Hacettepe Journal of Mathematics and Statistics
Volume 43 (6) (2014), 1063 – 1077

The Lindley-Poisson distribution in lifetime analysis and its properties

Wenhao Gui* , Shangli Zhang† and Xinman Lu‡

Abstract

In this paper, we introduce a new compounding distribution, named the Lindley-Poisson distribution. We investigate its characterization and statistical properties. The maximum likelihood inference using EM algorithm is developed. Asymptotic properties of the MLEs are discussed and simulation studies are performed to assess the performance of parameter estimation. We illustrate the proposed model with two real applications and it shows that the new distribution is appropriate for lifetime analyses.

2000 AMS Classification:

Keywords: Lindley distribution, Poisson distribution, Hazard function, Maximum likelihood estimation, EM algorithm, Fisher information matrix.

Received 08 : 08 : 2013 : Accepted 11 : 01 : 2014 Doi : 10.15672/HJMS.201427453

1. Introduction

The Lindley distribution was originally introduced by [16] to illustrate a difference between fiducial distribution and posterior distribution. It has attracted a wide applicability in survival and reliability. Its density function is given by

$$f(t) = \frac{\theta^2}{1+\theta}(1+t)e^{-\theta t}, \quad t, \theta > 0. \tag{1.1}$$

We denoted this by writing $LD(\theta)$. The density in (1.1) indicates that the Lindley distribution is a mixture of an exponential distribution with scale θ and a gamma distribution with shape 2 and scale θ, where the mixing proportion is $\theta/(1+\theta)$.

*Department of Mathematics and Statistics, University of Minnesota Duluth, Duluth, MN 55812, USA.
Email: wgui@d.umn.edu
†Department of Mathematics, Beijing Jiaotong University, Beijing 100044, China.
‡Business School,Nankai University, Tianjin 300071, China.

[11] provided a comprehensive treatment of the statistical properties of the Lindley distribution and showed that in many ways it performs better than the well-known exponential distribution. [20] discussed the discrete Poisson–Lindley distribution by compounding the Poission distribution and the Lindley distribution. [10] investigated the properties of the zero-truncated Poisson—Lindley distribution. [3] extended the Lindley distribution by exponentiation. [22] introduced and analyzed a three-parameter generalization of the Lindley distribution, which was used by [17] to derive an extended version of the compound Poisson distribution. [21] introduced a two-parameter Lindley distribution of which the one-parameter $LD(\theta)$ is a particular case, for modeling waiting and survival times data. [9] introduced a two-parameter power Lindley distribution (PL) and discussed its properties. [18] proposed a generalized Lindley distribution (GL) and provided comprehensive account of the mathematical properties of the distribution.

On the other hand, the studies and analysis of lifetime data play a central role in a wide variety of scientific and technological fields. There have been developed several distributions by compounding some useful life distributions. [1] introduced a two-parameter exponential-geometric (EG) distribution with decreasing failure rate by compounding an exponential with a geometric distribution. [15] proposed an exponential-Poisson (EP) distribution by mixing an exponential and zero truncated Poisson distribution and discussed its various properties. [5] introduced a new two-parameter distribution family with decreasing failure rate by mixing power-series distribution and exponential distribution.

The aim of this paper is to propose an extension of the Lindley distribution which offers a more flexible distribution for modeling lifetime data. In this paper, we introduce an extension of the Lindley distribution by mixing Lindley and zero truncated Poisson distribution. It differs from the discrete Poisson–Lindley distribution proposed by [20]. Since the Lindley distribution is not a generalization of exponential distribution, the model EP in [15] can not be obtained as a particular case of the new model in this paper. An interpretation of the proposed model is as follows: a situation where failure occurs due to the presence of an unknown number, Z, of initial defects of same kind. Z is a zero truncated Poisson variable. Their lifetimes, Y's, follow a Lindley distribution. Then for modeling the first failure X, the distribution leads to the Lindley–Poisson distribution. We aim to discuss some properties of the proposed distribution.

The rest of this paper is organized as follows: in Section 2, we present the new Lindley-Poisson distribution and investigate its basic properties, including the shape properties of its density function and the hazard rate function, stochastic orderings and representation, moments and measurements based on the moments. Section 3 discusses the distributions of some extreme order statistics. The maximum likelihood inference using EM algorithm and asymtotical properties of the estimates are discussed in Section 4. Simulation studies are also conducted in this Section. Section 5 gives a real illustrative application and reports the results. Our work is concluded in Section 6.

2. Lindley-Poisson Distribution and its Properties

2.1. Density and hazard function. The new distribution can be constructed as follows. Suppose that the failure of a device occurs due to the presence of Z (unknown number) initial defects of some kind. Let $Y_1, Y_2, ..., Y_Z$ denote the failure times of the initial defects, then the failure time of this device is given by $X = \min(Y_1, ..., Y_Z)$.

Suppose the failure times of the initial defects $Y_1, Y_2, ..., Y_Z$ follow a Lindley distribution $LD(\theta)$ and Z has a zero truncated Poisson distribution with probability mass function as follows:

$$p(Z = z) = \frac{\lambda^z e^{-\lambda}}{z!(1 - e^{-\lambda})}, \quad \lambda > 0, z = 1, 2, ... \tag{2.1}$$

By assuming that the random variables Y_i and Z are independent, then the density of $X|Z = z$ is given by

$$f(x|z) = \frac{\theta^2(x+1)ze^{-xz\theta}(\theta+\theta x+1)^{z-1}}{(\theta+1)^z}, \quad x > 0,$$

and the marginal probability density function of X is

$$f(x) = \frac{\theta^2\lambda(x+1)e^{\frac{\lambda e^{-\theta x}(\theta+\theta x+1)}{\theta+1}-\theta x}}{(\theta+1)\left(e^\lambda - 1\right)}, \quad \theta > 0, \lambda > 0, x > 0. \tag{2.2}$$

In the sequel, the distribution of X will be refered to as the LP, which is customary for such a name given to the distribution arising via the operation of compounding in the literature.

2.1. Theorem. *Considering the LP distribution with the probability density function in (2.2), we have the following properties:*

(1) *As λ goes to zero, $LP(\theta, \lambda)$ leads to the Lindley distribution $LD(\theta)$.*
(2) *If $\theta^2(\lambda+1) \geq 1$, $f(x)$ is decreasing in x. If $\theta^2(\lambda+1) < 1$, $f(x)$ is a unimodal function at x_0, where x_0 is the solution of the equation $\theta^2\lambda(x+1)^2+(\theta+1)e^{\theta x}(\theta+\theta x-1) = 0$.*

Proof. 1. As λ goes to zero, then

$$\begin{aligned}
\lim_{\lambda\to 0} f(x) &= \lim_{\lambda\to 0} \frac{\theta^2\lambda(x+1)e^{\frac{\lambda e^{-\theta x}(\theta+\theta x+1)}{\theta+1}-\theta x}}{(\theta+1)\left(e^\lambda - 1\right)} \\
&= \frac{\theta^2(x+1)e^{-\theta x}}{\theta+1},
\end{aligned}$$

which is the probability density distribution of $LD(\theta)$.

2. $f(0) = \frac{\theta^2 e^\lambda \lambda}{(\theta+1)\left(e^\lambda-1\right)}$ and $f(\infty) = 0$. The first derivative of $\log f(x)$ is

$$\frac{d\log f(x)}{dx} = -\frac{e^{-\theta x}\left[\theta^2\lambda(x+1)^2+(\theta+1)e^{\theta x}(\theta+\theta x-1)\right]}{(\theta+1)(x+1)}.$$

Let $s(x) = \theta^2\lambda(x+1)^2+(\theta+1)e^{\theta x}(\theta+\theta x-1)$, then $s(0) = \theta^2(\lambda+1)-1$ and $s(\infty) = \infty$, $s'(x) = \theta^2(x+1)\left[2\lambda+(\theta+1)e^{\theta x}\right] > 0$.

If $\theta^2(\lambda+1) \geq 1$, then $s(x) \geq 0$, $\frac{d\log f(x)}{dx} \leq 0$, i.e., $f(x)$ is decreasing in x. If $\theta^2(\lambda+1) < 1$, $f(x)$ is a unimodal function at x_0, where x_0 is the solution of the equation $s(x) = 0$.

□

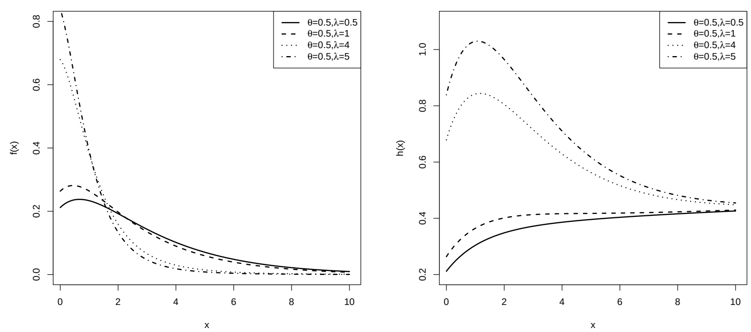
The cumulative distribution of the LP distribution is given by

$$F(x) = \frac{e^\lambda - e^{\frac{\lambda e^{-\theta x}(\theta+\theta x+1)}{\theta+1}}}{e^\lambda - 1}, \quad x > 0. \tag{2.3}$$

The hazard rate function of the $LP(\theta, \lambda)$ distribution is given by

$$h(x) = \frac{\theta^2\lambda(x+1)e^{\frac{\lambda e^{-\theta x}(\theta+\theta x+1)}{\theta+1}-\theta x}}{(\theta+1)\left[e^{\frac{\lambda e^{-\theta x}(\theta+\theta x+1)}{\theta+1}}-1\right]}, \quad x > 0. \tag{2.4}$$

2.2. Theorem. *Considering the hazard function of the LP distribution, we have the following properties:*



(a) density function **(b)** hazard function

Figure 1. Plots of the LP density and hazard function for some parameter values.

(1) *If* $-\theta^3\lambda+\theta^2\lambda+\theta+1>0$ *and the equation* $(\theta+1)e^{\theta x}-\theta^2\lambda(x+1)^2(\theta+\theta x-1)=0$ *has no real roots, then the hazard function is increasing.*
(2) *If* $-\theta^3\lambda+\theta^2\lambda+\theta+1<0$ *and the equation* $(\theta+1)e^{\theta x}-\theta^2\lambda(x+1)^2(\theta+\theta x-1)=0$ *has one real roots, then the hazard function is bathtub shaped.*

Proof. $h(0)=\frac{\theta^2 e^{\lambda}\lambda}{(\theta+1)(e^{\lambda}-1)}$. For the LP distribution, we have

$$\eta(x)=-\frac{f'(x)}{f(x)}=\frac{e^{-\theta x}\left[\theta^2\lambda(x+1)^2+(\theta+1)e^{\theta x}(\theta+\theta x-1)\right]}{(\theta+1)(x+1)},$$

and its first derivative is

$$\eta'(x)=\frac{e^{-\theta x}\left[(\theta+1)e^{\theta x}-\theta^2\lambda(x+1)^2(\theta+\theta x-1)\right]}{(\theta+1)(x+1)^2}.$$

Let $t(x)=(\theta+1)e^{\theta x}-\theta^2\lambda(x+1)^2(\theta+\theta x-1)$, then $t(0)=-\theta^3\lambda+\theta^2\lambda+\theta+1$ and $t(\infty)=\infty$, the sign of $\eta'(x)$ is the sign of $t(x)$ and $\eta'(x)=0$ if $t(x)=0$. The properties follow from the results in [12]. $\square$

For the Lindley distribution $LD(\theta)$, its hazard function $h(x)=\frac{\theta^2(1+x)}{\theta+1+\theta x}$ which is increasing. For the exponential distribution, its hazard function $h(x)=\theta$ which is a constant. (2.4) shows the flexibility of the LP distribution over the Lindley and exponential distribution.

Figure 1a shows some density functions of the $LP(\theta,\lambda)$ distribution with various parameters. Figure 1b shows some shapes of the $LP(\theta,\lambda)$ hazard function with various parameters.

2.2. Stochastic Ordering. In probability theory and statistics, a stochastic order quantifies the concept of one random variable being “bigger" than another. A random variable X is less than Y in the ususal stochastic order (denoted by $X \prec_{st} Y$) if $F_X(x)\geq F_Y(x)$ for all real x. X is less than Y in the hazard rate order (denoted by $X \prec_{hr} Y$) if $h_X(x)\geq h_Y(x)$, for all $x\geq 0$. X is less than Y in the likelihood ratio order (denoted by $X \prec_{lr} Y$) if $f_X(x)/f_Y(x)$ increases in x over the union of the supports of X and Y. It is known that $X \prec_{lr} Y \Rightarrow X \prec_{hr} \Rightarrow X \prec_{st} Y$, see [19].

2.3. Theorem. *If* $X \sim LP(\theta, \lambda_1)$ *and* $Y \sim LP(\theta, \lambda_2)$, *and* $\lambda_1 < \lambda_2$, *then* $Y \prec_{lr} X$, $Y \prec_{hr} X$ *and* $Y \prec_{st} X$.

Proof. The density ratio is given by

$$U(x) = \frac{f_X(x)}{f_Y(x)} = \frac{\left(e^{\lambda_2} - 1\right)\lambda_1 \exp\left(\frac{\lambda_1 e^{-\theta x}(\theta+\theta x+1)}{\theta+1} - \frac{\lambda_2 e^{-\theta x}(\theta+\theta x+1)}{\theta+1}\right)}{\left(e^{\lambda_1} - 1\right)\lambda_2}.$$

Taking the derivative with respect to x,

$$U'(x) = -\frac{\theta^2\left(e^{\lambda_2} - 1\right)\lambda_1\left(\lambda_1 - \lambda_2\right)(x+1)\exp\left(-\frac{e^{-\theta x}\left(-\lambda_1(\theta+\theta x+1)+\lambda_2(\theta+\theta x+1)+\theta(\theta+1)xe^{\theta x}\right)}{\theta+1}\right)}{(\theta+1)\left(e^{\lambda_1} - 1\right)\lambda_2}.$$

If $\lambda_1 < \lambda_2$, $U'(x) > 0$, $U(x)$ is an increasing function of x. The results follow. □

2.3. Moments and Measures based on moments. In this section, we consider the moments and measures of the LP distribution $X \sim LP(\theta, \lambda)$. The k-th raw moment of X is given by, for $k = 1, 2, ...$,

$$\mu_k = \mathbb{E}(X^k) = k\int_0^\infty x^{k-1}\bar{G}(x)dx = \int_0^\infty \frac{kx^{k-1}[e^{\frac{\lambda e^{-\theta x}(\theta+\theta x+1)}{\theta+1}} - 1]}{e^\lambda - 1}\,dx.$$

$\mathbb{E}(X^k)$ cannot be expressed in a simple closed-form and need be calculated numerically. Using numerical integration, we can find some measures based on the moments such as mean, variance, skewness and kurtosis etc. For the skewness and kurtosis coefficients, $\sqrt{\beta_1} = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{3/2}}$ and $\beta_2 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2}$.

The cumulative distribution of the LP distribution is given in (2.3). The qth ($0 \le q \le 1$) quantile $x_q = F^{-1}(q)$ of the $LP(\theta, \lambda)$ distribution is

$$x_q = \frac{-\theta - W\left(-\frac{e^{-\theta-1}(\theta+1)\log\left(e^\lambda - e^\lambda q + q\right)}{\lambda}\right) - 1}{\theta},$$

where $W(a)$ giving the principal solution for w in $a = we^w$ is pronounced as Lambert W function, see [14].

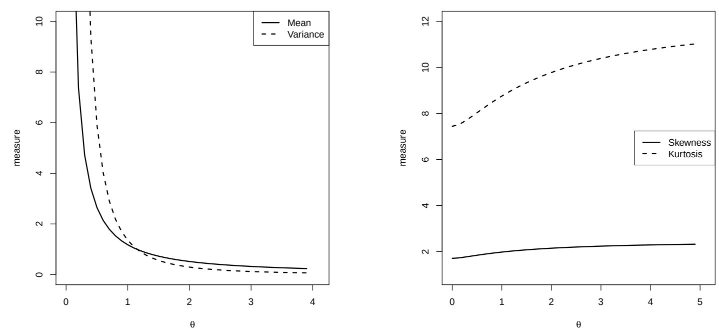
In particular, the median of the $LP(\theta, \lambda)$ distribution is given by

$$x_m = \frac{-\theta - W\left(-\frac{e^{-\theta-1}(\theta+1)\log\left(\frac{1}{2}\left(e^\lambda + 1\right)\right)}{\lambda}\right) - 1}{\theta}. \tag{2.5}$$

Figure 2a displays the mean and variance of the $LP(\theta, \lambda = 1)$ distribution. Figure 2b shows the skewness and kurtosis coefficients of the $LP(\theta, \lambda = 1)$ distribution. From the figures, it is found that the $LP(\theta, \lambda = 1)$ distribution has positive skewness and kurtosis coefficients. The coefficients are increasing functions of θ.

3. Distributions of Order Statistics

Let $X_1, X_2, ..., X_n$ be a random sample of size n from the $LP(\theta, \lambda)$ distribution. By the usual central limit theorem, the same mean $(X_1 + ... + X_n)/n$ approaches the normal distribution as $n \to \infty$. Sometimes one would be interested in the asymptotics of the sample minima $X_{1:n} = \min(X_1, ..., X_n)$ and the sample maxima $X_{n:n} = \max(X_1, ..., X_n)$. These extreme order statistics represent the life of series and parallel system and have important applications in probability and statistics.



(a) **(b)**

Figure 2. (a) Plot of mean and variance of the $LP(\theta, \lambda = 1)$ distribution; (b) Plot of skewness and kurtosis coefficients of the $LP(\theta, \lambda = 1)$ distribution.

3.1. Theorem. *Let $X_{1:n}$ and $X_{n:n}$ be the smallest and largest order statistics from the $LP(\theta, \lambda)$ distribution. Then*
(1) $\lim_{n\to\infty} P(X_{1:n} \le b_n^* t) = 1 - e^{-t}, t > 0,$ *where* $b_n^* = F^{-1}(1/n)$.
(2) $\lim_{n\to\infty} P(X_{n:n} \le b_n t) = e^{-t^{-1}}, t > 0,$ *where* $b_n = F^{-1}(1 - 1/n)$.

Proof. We apply the following asymptotical results for $X_{1:n}$ and $X_{n:n}$ ([2]).
(1) For the smallest order statistic $X_{1:n}$, we have

$$\lim_{n\to\infty} P(X_{1:n} \le a_n^* + b_n^* t) = 1 - e^{-t^c}, \quad t > 0, c > 0,$$

(of the Weibull type) where $a_n^* = F^{-1}(0)$ and $b_n^* = F^{-1}(1/n) - F^{-1}(0)$ if and only if $F^{-1}(0)$ is finite and for all $t > 0$ and $c > 0$,

$$\lim_{\epsilon\to 0^+} \frac{F(F^{-1}(0) + \epsilon t)}{F(F^{-1}(0) + \epsilon)} = t^c.$$

For the $LP(\theta, \lambda)$ distribution, its cumulative distribution function is

$$F(x) = \frac{e^{\lambda} - e^{\frac{\lambda e^{-\theta x}(\theta + \theta x + 1)}{\theta + 1}}}{e^{\lambda} - 1}, \quad \theta > 0, \lambda > 0, x > 0.$$

Let $F(x) = 0$, we have $\theta + \theta x + 1 = e^{\theta x}(\theta + 1) \ge (1 + \theta x)(\theta + 1), \theta x^2 \le 0$. Thus $F^{-1}(0) = 0$ is finite. Furthermore,

$$\lim_{\epsilon\to 0^+} \frac{F(0 + \epsilon t)}{F(0 + \epsilon)} = t \lim_{\epsilon\to 0^+} \frac{f(\epsilon t)}{f(\epsilon)} = t.$$

Therefore, we obtain that $c = 1$, $a_n^* = 0$ and $b_n^* = F^{-1}(1/n)$ which is the $\frac{1}{n}$th quantile.
(2) For the largest order statistic $X_{n:n}$, we have

$$\lim_{n\to\infty} P(X_{n:n} \le a_n + b_n t) = e^{-t^{-d}}, \quad t > 0, d > 0$$

(of the Fréchet type) where $a_n = 0$ and $b_n^* = F^{-1}(1 - 1/n)$ if and only if $F^{-1}(1) = \infty$ and there exists a constant $d > 0$ such that

$$\lim_{x\to\infty} \frac{1 - F(xt)}{1 - F(x)} = t^{-d}.$$

For the $LP(\theta, \lambda)$ distribution, let $F(x) = 1$, then $\frac{\lambda e^{-\theta x}(\theta+\theta x+1)}{\theta+1} = 0$, we have the solution $x = \infty$. Thus $F^{-1}(1) = \infty$. Furthermore,

$$\lim_{x\to\infty} \frac{1 - F(xt)}{1 - F(x)} = t^{-1}.$$

Therefore, we obtain that $d = 1$, $a_n = 0$ and $b_n = F^{-1}(1-1/n)$ which is the the $(1-\frac{1}{n})$th quantile.

□

3.2. Remark. Let $Q^*(t)$ and $Q(t)$ denote the limiting distributions of the random variables $(X_{1:n} - a_n^*)/b_n^*$ and $(X_{n:n} - a_n)/b_n$ respectively, then for $k > 1$, the limiting distributions of $(X_{k:n} - a_n^*)/b_n^*$ and $(X_{n-k+1:n} - a_n)/b_n$ are given by, see [2],

$$\lim_{n\to\infty} P(X_{k:n} \le a_n^* + b_n^* t) = 1 - \sum_{j=0}^{k-1} (1 - Q^*(t)) \frac{[-\log(1 - Q^*(t))]^j}{j!},$$

$$\lim_{n\to\infty} P(X_{n-k+1:n} \le a_n + b_n t) = \sum_{j=0}^{k-1} Q(t) \frac{[-\log Q(t)]^j}{j!}.$$

4. Estimation and inference

4.1. Maximum likelihood estimation. Here, we consider the maximum likelihood estimation about the parameters (θ, λ) of the LP model. Suppose $y_{obs} = \{x_1, x_2, ..., x_n\}$ is a random sample of size n from the $LP(\theta, \lambda)$ distribution. Then the log-likelihood function is given by

$$\begin{aligned} l &= \log \prod_{i=1}^{n} f_X(x_i) \\ &= \lambda \sum_{i=1}^{n} e^{-\theta x_i} + \frac{\theta\lambda \sum_{i=1}^{n} x_i e^{-\theta x_i}}{\theta + 1} - \theta \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \log(x_i + 1) \\ &\quad + 2n\log(\theta) - n\log(\theta+1) - n\log\left(e^{\lambda} - 1\right) + n\log(\lambda). \end{aligned} \tag{4.1}$$

The associated gradients are found to be

$$\frac{\partial l}{\partial \theta} = -\sum_{i=1}^{n} x_i + \frac{2n}{\theta} - \frac{n}{\theta+1} - \frac{\theta(\theta+2)\lambda \sum_{i=1}^{n} x_i e^{-\theta x_i}}{(\theta+1)^2} - \frac{\theta\lambda \sum_{i=1}^{n} x_i^2 e^{-\theta x_i}}{\theta+1}, \tag{4.2}$$

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^{n} e^{-\theta x_i} + \frac{\theta \sum_{i=1}^{n} x_i e^{-\theta x_i}}{\theta+1} - \frac{n e^{\lambda}}{e^{\lambda} - 1} + \frac{n}{\lambda}. \tag{4.3}$$

The estimates of the parameters maximize the likelihood function. Equalizing the obtained gradients expressions to zero yield the likelihood equations. However, they do not lead to explicit analytical solutions for the parameters. Thus, the estimates can be obtained by means of numerical procedures such as Newton-Raphson method. The program R provides the nonlinear optimization routine *optim* for solving such problems.

The equation $\frac{\partial l}{\partial \theta} = 0$ could be solved exactly for λ, namely

$$(4.4)\hat{\lambda} = \frac{(\hat{\theta}+1)\left[\hat{\theta}(\hat{\theta}+1)\sum_{i=1}^{n} x_i - (\hat{\theta}+2)n\right]}{\hat{\theta}\left[-(\hat{\theta}+1)^2\sum_{i=1}^{n} x_i e^{-\hat{\theta}x_i} - \hat{\theta}(\hat{\theta}+1)\sum_{i=1}^{n} x_i^2 e^{-\hat{\theta}x_i} + \sum_{i=1}^{n} x_i e^{-\hat{\theta}x_i}\right]},$$

conditional on the value of $\hat{\theta}$, where $\hat{\theta}$ and $\hat{\lambda}$ are the maximum likelihood estimators for the parameters θ and λ, respectively.

In the following, Theorem 4.1 gives the condition for the existence and uniqueness of $\hat{\lambda}$ when θ is known.

4.1. Theorem. *For the MLEs, let $l_2(\lambda;\theta,y_{obs})$ denote the function on the RHS of the expression in (4.3), if θ is known, then the root of $l_2(\lambda;\theta,y_{obs}) = 0$, $\hat{\lambda}$, uniquely exists if* $\sum_{i=1}^{n} e^{-\theta x_i} + \frac{\theta\sum_{i=1}^{n} x_i e^{-\theta x_i}}{\theta+1} > \frac{n}{2}$.

Proof. Notice that $\lim_{\lambda\to 0} l_2(\lambda;\theta,y_{obs}) = \sum_{i=1}^{n} e^{-\theta x_i} + \frac{\theta\sum_{i=1}^{n} x_i e^{-\theta x_i}}{\theta+1} - \frac{n}{2} > 0$ when $\sum_{i=1}^{n} e^{-\theta x_i} + \frac{\theta\sum_{i=1}^{n} x_i e^{-\theta x_i}}{\theta+1} > \frac{n}{2}$. On the other hand, we can show that $\lim_{\lambda\to\infty} l_2(\lambda;\theta,y_{obs}) = \sum_{i=1}^{n} e^{-\theta x_i} + \frac{\theta\sum_{i=1}^{n} x_i e^{-\theta x_i}}{\theta+1} - n$. Consider $g(x) = e^{-\theta x} + \frac{\theta}{\theta+1} x e^{-\theta x} - 1$, $g(0) = 0$ and $g(\infty) = -1$, $g'(x) = -\frac{\theta^2(x+1)e^{\theta(-x)}}{\theta+1} < 0$, therefore, $\lim_{\lambda\to\infty} l_2(\lambda;\theta,y_{obs}) < 0$, there is at least one root of $l_2(\lambda;\theta,y_{obs}) = 0$. We need to prove that the function $l_2(\lambda;\theta,y_{obs})$ is decreasing in λ. Taking the first derivative

$$l_2'(\lambda;\theta,y_{obs}) = -\frac{\left[-e^{\lambda}\left(\lambda^2+2\right) + e^{2\lambda} + 1\right]n}{\left(e^{\lambda}-1\right)^2\lambda^2} = -\frac{e^{\lambda}\left[-\left(\lambda^2+2\right) + e^{\lambda} + e^{-\lambda}\right]n}{\left(e^{\lambda}-1\right)^2\lambda^2} < 0.$$

This completes the proof.

□

4.2. An EM algorithm. An expectation–maximization (EM) algorithm ([7]) is a powerful method for finding maximum likelihood estimates of parameters in statistical models, where the model depends on unobserved latent variables. The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter estimates are then used to determine the distribution of the latent variables in the next E step. We propose the use of the EM algorithm in this section.

Assume that (X,Z) denotes a random vector, where X denotes the observed data and Z denotes the missing data. To implement the algorithm we define the hypothetical complete–data distribution with density function

$$f(x,z) = p(z)f(x|z) = \frac{\theta^2(x+1)ze^{-xz\theta}(\theta+\theta x+1)^{z-1}}{(\theta+1)^z}\frac{\lambda^z e^{-\lambda}}{z!(1-e^{-\lambda})}, x>0, z=1,2,...,$$

where $\theta > 0$ and $\lambda > 0$ are parameters. It is straightforward to verify that the computation of the conditional expectation of $(Z|X)$ using the pdf

$$p(z|x) = \frac{(\theta+1)^{1-z}\lambda^{z-1}(\theta+\theta x+1)^{z-1}\exp\left(-\frac{\lambda e^{-\theta x}(\theta+\theta x+1)}{\theta+1} + \theta x - \theta x z\right)}{(z-1)!}, z = 1,2,...$$

Then we have

$$\mathbb{E}(Z|X) = 1 + \frac{\lambda e^{-\theta x}(\theta+\theta x+1)}{\theta+1}.$$

The cycle is completed with the M–step which is essentially-full data maximum likelihood over the parameters, with the missing $Z's$ replaced by their conditional expectations $\mathbb{E}(Z|X)$. Thus, an EM iteration is given by

$$\theta^{(t+1)} = 2n[\sum_{i=1}^{n} \frac{x_i+1}{\theta^{(t)}+\theta^{(t)}x_i+1} - \sum_{i=1}^{n} \frac{(x_i+1)\,w_i^{(t)}}{\theta^{(t)}+\theta^{(t)}x_i+1} + \sum_{i=1}^{n} x_i w_i^{(t)} + \frac{\sum_{i=1}^{n} w_i^{(t)}}{\theta^{(t)}+1}]^{-1},$$

$$\lambda^{(t+1)} = n^{-1}[1-e^{-\lambda^{(t)}}]\sum_{i=1}^{n} w_i^{(t)},$$

where $w_i^{(t)} = 1 + \frac{\lambda^{(t)}e^{-\theta^{(t)}x_i}(\theta^{(t)}+\theta^{(t)}x_i+1)}{\theta^{(t)}+1}$.

4.3. Asymtotic variance and covariance of MLEs. It is known that under some regular conditions, as the sample size increases, the distribution of the MLE tends to the bivariate normal distribution with mean (θ, λ) and covariance matrix equal to the inverse of the Fisher information matrix, see [6]. The bivariate normal distribution can be used to construct approximate confidence intervals for the parameters θ and λ.

Let $I = I(\theta, \lambda; y_{obs})$ be the observed matrix with elements I_{ij} with $i, j = 1, 2$. The elements of the observed information matrix are found as follows:

$$I_{11} = -\frac{\left((\theta+1)^2-2\lambda\right)\sum_{i=1}^{n} x_i^2 e^{-\theta x_i}}{(\theta+1)^2} - \frac{\theta\lambda\sum_{i=1}^{n} x_i^3 e^{-\theta x_i}}{\theta+1} + \frac{2\lambda\sum_{i=1}^{n} x_i e^{-\theta x_i}}{(\theta+1)^3} + \frac{2n}{\theta^2} - \frac{n}{(\theta+1)^2},$$

$$l_{12} = l_{21} = \frac{\theta(\theta+2)\sum_{i=1}^{n} x_i e^{-\theta x_i}}{(\theta+1)^2} + \frac{\theta\sum_{i=1}^{n} x_i^2 e^{-\theta x_i}}{\theta+1},$$

$$l_{22} = -\frac{e^{\lambda}n}{(e^{\lambda}-1)^2} + \frac{n}{\lambda^2}.$$

The expectation $J = \mathbb{E}(I(\theta, \lambda; y_{obs}))$ is taken with respect to the distribution of X. The Fisher information matrix is given by

$$J(\theta, \lambda) = n\begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}$$

where

$$J_{11} = -\frac{\left((\theta+1)^2-2\lambda\right)\mathbb{E}(X^2e^{-\theta X})}{(\theta+1)^2} - \frac{\theta\lambda\mathbb{E}(X^3e^{-\theta X})}{\theta+1} + \frac{2\lambda\mathbb{E}(Xe^{-\theta X})}{(\theta+1)^3} + \frac{2}{\theta^2} - \frac{1}{(\theta+1)^2},$$

$$J_{12} = J_{21} = \frac{\theta(\theta+2)\mathbb{E}(Xe^{-\theta X})}{(\theta+1)^2} + \frac{\theta\mathbb{E}(X^2e^{-\theta X})}{\theta+1},$$

$$J_{22} = \frac{1}{\lambda^2} - \frac{e^{\lambda}}{(e^{\lambda}-1)^2}.$$

The inverse of $J(\theta, \lambda)$, evaluated at $\hat{\theta}$ and $\hat{\lambda}$ provides the asymptotic variance–covariance matrix of the MLEs. Alternative estimates can be obtained from the inverse of the observed information matrix since it is a consistent estimator of J^{-1}.

4.4. Simulation study. The random data X from the proposed distribution can be generated as follows:

(1) Generate $Z \sim$ zero truncated Poisson (λ).
(2) Generate $U_i \sim$ Uniform$(0, 1)$, $i = 1, ..., Z$.
(3) Generate $V_i \sim$ Exponential(θ), $i = 1, ..., Z$.
(4) Generate $W_i \sim$ Gamma$(2, \theta)$, $i = 1, ..., Z$.
(5) If $U_i \leq \theta/(1+\theta)$, then set $Y_i = V_i$, otherwise, set $Y_i = W_i$, $i = 1, ..., Z$.
(6) Set $X = \min(Y_1, ..., Y_Z)$.

In order to assess the performance of the approximation of the variances and covariances of the MLEs determined from the information matrix, a simulation study (based on 10000 simulations) has been conducted.

For each value of (θ, λ), the parameter estimates have been obtained by the EM iteration in Section 4.2 with different initial values. The convergence is assumed when the absolute differences between successive estimates are less than 10^{-5}.

The simulated values of $Var(\hat{\theta})$, $Var(\hat{\lambda})$ and $Cov(\hat{\theta}, \hat{\lambda})$ as well as the approximate values determined by averaging the corresponding values obtained from the expected and observed information matrices are given in Table 1. We can see that for large values of n, the approximate values determined from expected and observed information matrices are quite close to the corresponding simulated values. The approximation becomes quite accurate as n increases. As expected, variances and covariances of the MLEs obtained from the observed information matrix are quite close to that of the expected information matrix for large values of n.

Table 1. Variances and covariances of the MLEs.

n	(θ, λ)	Simulated			From expected information			From observed information		
		$Var(\hat{\theta})$	$Var(\hat{\lambda})$	$Cov(\hat{\theta}, \hat{\lambda})$	$Var(\hat{\theta})$	$Var(\hat{\lambda})$	$Cov(\hat{\theta}, \hat{\lambda})$	$Var(\hat{\theta})$	$Var(\hat{\lambda})$	$Cov(\hat{\theta}, \hat{\lambda})$
50	(0.5, 1.0)	0.1263	5.1246	-0.5943	0.0669	4.6146	-0.5249	0.0675	5.5077	-0.5827
50	(1.0, 0.5)	0.1809	2.3515	-0.7432	0.2009	2.6355	-0.8070	0.1112	1.9026	-0.7148
50	(0.5, 2.0)	0.0854	3.0085	-0.4022	0.0755	3.4085	-0.4615	0.0503	2.6381	-0.3235
50	(2.0, 0.5)	0.7783	3.3421	-1.5915	0.7578	3.0401	-1.3959	0.8382	3.6288	-1.6234
50	(2.0, 2.0)	0.7069	2.5474	-1.1334	0.7001	2.0854	-1.0336	0.7149	3.4743	-1.2402
100	(0.5, 1.0)	0.0365	2.9019	-0.3419	0.0476	2.9411	-0.3599	0.0334	2.3195	-0.3281
100	(1.0, 0.5)	0.0901	1.7915	-0.3829	0.0996	1.9011	-0.4122	0.0925	1.643	-0.3645
100	(0.5, 2.0)	0.0234	1.4168	-0.1738	0.0289	1.4896	-0.1882	0.0252	1.2935	-0.162
100	(2.0, 0.5)	0.2743	1.2773	-0.5513	0.2824	1.2676	-0.5510	0.2605	1.2929	-0.511
100	(2.0, 2.0)	0.3602	1.0218	-0.5014	0.3588	1.0148	-0.5218	0.349	0.9358	-0.4904
500	(0.5, 1.0)	0.0064	0.4256	-0.0506	0.0063	0.4238	-0.0496	0.0065	0.4462	-0.052
500	(1.0, 0.5)	0.0545	0.943	-0.2255	0.0522	0.9426	-0.2201	0.0567	0.9446	-0.2278
500	(0.5, 2.0)	0.0028	0.2001	-0.0211	0.0027	0.2009	-0.0209	0.0029	0.1998	-0.0213
500	(2.0, 0.5)	0.0899	0.3562	-0.1753	0.0888	0.3596	-0.1761	0.0938	0.3548	-0.1749
500	(2.0, 2.0)	0.0419	0.1672	-0.0723	0.0418	0.1672	-0.0733	0.0416	0.1673	-0.0723

In addition, simulations have been conduced to investigate the convergence of the proposed EM algorithm in Section 4.2. Ten thousand samples of size 100 and 500 of which are randomly sampled from the LP distribution for each of the five values of (θ, λ) are generated.

The results are presented in Table 2, which gives the averages of the 10000 MLEs, $av(\hat{\theta}), av(\hat{\lambda})$, and average number of iterations to convergence, $av(h)$, together with their standard errors, where

$$av(\hat{\theta}) = \frac{1}{10000}\sum_{i=1}^{10000}\hat{\theta}_i, \quad se(\hat{\theta}) = \sqrt{\frac{1}{10000}\sum_{i=1}^{10000}(\hat{\theta}_i - av(\hat{\theta}))^2},$$

$$av(\hat{\lambda}) = \frac{1}{10000}\sum_{i=1}^{10000}\hat{\lambda}_i, \quad se(\hat{\lambda}) = \sqrt{\frac{1}{10000}\sum_{i=1}^{10000}(\hat{\lambda}_i - av(\hat{\lambda}))^2},$$

$$av(\hat{h}) = \frac{1}{10000}\sum_{i=1}^{10000}\hat{h}_i, \quad se(\hat{h}) = \sqrt{\frac{1}{10000}\sum_{i=1}^{10000}(\hat{h}_i - av(\hat{h}))^2}.$$

From Table 2, it is observed that convergence has been achieved in all cases, even when the initial values are far from the true values and this endorses the numerical stability of the proposed EM algorithm. The EM estimates performed consistently. Standard errors of the MLEs decrease when sample size n increases.

Table 2. The means and standard errors of the EM estimator and iterations to convergence with initial values $(\theta^{(0)}, \lambda^{(0)})$ from 10000 samples.

n	θ	λ	$\theta^{(0)}$	$\lambda^{(0)}$	$av(\hat{\theta})$	$av(\hat{\lambda})$	$se(\hat{\theta})$	$se(\hat{\lambda})$	$av(h)$	$se(h)$
100	0.5	1	0.5	1	0.470	1.493	0.103	1.206	481.949	423.532
100	1	0.5	1	0.5	0.897	0.733	0.171	1.070	435.405	363.209
100	0.5	2	0.5	2	0.525	2.061	0.142	1.469	516.551	318.547
100	2	0.5	2	0.5	1.840	0.854	0.364	1.078	404.442	412.249
100	2	2	2	2	2.093	2.123	0.593	1.346	484.928	489.149
100	0.5	1	0.1	0.1	0.481	1.406	0.107	1.249	537.204	452.071
100	1	0.5	0.1	0.1	0.920	0.807	0.179	1.086	453.290	382.990
100	0.5	2	0.1	0.1	0.523	2.011	0.133	1.288	589.371	498.996
100	2	0.5	0.1	0.1	1.780	0.724	0.366	1.143	445.348	379.776
100	2	2	0.1	0.1	2.130	1.981	0.583	1.271	534.251	462.154
500	0.5	1	0.5	1	0.496	1.106	0.068	0.781	443.485	405.746
500	1	0.5	1	0.5	0.977	0.631	0.085	0.415	327.897	145.757
500	0.5	2	0.5	2	0.507	2.061	0.094	0.979	592.532	380.115
500	2	0.5	2	0.5	1.970	0.576	0.165	0.341	293.798	112.133
500	2	2	2	2	2.020	2.087	0.387	0.954	560.947	576.358
500	0.5	1	0.1	0.1	0.495	1.097	0.066	0.705	572.473	428.584
500	1	0.5	0.1	0.1	0.989	0.586	0.083	0.453	377.760	171.738
500	0.5	2	0.1	0.1	0.508	2.057	0.096	0.952	823.717	605.764
500	2	0.5	0.1	0.1	1.969	0.591	0.167	0.383	347.611	175.877
500	2	2	0.1	0.1	2.041	2.053	0.401	0.962	736.315	735.316

5. Illustrative Examples

In this section, we consider two numerical applications to test the performance of the new distribution. First, we consider the time intervals of the successive earthquakes taken from University of Bosphoros, Kandilli Observatory and Earthquake Research Institute-National Earthquake Monitoring Center. The data set has been previously studied by

[15]. The second dataset originally due to [4], which has also been analyzed previously by [13]. The data represent the survival times of guinea pigs injected with different doses of tubercle bacilli.

Table 3. Maximum likelihood parameter estimates(with (SE)) of the LP, LD, PL and GL models for the two datasets.

Example	Model	Estimations		loglik	AIC	K-S statistic	p-value
	LP	0.6515	2.7778	-32.0766	68.1532	0.1667	0.9024
		(0.2112)	(0.1578)				
	LD	1.0420	–	-34.5092	71.0184	0.2500	0.4490
		(0.1612)	–				
1 ($n = 24$)	PL	0.6215	1.0898	-32.6134	69.2268	0.2083	0.6860
		(0.1026)	(0.1745)				
	GL	0.5940	0.7701	-32.3633	68.7266	0.1667	0.9024
		(0.1567)	(0.1895)				
	LP	0.0112	2.9545	-392.4274	788.8548	0.1111	0.7658
		(0.0033)	(0.1496)				
	LD	0.0198	–	-394.5197	791.0394	0.1528	0.3701
		(0.0016)	–				
2 ($n = 72$)	PL	0.8451	0.0387	-396.8082	797.6164	0.1667	0.2700
		(0.0503)	(0.1745)				
	GL	1.1389	0.0212	-394.2822	792.5644	0.1528	0.3701
		(0.2101)	(0.0026)				

We fit the data sets with the Lindley–Poisson distribution $LP(\theta, \lambda)$, Lindley distribution $LD(\theta)$, Power Lindley distribution $PL(\alpha, \beta)$ and generalized Lindley distribution $GL(\alpha, \lambda)$ and examine the performances of the distributions.

Those probability density functions are given below:

$$PL: \qquad f(x|\Theta_1) = \frac{\alpha\beta^2}{\beta+1}(1+x^\alpha)x^{\alpha-1}e^{-\beta x^\alpha}, \quad \Theta_1 = (\alpha, \beta), \quad x > 0,$$

$$GL: \qquad f(x|\Theta_2) = \frac{\alpha\lambda^2}{1+\lambda}(1+x)[1 - \frac{1+\lambda+\lambda x}{1+\lambda}e^{-\lambda x}]^{\alpha-1}, \quad \Theta_2 = (\alpha, \lambda), \quad x > 0.$$

The maximum likelihood estimates of the parameters are obtained and the results are reported in Table 3. The Akaike information criterion (AIC) is computed to measure the goodness of fit of the models. $AIC = 2k - 2\log L$, where k is the number of parameters in the model and L is the maximized value of the likelihood function for the estimated model. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. The Kolmogorov-Smirnov (K-S) statistics and the p-values for these models are also presented. The K-S test compares an empirical and a theoretical model by computing the maximum absolute difference between the empirical and theoretical distribution functions: $D = \max_x |F_n(x) - F(x)|$. The associated the p-value is the chance that the value of the Komogorov-Smirnov D statistic would be as large or larger than observed. The computation of p-value can be found in [8].

For the first dataset, the K-S statistics for the LP and GL models are same and smaller than those for the LD and PL models. For the LP model, AIC=68.1532 is smaller than that obtained for the GL model. Log-likelihood value=-32.0766 is larger than those for the GL model. It indicates that the LP model performs a best fit for this dataset. The good performance of the LP model can also be supported by the second dataset.

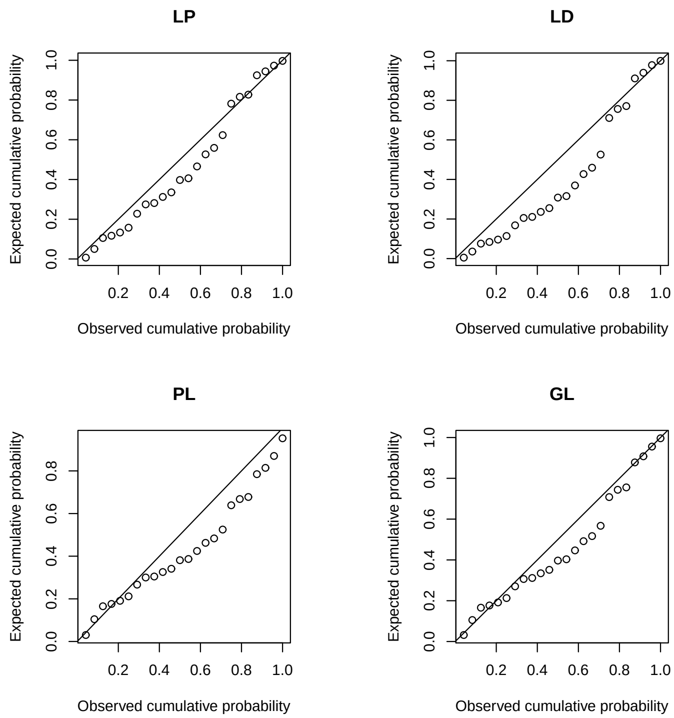


Figure 3. P-P plots for the first dataset.

Figure 3 and 4 display the probability-probability (P-P) plot for the two datasets.

6. Concluding Remarks

In this article, we have introduced a continuous Lindley-Poisson distribution by compounding the Lindley distribution and zero truncated Poisson distribution. The properties, including the shape properties of its density function and the hazard rate function, stochastic orderings, moments and measurements based on the moments are investigated. The distributions of some extreme order statistics are also derived. Maximum likelihood estimation method using EM algorithm is developed for estimating the parameters. Asymptotic properties of the MLEs are studied. We conduct intensive simulations and the results show that the estimation performance is satisfied as expected. We apply the model to two real datasets and the results demonstrate that the proposed model is appropriate for the datasets.

7. Acknowledgements

We thank the editor and two anonymous reviewers for their very valuable guidance and comments that helped us improve the manuscript. The first author's work was partially conducted while he was a postdoctorate researcher at Cornell University, supported

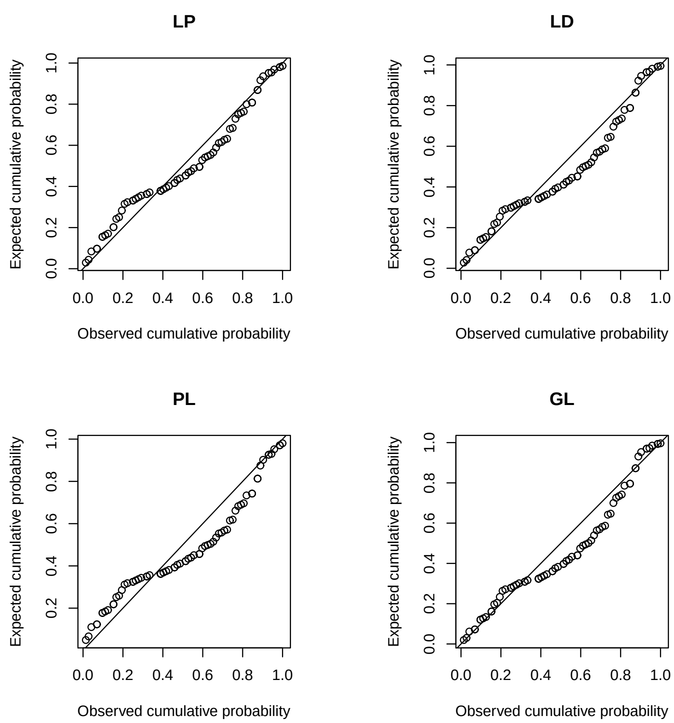


Figure 4. P-P plots for the second dataset.

by NSF-DMS-0808864, ONR-YIP-N000140910911, and a grant from Microsoft. The second author's work was partially supported by the Fundamental Research Funds for the Central Universities (2014JBM125).

References

[1] K Adamidis and S Loukas. A lifetime distribution with decreasing failure rate. *Statistics & Probability Letters*, 39(1):35–42, 1998.

[2] B.C. Arnold, N. Balakrishnan, and H.N. Nagaraja. *A first course in order statistics*, volume 54. Society for Industrial Mathematics, 2008.

[3] H.S. Bakouch, B.M. Al-Zahrani, A.A. Al-Shomrani, V.A.A. Marchi, and F. Louzada. An extended lindley distribution. *Journal of the Korean Statistical Society*, 41(1):75–85, 2012.

[4] T. Bjerkedal. Acquisition of resistance in guinea pies infected with different doses of virulent tubercle bacilli. *American Journal of Hygiene*, 72(1):130–148, 1960.

[5] M Chahkandi and M Ganjali. On some lifetime distributions with decreasing failure rate. *Computational Statistics & Data Analysis*, 53(12):4433–4440, 2009.

[6] D.R. Cox and D.V. Hinkley. *Theoretical statistics*. Chapman & Hall, London., 1979.

[7] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[8] William Feller. On the kolmogorov-smirnov limit theorems for empirical distributions. *The annals of mathematical statistics*, 19(2):177–189, 1948.

[9] ME Ghitany, DK Al-Mutairi, N Balakrishnan, and LJ Al-Enezi. Power lindley distribution and associated inference. *Computational Statistics & Data Analysis*, 64:20–33, 2013.

[10] ME Ghitany, DK Al-Mutairi, and S. Nadarajah. Zero-truncated poisson–lindley distribution and its application. *Mathematics and Computers in Simulation*, 79(3):279–287, 2008.

[11] ME Ghitany, B. Atieh, and S. Nadarajah. Lindley distribution and its application. *Mathematics and Computers in Simulation*, 78(4):493–506, 2008.

[12] Ronald E Glaser. Bathtub and related failure rate characterizations. *Journal of the American Statistical Association*, 75(371):667–672, 1980.

[13] Ramesh C Gupta, Nandini Kannan, and Aparna Raychaudhuri. Analysis of lognormal survival data. *Mathematical biosciences*, 139(2):103–115, 1997.

[14] P Jodrá. Computer generation of random variables with lindley or poisson–lindley distribution via the lambert W function. *Mathematics and Computers in Simulation*, 81(4):851–859, 2010.

[15] Coşkun Kuş. A new lifetime distribution. *Computational Statistics & Data Analysis*, 51(9):4497–4509, 2007.

[16] D.V. Lindley. Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(1):102–107, 1958.

[17] E. Mahmoudi and H. Zakerzadeh. Generalized poisson–lindley distribution. *Communications in Statistics Theory and Methods*, 39(10):1785–1798, 2010.

[18] Saralees Nadarajah, Hassan S Bakouch, and Rasool Tahmasbi. A generalized lindley distribution. *Sankhya B*, 73(2):331–359, 2011.

[19] C.G. Ramesh and S. Kirmani. On order relations between reliability measures. *Stochastic Models*, 3(1):149–156, 1987.

[20] M. Sankaran. The discrete poisson-lindley distribution. *Biometrics*, 26(1):145–149, 1970.

[21] Rama Shanker, Shambhu Sharma, and Ravi Shanker. A two-parameter lindley distribution for modeling waiting and survival times data. *Applied Mathematics*, 4:363–368, 2013.

[22] H. Zakerzadeh and A. Dolati. Generalized lindley distribution. *Journal of Mathematical Extension*, 3(2):13–25, 2009.