures in these intervals.

The statistic of Bagdanovicius and Nikulin given as

$$Y^2 = Z^T \hat{\Sigma}^- Z,$$

where $\hat{\Sigma}^-$ is the general inverse matrix of the covariance matrix $\hat{\Sigma}$,

$$\begin{aligned} \hat{\Sigma} &= \hat{A} - \hat{C}^T \hat{I}^{-1} \hat{C}, \\ \hat{\Sigma}^{-1} &= \hat{A}^{-1} + \hat{A}^{-1} \hat{C}^T \hat{G}^{-1} \hat{C} \hat{A}^{-1}, \\ \hat{G} &= \hat{I} - \hat{C} \hat{A}^{-1} \hat{C}^T, \end{aligned}$$

where \hat{A} is the diagonal $k \times k$ matrix with elements $A_j = \frac{U_j}{n}$ on the diagonal, \hat{A}^{-1} is inverse matrix of \hat{A} , and

$$(4.2) \quad \hat{C} = [\hat{C}_{lj}]_{d \times k}, \text{ with } \hat{C}_{lj} = \frac{1}{n} \sum_{i: x_i \in I_j} \delta_i \frac{\partial \ln \lambda(x_i, \hat{\theta})}{\partial \theta_l}, l = 1, 2, \dots, d; j = 1, 2, \dots, k.$$

$$(4.3) \quad \hat{I} = [\hat{i}_{ll'}]_{d \times d}, \text{ with } \hat{i}_{ll'} = \frac{1}{n} \sum_{i=1}^n \delta_i \frac{\partial \ln \lambda(x_i, \hat{\theta})}{\partial \theta_l} \frac{\partial \ln \lambda(x_i, \hat{\theta})}{\partial \theta_{l'}}, l, l' = 1, 2, \dots, d$$

From the definition of Z in (4.1), the test statistic Y^2 should be written as

$$Y^2 = X^2 + Q$$

where

$$X^2 = \sum_{j=1}^k \frac{(U_j - e_j)}{U_j}, Q = \hat{W}^T \hat{G}^{-1} \hat{W} \text{ and } W = \hat{C} \hat{A}^{-1} Z.$$

Under hypothesis H_0 , the statistic Y^2 is chi square with $r = \text{rank}(\Sigma)$ degrees of freedom. In most cases, Σ has full rank k , but in particular class of cases, when the matrix G is degenerated we have $\text{rank}(\Sigma) = k - 1$.

The null hypothesis H_0 is rejected with approximate significance level α if $Y^2 > \mathcal{X}_\alpha^2(r)$ or $Y^2 < \mathcal{X}_{1-\alpha}^2(r)$, where $\mathcal{X}_\alpha^2(r)$ and $\mathcal{X}_{1-\alpha}^2(r)$ are the quantities of chi-square with r degrees of freedom.

5 Chi-squared type test for gamma lindley distribution in censored data case

In particular, we shall give chi-squared test NRR for the hypothesis H_0 , that the data T_i are coming from the Gamma Lindley distribution, with the probability density, cumulative distribution, hazard rate, survival, cumulative hazard and log-likelihood functions are given in formulas (2.1), (2.2), (2.4), (2.3), (2.5) and (3.2), respectively.

Let $\hat{\theta} = (\hat{\theta}, \hat{\beta})^T$ be maximum likelihood estimations which are solutions of the non-linear system equations (3.3) and (3.4).

Using the formula (4.3) and (4.2), the elements of the information matrix $\hat{I} = [\hat{i}_{uv}]_{2 \times 2}$ are

$$\begin{aligned} \hat{i}_{11} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{(\hat{\beta} + 1) t_i}{(\hat{\beta} + \hat{\theta} \hat{\beta} - \hat{\theta}) t_i + 1} - \frac{2}{\hat{\theta}} - \frac{(\hat{\beta} + 2\hat{\theta} \hat{\beta} - 2\hat{\theta}) t_i + \hat{\beta}}{\hat{\theta} (\hat{\beta} + \hat{\theta} \hat{\beta} - \hat{\theta}) t_i + \hat{\beta} + \hat{\theta} \hat{\beta}} \right)^2, \\ \hat{i}_{22} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{(1 + \hat{\theta}) t_i}{(\hat{\beta} + \hat{\theta} \hat{\beta} - \hat{\theta}) t_i + 1} - \frac{\hat{\theta} ((1 + \hat{\theta}) t_i + 1) + 1}{\hat{\theta} (\hat{\beta} + \hat{\theta} \hat{\beta} - \hat{\theta}) t_i + \hat{\beta} + \hat{\theta} \hat{\beta}} \right)^2, \\ \hat{i}_{12} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{(\hat{\beta} + 1) t_i}{(\hat{\beta} + \hat{\theta} \hat{\beta} - \hat{\theta}) t_i + 1} - \frac{2}{\hat{\theta}} - \frac{(\hat{\beta} + 2\hat{\theta} \hat{\beta} - 2\hat{\theta}) t_i + \hat{\beta}}{\hat{\theta} (\hat{\beta} + \hat{\theta} \hat{\beta} - \hat{\theta}) t_i + \hat{\beta} + \hat{\theta} \hat{\beta}} \right) \\ &\quad \left(\frac{(1 + \hat{\theta}) t_i}{(\hat{\beta} + \hat{\theta} \hat{\beta} - \hat{\theta}) t_i + 1} - \frac{\hat{\theta} ((1 + \hat{\theta}) t_i + 1) + 1}{\hat{\theta} (\hat{\beta} + \hat{\theta} \hat{\beta} - \hat{\theta}) t_i + \hat{\beta} + \hat{\theta} \hat{\beta}} \right). \end{aligned}$$

And the elements of the matrix $\hat{C} = [\hat{C}_{lj}]_{2 \times k}$

$$\hat{C}_{1j} = \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left(\frac{(\hat{\beta} + 1)t_i}{(\hat{\beta} + \hat{\theta}\hat{\beta} - \hat{\theta})t_i + 1} - \frac{2}{\hat{\theta}} - \frac{(\hat{\beta} + 2\hat{\theta}\hat{\beta} - 2\hat{\theta})t_i + \hat{\beta}}{\hat{\theta}(\hat{\beta} + \hat{\theta}\hat{\beta} - \hat{\theta})t_i + \hat{\beta} + \hat{\theta}\hat{\beta}} \right),$$

$$\hat{C}_{2j} = \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left(\frac{(1 + \hat{\theta})t_i}{(\hat{\beta} + \hat{\theta}\hat{\beta} - \hat{\theta})t_i + 1} - \frac{\hat{\theta}((1 + \hat{\theta})t_i + 1) + 1}{\hat{\theta}(\hat{\beta} + \hat{\theta}\hat{\beta} - \hat{\theta})t_i + \hat{\beta} + \hat{\theta}\hat{\beta}} \right).$$

Then, we obtain the statistic

$$Y^2 = \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + Q.$$

6 Applications

A data frame from a trial of 42 leukemia patients. Some were treated with the drug 6-mercaptopurine and the rest are controls. The trial was designed as matched pairs, both withdrawn from the trial when either came out of remission. This data included 30 observations and 12 censoring times

1	10	22	7	3	32*	12
23	8	22	17	6	2	16
11	34*	8	32*	12	25*	2
11*	5	20*	4	19*	15	6
8	17*	23	35*	5	6	11
13	4	9*	1	6*	8	10*

* means censored data

Using R statistical software (the BB package), we find the values of the maximum likelihood estimators of Gamma Lindley distribution parameters: $\hat{\theta} = 0.05696237$, $\hat{\beta} = 2.256831$.

We choose for example 6 intervals, ie $k = 6$. Further results to calculate Y^2 are shown below

a_j	4.953122	7.73171	11.00439	12.73609	17.41944	35
U_j	7	6	7	2	4	4
Z_j	0.8264873	0.6721839	0.8264873	0.05497055	0.3635772	0.3635772
e_j	1.64375	1.64375	1.64375	1.64375	1.64375	1.64375

$$\hat{C} = \begin{pmatrix} 5.43729101 & 4.355040987 & 4.822547609 & 1.3252370717 & 2.5568083410 & 2.3972791918 \\ -0.01219789 & -0.003339234 & -0.002265507 & -0.0004385152 & -0.0006373237 & -0.0003524517 \end{pmatrix}$$

Hence

$$\hat{W} = (91.81183956, -0.09171966)$$

The estimated Fisher's information matrix

$$\hat{I} = \begin{pmatrix} 615.759051 & 8.28310721 \\ 8.28310721 & 0.001232983 \end{pmatrix}$$

and

$$\hat{G} = \begin{pmatrix} 0.202192 & 8.886584001 \\ 8.886584001 & 0.0002217965 \end{pmatrix}$$

We obtain $Q = -1.918903$, $X^2 = 14.19921$ and the value of statistic test is $Y^2 = 12.28031$. Since $Y^2 = 12.28031 < \chi_{0.05}^2(6) = 12.5916$, so we concluded that the Gamma Lindley distribution can fit this data.

7 Conclusions

According to Zeghdoudi and Nedjar (2016a,2016b), a Gamma Lindley distribution provides more flexibility to survival analysis, analyze skewed data and lifetime data. In this work, modified chi-squared goodness-of-fit tests based on the NRR statistic, proposed by Bagdanovicius and Nikulin (2011) for censored data. An application of this distribution in survival analysis is given to discussing the potential of the presented test. The criteria tests are derived from both complete and right-censored data. The results obtained showed the practice of the proposed tests and can be used to check whether observed data, from reliability and survival analysis, behave according to this model.

References

- [1] A. Asgharzadeh, H.S. Bakouch, L. Esmaeili, Pareto, *Poisson-Lindley distribution and its application*, Journal of Applied Statistics (2013), 1-18.
- [2] V.B. Bagdonavičius, and M. Nikulin, *Chi-square goodness-of-fit test for right censored data* Int. J. Appl. Math. Stat. 24, SI-11A (2011), 30-50.
- [3] V.B. Bagdonavičius, J. Kruopis, and M. Nikulin *Nonparametric Tests for Censored Data*, ISTE and J.Wiley, 2010.
- [4] M.E. Ghitany, D.K. Al-Mutairi, and S. Nadarajah, *Zero-truncated Poisson-Lindley distribution and its application*, Math. Comput. Simulation 79, (2008a), 279-287.
- [5] V.B. Bagdonavičius and M.S. Nikulin, *Statistical models to analyze failure, wear, fatigue, and degradation data with explanatory variables*, Communications in Statistics - Theory and Methods, 38, 16-17 (2009), 3031-3047.
- [6] V.B. Bagdonavičius and M.S. Nikulin, *Chi-squared tests for general composite hypotheses from censored samples*, Comptes Rendus Mathematique, 349, 3 (2011), 219-223.
- [7] V.B. Bagdonavičius and M.S. Nikulin, *Chi-squared goodness-of-fit test for right censored data*, International Journal of Applied Mathematics and Statistics 24, SI-11A (2011), 30-50.
- [8] V.B. Bagdonavičius, J. Krupois and M.S. Nikulin, *Non-parametric Tests for Censored Data*, Wiley, 2011.

- [9] V.B. Bagdonavičius, R.J. Levulienė and M.S. Nikulin, *Exact goodness-of-fit tests for shape-scale families and type II censoring*, Lifetime Data Analysis, (2012), 1-23.
- [10] V.B. Bagdonavičius and M.S. Nikulin, *Statistical methods to analyse failures of complex systems in presence of wear, fatigue and degradation: An engineering perspective in accelerated trials*, In Electronic Instrument Engineering. APEIE, 2008.
- [11] M. E. Ghitany, B. Atieh, and S. Nadarajah, *Lindley distribution and its applications*, Mathematics and Computers in Simulation, 78 (2008b), 493-506.
- [12] D.V. Lindley, *Fiducial distributions and Bayes' theorem*, Journal of the Royal Society, Series B, 20 (1958), 102-107.
- [13] M.S. Nikulin, *Chi-square test for normality*, In Proceedings of International Vilnius Conference on Probability Theory and Mathematical Statistics, 1973a.
- [14] M.S. Nikulin, *On a chi-square test for continuous distribution*, Theory of Probability and its Application, 19 (1973b), 638-639.
- [15] M.S. Nikulin, *Chi-square test for continuous distributions with location and scale parameters*, Teoriya Veroyatnostei i ee Primeneniya, 18, 3 (1973c), 583-591.
- [16] M.S. Nikulin, N. Saaidia, and R. Tahir, *Reliability analysis of redundant systems by simulation for data with unimodal hazard rate functions*, Journal MESA, 2, 3 (2011), 277-286.
- [17] M.S. Nikulin, N. Saaidia, and R. Tahir, *Recent results in the analysis of redundant systems*, Recent Advances in System Reliability (2012), 181-193.
- [18] M. Sankaran, *The discrete Poisson-Lindley distribution*, Biometrics 26 (1970), 145-149.
- [19] H. Zeghdoudi, S. Nedjar *Gamma Lindley distribution and its application*, Journal of Applied Probability and Statistics 11, 1 (2016a), 129-138.
- [20] H. Zeghdoudi, S. Nedjar, *On gamma Lindley distribution: properties and simulations*, Journal of Computational and Applied Mathematics 298 (2016b), 167-174.

Authors' addresses:

Abderrahmane Bensid
Department of Mathematics, Faculty of Exact Sciences,
University the Brothers Mentouri Constantine, 25019, Algeria.
E-mail: abderrahmanebensid4@gmail.com

Halim Zeghdoudi
LaPS laboratory, Badji-Mokhtar University,
Box 12, Annaba, 23000, Algeria.
E-mail: hzeghdoudi@yahoo.fr

Azzedine Benchettah
LANOS laboratory, Badji-Mokhtar University, Box 12,
Annaba, 23000, Algeria.
E-mail: abenchettah@hotmail.com