Sankhyā : The Indian Journal of Statistics
2017, Volume 79-B, Part 2, pp. 316-335

The Complementary Lindley-Geometric Distribution and Its Application in Lifetime Analysis

Wenhao Gui and Lei Guo
Beijing Jiaotong University, Beijing, China
Huainian Zhang
Beijing Institute of Petro-chemical Technology, Beijing, China

Abstract

In this paper, we propose a new compounding distribution, named the complementary Lindley-geometric distribution. It arises on a latent complementary risks scenarios where only the maximum lifetime value among all risks instead of a particular risk is observed. Its characterization and statistical properties are investigated. 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 a real application and it shows that the new distribution is appropriate and potential for lifetime analyses.

AMS (2000) *subject classification.* Primary 62E15; Secondary 62F10.
Keywords and phrases. Lindley distribution, Geometric distribution, Hazard function, Maximum likelihood estimation, EM algorithm, Fisher information matrix

1 Introduction

Lindley (1958) originally introduced the Lindley distributionto illustrate a difference between fiducial distribution and posterior distribution. This distribution has 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 $p = \theta/(1+\theta)$.

A comprehensive treatment of the statistical properties of the Lindley distribution was provided by Ghitany et al. (2008). Sankaran (1970) proposed the discrete Poisson–Lindley distribution. Ghitany et al. (2008) studied the properties of the zero-truncated Poisson—Lindley distribution. Bakouch et al. (2012) extended the Lindley distribution by exponentiation. Shanker et al. (2013) introduced a two-parameter Lindley distribution of which the one-parameter $LD(\theta)$ is a particular case, for modeling waiting and survival times data. Ghitany et al. (2013) proposed a two-parameter power Lindley distribution (PL) and discussed its properties. Nadarajah et al. (2011) introduced a generalized Lindley distribution (GL) and provided comprehensive account of the mathematical properties of the distribution.

Lifetime data analysis plays a critical role in a wide variety of scientific fields and there have been several distributions developed by compounding some useful life distributions to support analysis. Adamidis and Loukas (1998) introduced a two-parameter exponential-geometric (EG) distribution with decreasing failure rate by compounding an exponential with a geometric distribution. Kuṡ (2007) proposed an exponential-Poisson (EP) distribution by mixing an exponential and zero truncated Poisson distribution and discussed its various properties.

The aim of this paper is to propose an extension of the Lindley distribution which offers a more flexible distribution for modeling lifetime data based on a complementary risk problem (Basu and Klein, 1982) in presence of latent risks. Latent risks suggest that there is no information about which factor was responsible for the component failure but only the maximum lifetime value among all risks is observed. The extended distribution is a mixture of the Lindley and geometric distributions. 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 some kind. Z is a geometric variable. Their lifetimes, $Y's$, follow a Lindley distribution. Then for modeling the maximum failure X, the distribution leads to the complementary Lindley–geometric distribution.

The rest of this paper is organized as follows: in Section 2, we introduce the new Lindley–geometric 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 The Complementary Lindley-geometric Distribution and Its Properties

2.1. Density and Hazard Function 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 $X = \max(Y_1, ..., Y_Z)$.

Suppose the failure times of the initial defects $Y_1, Y_2, ..., Y_Z$ follow a Lindley distribution $LD(\theta)$ and Z is a geometric random variable with the probability mass function:

$$p(Z = z) = p^{z-1}(1-p), \quad 0 < p < 1, z = 1, 2, ... \tag{2.1}$$

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

$$f(x|z) = \theta^2(x+1)ze^{-\theta x}(\theta+1)^{-z}\left[\theta+1-e^{-\theta x}(\theta+\theta x+1)\right]^{z-1}, \quad x > 0,$$

and the marginal probability density function of X is

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

We denoted this by writing $LG(\theta, p)$. The parameter θ can be interpreted as a limit of the failure rate function, which is an important characteristic for lifetime models. The parameter $1/p$ represents the average number of initial defects.

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

1. *As p goes to zero, $LG(\theta, p)$ leads to the Lindley distribution $LD(\theta)$.*

2. *If $\theta \geq \sqrt{\frac{1-p}{1+p}}$, $f(x)$ is decreasing in x. If $\theta < \sqrt{\frac{1-p}{1+p}}$, $f(x)$ is a unimodal function at x_0, where x_0 is the solution of the equation* $p\left(\theta^2(x+1)^2+1\right) + (\theta+1)e^{\theta x}(\theta+\theta x-1) = 0$.

PROOF.

1. As p goes to zero, then

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

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

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

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

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

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

From Theorem 2.1, it can be seen that when p goes to zero, i.e., in the case of Lindley distribution, $f(x)$ is decreasing (unimodal) if $\theta \geq 1 (0 < \theta < 1)$ with $f(0) = \frac{\theta^2}{\theta+1}$ and $f(\infty) = 0$.

Figure 1a shows some density functions of the $LG(\theta, p)$ distribution with various parameters, displaying all the shapes established in Theorem 2.1.

The cumulative distribution of the LG distribution is given by

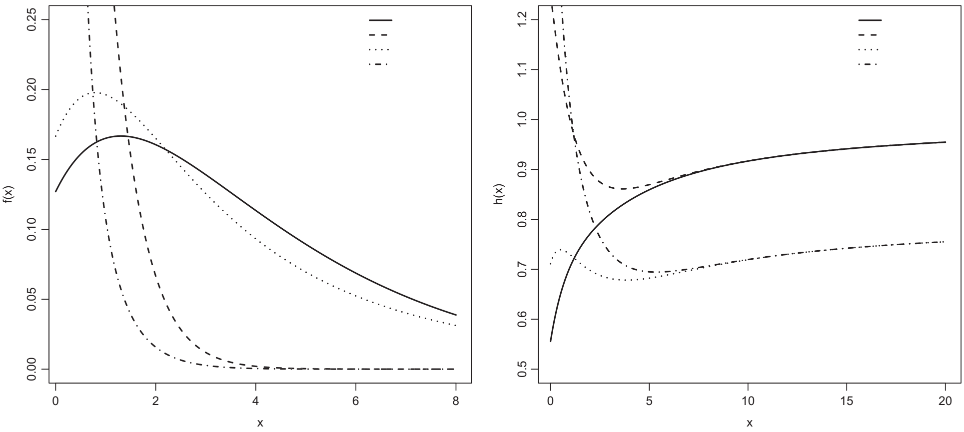
$$F(x) = \frac{(\theta+1)e^{\theta x} - \theta(x+1) - 1}{(\theta+1)e^{\theta x} - p(\theta+\theta x+1)}, \quad x > 0. \tag{2.3}$$

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

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

THEOREM 2.2. *Considering the hazard function of the LG distribution, we have the following properties:*

1. *If* $p < \frac{1}{1+\theta^2}$ *and the equation* $(\theta+1)e^{\theta x} - p\left[\theta^3(x+1)^3 + \theta^2(x+1)^2 + \theta (x+1) + 1\right] = 0$ *has no real roots, then the hazard function is increasing. If the equation* $(\theta+1)e^{\theta x} - p\left[\theta^3(x+1)^3 + \theta^2(x+1)^2 + \theta(x+1) +1\right] = 0$ *has one real root, the hazard function is increasing–decreasing–increasing shaped.*



(a) density function (b) hazard function

Figure 1: Plots of the LG density and hazard function for some parameter values

2. If $p \geq \frac{1}{1+\theta^2}$ and the equation $(\theta+1)e^{\theta x}-p\left[\theta^3(x+1)^3+\theta^2(x+1)^2+\theta(x+1)+1\right]=0$ has one real root, then the hazard function is decreasing–increasing shaped.

Proof. Notice that $h(0)=\frac{\theta^2}{(\theta+1)(1-p)}$, $h(\infty)=\theta$. The first derivative of the hazard function is

$$h'(x)=\frac{\theta^2(\theta+1)e^{\theta x}t(x)}{(\theta+\theta x+1)^2\left[(\theta+1)e^{\theta x}-p(\theta+\theta x+1)\right]^2},$$

where

$$t(x)=(\theta+1)e^{\theta x}-p\left[\theta^3(x+1)^3+\theta^2(x+1)^2+\theta(x+1)+1\right].$$

Here $t(0)=\theta+1-\left(\theta^3+\theta^2+\theta+1\right)p$ and $t(\infty)=\infty$. If $p<\frac{1}{1+\theta^2}$, $t(0)>0$ and the equation $(\theta+1)e^{\theta x}-p\left[\theta^3(x+1)^3+\theta^2(x+1)^2+\theta(x+1)+1\right]=0$ has no real roots, then $h'(x)>0$, the hazard function is increasing. If the equation has one real root, $h'(x)$ changes sign from positive to negative to positive, the hazard function is increasing–decreasing–increasing shaped.

If $p\geq\frac{1}{1+\theta^2}$, $t(0)\leq 0$ and the equation $(\theta+1)e^{\theta x}-p[\theta^3(x+1)^3+\theta^2(x+1)^2+\theta(x+1)+1]=0$ has one real root, then $h'(x)$ changes sign from negative to positive, the hazard function $h(x)$ is decreasing–increasing shaped.

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. Theorem 2.2 indicates the flexibility of the LG distribution over the Lindley and exponential distribution.

Figure 1b shows some shapes of the $LG(\theta, p)$ 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 usual 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 Ramesh and Kirmani (1987).

THEOREM 2.3. *If $X \sim LG(\theta, p_1)$ and $Y \sim LG(\theta, p_2)$, and $0 < p_1 < p_2 < 1$, then $Y \prec_{lr} X$, $Y \prec_{hr} X$ and $Y \prec_{st} X$.*

PROOF. The density ratio is given by

$$U(x) = \frac{f_X(x)}{f_Y(x)} = \frac{(1-p_1)\left((\theta+1)e^{\theta x} - p_2(\theta+\theta x+1)\right)^2}{(1-p_2)\left((\theta+1)e^{\theta x} - p_1(\theta+\theta x+1)\right)^2}$$

Taking the derivative with respect to x,

$$U'(x) = \frac{2\theta^2(\theta+1)(1-p_1)(p_2-p_1)(x+1)e^{\theta x}\left[(\theta+1)e^{\theta x} - p_2(\theta+\theta x+1)\right]}{(1-p_2)\left[(\theta+1)e^{\theta x} - p_1(\theta+\theta x+1)\right]^3}$$

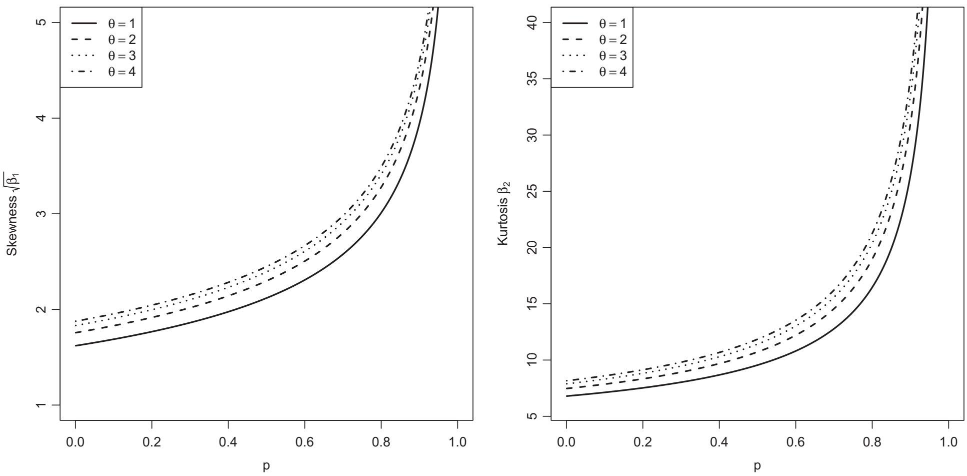
If $p_1 < p_2$, $U'(x) > 0$, $U(x)$ is an increasing function of x. The results follow.

2.3. Moments and Measures Based on Moments The kth raw moment of $X \sim LG(\theta, p)$ is given by, for $k = 1, 2, ...$

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

$\mathbb{E}(X^k)$ can not 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. 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 $\beta_2 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2-\mu_1^2)^2}$.

Figure 2a, b displays the skewness and kurtosis coefficients of the the $LG(\theta, p)$ distribution with various parameters. From the figures, it is found that the $LG(\theta, p)$ distribution has positive skewness and kurtosis coefficients. The coefficients are increasing as θ or p goes up.



(a) (b)

Figure 2: **a** Plot of skewness of the $LG(\theta, p)$ distribution. **b** Plot of kurtosis of the $LG(\theta, p)$ distribution

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

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

where $W(a)$ giving the principal solution for w in $a = we^w$ is pronounced as Lambert W function, see Jodrá (2010).

In particular, the median of the $LG(\theta, p)$ distribution is given by

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

3 Distributions of Order Statistics

Let $X_1, X_2, ..., X_n$ be a random sample of size n from the $LG(\theta, p)$ distribution. The sample 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 and maxima are defined as $X_{1:n} = \min(X_1, ..., X_n)$ and $X_{n:n} = \max(X_1, ..., X_n)$. These extreme order statistics represent the lifetimes of series and parallel system and have important applications in probability and statistics.

THEOREM 3.1. *Let $X_{1:n}$ and $X_{n:n}$ be the smallest and largest order statistics from the $LG(\theta, p)$ 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) $\lim_{n\to\infty} P(X_{n:n} \le a_n + b_n t) = \exp\{-e^{-t}\}, -\infty < t < \infty,$ *where* $a_n = F^{-1}(1 - 1/n), b_n = 1/[nf(a_n)]$.

PROOF. We apply the following asymptotical results for $X_{1:n}$ and $X_{n:n}$ (Arnold et al., 2008).

(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 $LG(\theta, p)$ distribution, its cumulative distribution function is

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

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

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

(of the Gumbel type) where $a_n = F^{-1}(1 - 1/n)$ and $b_n = 1/[nf(a_n)]$ if

$$\lim_{x\to F^{-1}(1)} \frac{d}{dx}\left\{\frac{1}{h(x)}\right\} = 0.$$

For the $LG(\theta, p)$ distribution, let $F(x)=1$, then $x=\infty$. Furthermore,

$$\begin{aligned}
&\lim_{x\to F^{-1}(1)} \frac{d}{dx}\left\{\frac{1}{h(x)}\right\} \\
=&\lim_{x\to F^{-1}(1)} \frac{e^{-\theta x}\left\{p\left[\theta^3(x+1)^3+\theta^2(x+1)^2+\theta(x+1)+1\right]-(\theta+1)e^{\theta x}\right\}}{\theta^2(\theta+1)(x+1)^2} \\
=&\ 0.
\end{aligned}$$

The statement follows.

REMARK 3.2. 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 Arnold et al. (2008),

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

4 Estimation and Inference

4.1. Maximum Likelihood Estimation In this section, we consider the maximum likelihood estimation about the parameters (θ, p) of the LG model. Suppose $y_{obs}=\{x_1,x_2,...,x_n\}$ is a random sample of size n from the $LG(\theta,p)$ distribution. Then the log-likelihood function is given by

$$\begin{aligned}
l &= \log\prod_{i=1}^n f_X(x_i) \\
&= -2\sum_{i=1}^n \log\left[(\theta+1)e^{\theta x_i}-p\left(\theta+\theta x_i+1\right)\right]+\theta\sum_{i=1}^n x_i+\sum_{i=1}^n\log\left(x_i+1\right) \\
&\quad +2n\log(\theta)+n\log(\theta+1)+n\log(1-p) \qquad (4.1)
\end{aligned}$$

The associated gradients are found to be

$$\frac{\partial l}{\partial\theta} = -2\sum_{i=1}^n\frac{-p\left(x_i+1\right)+(\theta+1)x_ie^{\theta x_i}+e^{\theta x_i}}{(\theta+1)e^{\theta x_i}-p\left(\theta+\theta x_i+1\right)}+\sum_{i=1}^n x_i+\frac{2n}{\theta}+\frac{n}{\theta+1} \qquad (4.2)$$

$$\frac{\partial l}{\partial p} = 2\sum_{i=1}^n\frac{\theta+\theta x_i+1}{(\theta+1)e^{\theta x_i}-p\left(\theta+\theta x_i+1\right)}-\frac{n}{1-p} \qquad (4.3)$$

The estimates of the parameters maximize the likelihood function. Equalizing the obtained gradients expressions to zero yield to 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.

In the following, we present when one parameter is given, what the conditions are there to achieve the existence and uniqueness of the MLE.

THEOREM 4.1. *Let* $l_i(\theta, p, y_{obs}), i = 1, 2$ *denote the right hand side (RHS) of* Eqs. (4.2)–(4.3)*, respectively, and let* $\tilde{x} = (1/n)\sum_{i=1}^{n} x_i^{\alpha}$*, then the following properties hold:*

(1) If p *is known, then the root of* $l_1(\theta, p, y_{obs}) = 0$*,* $\hat{\theta}$*, lies in the interval* $(0, \infty)$*.*

(2) If θ *is known, then for* $\sum_{i=1}^{n} e^{-\theta x_i}(\theta + \theta x_i + 1) > \frac{n(\theta+1)}{2}$*, the equation* $l_2(\theta, p, y_{obs}) = 0$ *has at least one root.*

PROOF.

(1) Since $\lim_{\theta\to 0} l_1(\theta, p, y_{obs}) = \infty$, $\lim_{\theta\to\infty} l_1(\theta, p, y_{obs}) = -\sum_{i=1}^{n} x_i < 0$. Therefore, there exists at least one root of $l_1(\theta, p, y_{obs}) = 0$ in the interval $(0, \infty)$. Thus, there exists a root in the interval $(0, \infty)$.

(2) Notice that $\lim_{p\to 0} l_2(\theta, p, y_{obs}) = 2\sum_{i=1}^{n} \frac{e^{-\theta x_i}(\theta+\theta x_i+1)}{\theta+1} - n > 0$ when $\sum_{i=1}^{n} e^{-\theta x_i}(\theta + \theta x_i + 1) > \frac{n(\theta+1)}{2}$. On the other hand, we can show that $\lim_{p\to 1} l_2(\theta, p, y_{obs}) = -\infty$. Therefore, there is at least one root of $l_2(\theta, p, y_{obs}) = 0$.

4.2. An EM Algorithm An expectation–maximization (EM) algorithm (Dempster et al., 1977) 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

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

where $\theta > 0$ and $0 < p < 1$ are parameters. It is straightforward to verify that the computation of the conditional expectation of $(Z|X)$ using the pdf

$$p(z|x) = zp^{z-1}(\theta+1)^{-z-1}\left[(\theta+1)e^{\theta x}-p(\theta+\theta x+1)\right]^2 e^{-\theta x(z+1)}(\theta+\theta x+1)^{z-1},$$
$$z = 1,2,...$$

Then we have

$$\mathbb{E}(Z|X) = 1+\frac{2p(\theta+\theta x+1)}{(\theta+1)e^{\theta x}-p(\theta+\theta x+1)}$$

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

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

$$p^{(t+1)} = 1-\frac{n}{\sum_{i=1}^{n}w_i^{(t)}}$$

where $w_i^{(t)} = 1+\frac{2p^{(t)}(\theta^{(t)}+\theta^{(t)}x_i+1)}{(\theta^{(t)}+1)e^{\theta^{(t)}x_i}-p^{(t)}(\theta^{(t)}+\theta^{(t)}x_i+1)}$.

4.3. Asymtotic Variance and Covariance of MLEs It is known that under some mild regular conditions, as the sample size increases, the distribution of the MLE tends to the bivariate normal distribution with mean (θ, p) and covariance matrix equal to the inverse of the Fisher information matrix, see Cox and Hinkley D (1979). The bivariate normal distribution can be used to construct approximate confidence intervals for the parameters θ and p.

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

$$
\begin{aligned}
I_{11} &= 2\sum_{i=1}^{n} \frac{(\theta+1)x_i^2 e^{\theta x_i} + 2x_i e^{\theta x_i}}{(\theta+1)e^{\theta x_i} - p\left(\theta + \theta x_i + 1\right)} \\
&\quad -2\sum_{i=1}^{n} \frac{\left[-p\left(x_i+1\right) + (\theta+1)x_i e^{\theta x_i} + e^{\theta x_i}\right]^2}{\left[(\theta+1)e^{\theta x_i} - p\left(\theta + \theta x_i + 1\right)\right]^2} + \frac{2n}{\theta^2} + \frac{n}{(\theta+1)^2} \\
l_{12} &= l_{21} = 2\sum_{i=1}^{n} \frac{\left(\theta + \theta x_i + 1\right)\left[-p\left(x_i+1\right) + (\theta+1)x_i e^{\theta x_i} + e^{\theta x_i}\right]}{\left[(\theta+1)e^{\theta x_i} - p\left(\theta + \theta x_i + 1\right)\right]^2} \\
&\quad -2\sum_{i=1}^{n} \frac{x_i + 1}{(\theta+1)e^{\theta x_i} - p\left(\theta + \theta x_i + 1\right)} \\
l_{22} &= \frac{n}{(1-p)^2} - 2\sum_{i=1}^{n} \frac{\left(\theta + \theta x_i + 1\right)^2}{\left[(\theta+1)e^{\theta x_i} - p\left(\theta + \theta x_i + 1\right)\right]^2}
\end{aligned}
$$

The expectation $J = \mathbb{E}(I(\theta, p; y_{obs}))$ 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

$$
\begin{aligned}
J_{11} &= 2\mathbb{E}\left\{ \frac{(\theta+1)X^2 e^{\theta X} + 2Xe^{\theta X}}{(\theta+1)e^{\theta X} - p\left(\theta + \theta X + 1\right)} \right\} \\
&\quad -2\mathbb{E}\left\{ \frac{\left[-p\left(X+1\right) + (\theta+1)Xe^{\theta X} + e^{\theta X}\right]^2}{\left[(\theta+1)e^{\theta X} - p\left(\theta + \theta X + 1\right)\right]^2} \right\} + \frac{2}{\theta^2} + \frac{1}{(\theta+1)^2} \\
J_{12} &= J_{21} = 2\mathbb{E}\left\{ \frac{\left(\theta + \theta X + 1\right)\left[-p\left(X+1\right) + (\theta+1)Xe^{\theta X} + e^{\theta X}\right]}{\left[(\theta+1)e^{\theta X} - p\left(\theta + \theta X + 1\right)\right]^2} \right\} \\
&\quad -2\mathbb{E}\left\{ \frac{X+1}{(\theta+1)e^{\theta X} - p\left(\theta + \theta X + 1\right)} \right\} \\
J_{22} &= \frac{1}{(1-p)^2} - 2\mathbb{E}\left\{ \frac{\left(\theta + \theta X + 1\right)^2}{\left[(\theta+1)e^{\theta X} - p\left(\theta + \theta X + 1\right)\right]^2} \right\}
\end{aligned}
$$

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 $Z \sim$ Geometric (p).

(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, p), 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 $Var(\hat{\theta})$, $Var(\hat{p})$ and $Cov(\hat{\theta}, \hat{p})$ 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.

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 LG distribution for each of the four values of (θ, p) are generated.

Table 1: Variances and covariances of the MLEs

n	(θ, p)	Simulated			From expected information			From observed information		
	$Var(\hat{\theta})$	$Var(\hat{p})$	$Cov(\hat{\theta}, \hat{p})$	$Var(\hat{\theta})$	$Var(\hat{p})$	$Cov(\hat{\theta}, \hat{p})$	$Var(\hat{\theta})$	$Var(\hat{p})$	$Cov(\hat{\theta}, \hat{p})$	
50	(1, 0.2)	0.0320	0.0496	− 0.0317	0.0706	0.1789	− 0.0986	0.0704	0.1781	− 0.0983
50	(2, 0.2)	0.1444	0.0495	− 0.0652	0.2706	0.1465	− 0.1713	0.2822	0.1519	− 0.1790
50	(1, 0.5)	0.0665	0.0570	− 0.0529	0.1001	0.1166	− 0.0935	0.1013	0.1178	− 0.0947
50	(2, 0.5)	0.2778	0.0530	− 0.1022	0.3794	0.0968	− 0.1633	0.4087	0.1036	− 0.1764
100	(1, 0.2)	0.0193	0.0354	− 0.0212	0.0348	0.0980	− 0.0518	0.0351	0.0981	− 0.0521
100	(2, 0.2)	0.0713	0.0312	− 0.0371	0.1343	0.0847	− 0.0931	0.1393	0.0877	− 0.0969
100	(1, 0.5)	0.0405	0.0415	− 0.0360	0.0493	0.0583	− 0.0472	0.0505	0.0595	− 0.0483
100	(2, 0.5)	0.1755	0.0359	− 0.0690	0.1890	0.0475	− 0.0822	0.2036	0.0512	− 0.0893
500	(1, 0.2)	0.0050	0.0136	− 0.0071	0.0069	0.0212	− 0.0109	0.0070	0.0215	− 0.0110
500	(2, 0.2)	0.0207	0.0122	− 0.0135	0.0270	0.0182	− 0.0196	0.0281	0.0190	− 0.0205
500	(1, 0.5)	0.0092	0.0096	− 0.0085	0.0097	0.0102	− 0.0090	0.0100	0.0106	− 0.0094
500	(2, 0.5)	0.0399	0.0088	− 0.0170	0.0367	0.0083	− 0.0156	0.0398	0.0091	− 0.0171
1000	(1, 0.2)	0.0028	0.0079	− 0.0041	0.0035	0.0107	− 0.0055	0.0035	0.0108	− 0.0056
1000	(2, 0.2)	0.0120	0.0074	− 0.0082	0.0135	0.0091	− 0.0098	0.0141	0.0096	− 0.0103
1000	(1, 0.5)	0.0046	0.0047	− 0.0042	0.0048	0.0049	− 0.0044	0.0050	0.0051	− 0.0046
1000	(2, 0.5)	0.0195	0.0041	− 0.0081	0.0183	0.0040	− 0.0076	0.0198	0.0044	− 0.0084

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

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

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

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

Table 2: The means and standard errors of the EM estimator and iterations to convergence with initial values $(\theta^{(0)}, p^{(0)})$ 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	1	0.2	1	0.2	0.967	0.255	0.140	0.191	208.773	77.608
100	1	0.5	1	0.5	1.037	0.443	0.198	0.209	177.110	89.455
100	2	0.2	2	0.2	1.909	0.268	0.276	0.187	204.285	69.019
100	2	0.5	2	0.5	2.043	0.463	0.408	0.191	182.748	77.270
100	1	0.2	2	0.4	0.972	0.249	0.137	0.185	210.988	83.786
100	1	0.5	2	0.7	1.025	0.455	0.203	0.206	176.130	83.547
100	2	0.2	3	0.4	1.941	0.255	0.286	0.182	207.890	88.546
100	2	0.5	3	0.7	2.059	0.462	0.425	0.195	188.137	81.998
500	1	0.2	1	0.2	0.999	0.194	0.071	0.117	194.091	85.388
500	1	0.5	1	0.5	1.007	0.486	0.096	0.100	123.716	53.958
500	2	0.2	2	0.2	1.994	0.204	0.143	0.109	178.555	78.897
500	2	0.5	2	0.5	2.014	0.489	0.198	0.093	136.151	45.783
500	1	0.2	2	0.4	0.998	0.197	0.070	0.117	197.712	99.443
500	1	0.5	2	0.7	1.006	0.487	0.094	0.098	119.841	53.573
500	2	0.2	3	0.4	1.993	0.204	0.149	0.114	201.703	90.034
500	2	0.5	3	0.7	2.017	0.487	0.207	0.099	152.584	49.632

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.

5 Illustrative Examples

In this section, we consider a numerical application to test the performance of the new distribution. 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 Kuṡ (2007).

We fit the data set with the Lindley–geometric distribution $LG(\theta, p)$, Lindley distribution $LD(\theta)$, Power Lindley distribution $PL(\alpha, \beta)$, Generalized Lindley distribution $GL(\alpha, \lambda)$, Complementary Exponential-geometric distribution $CEG(\lambda, \theta)$, Exponential-Poisson distribution $EP(\lambda, \beta)$, Weibull distribution $WE(\lambda, \beta)$ and Exponential distribution $EXP(\lambda)$ and examine the performances of the distributions.

Those probability density functions are given below:

$$\begin{aligned}
PL: &\quad f(x|\Theta_1) = \frac{\alpha\beta^2}{\beta+1}(1+x^\alpha)x^{\alpha-1}e^{-\beta x^\alpha}, \quad \Theta_1 = (\alpha, \beta), \quad x>0, \\
GL: &\quad f(x|\Theta_2) = \frac{\alpha\lambda^2}{1+\lambda}(1+x)[1-\frac{1+\lambda+\lambda x}{1+\lambda}e^{-\lambda x}]^{\alpha-1}, \quad \Theta_2 = (\alpha, \lambda), \\
&\quad x>0 \\
CEG: &\quad f(x|\Theta_3) = \frac{\lambda\theta e^{-\lambda x}}{[e^{-\lambda x}(1-\theta)+\theta]^2}, \quad \Theta_3 = (\lambda, \theta), \quad x>0 \\
EP: &\quad f(x|\Theta_4) = \frac{\lambda\beta}{1-e^{-\lambda}}e^{-\lambda-\beta x+\lambda\exp(-\beta x)}, \quad \Theta_4 = (\lambda, \beta), \quad x>0 \\
WE: &\quad f(x|\Theta_5) = \beta\lambda^\beta x^{\beta-1}e^{-(\lambda x)^\beta}, \quad \Theta_5 = (\lambda, \beta), \quad x>0 \\
EXP: &\quad f(x|\Theta_6) = \lambda e^{-\lambda x}, \quad \Theta_6 = \lambda, \quad x>0.
\end{aligned}$$

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 Feller (1948).

From Table 3, for the earthquake dataset, AIC displays that LG model is a best fit. It has the smallest AIC and the highest likelihood values. The K-S statistic takes the smallest value and the largest p value under the LG model. Figure 3 displays the probability-probability (P-P) plot for the earthquake dataset.

Table 3: Maximum likelihood parameter estimates(with (SE)) of the LG, LD, PL, GL, CEG, EP, WE and EXP models for the earthquake dataset

Model	Estimations		loglik	AIC	K-S statistic	p value
LG	0.3736	0.9096	− 30.7748	65.5496	0.0833	0.9999
	(0.3157)	(0.1387)				
LD	1.0420	–	− 34.5092	71.0184	0.2500	0.4490
	(0.1612)	–				
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)				
CEG	0.7024	0.99	− 32.6091	69.2182	0.2083	0.6860
	(0.1882)	(0.4034)				
EP	0.3368	0.6994	− 30.9988	65.9976	0.1667	0.9024
	(0.2298)	(0.1427)				
WE	0.8125	0.7854	− 31.1879	66.3758	0.1250	0.9242
	(0.2232)	(0.1214)				
EXP	0.6994	–	− 32.5778	69.1556	0.2083	0.6860
	(0.1427)	–				

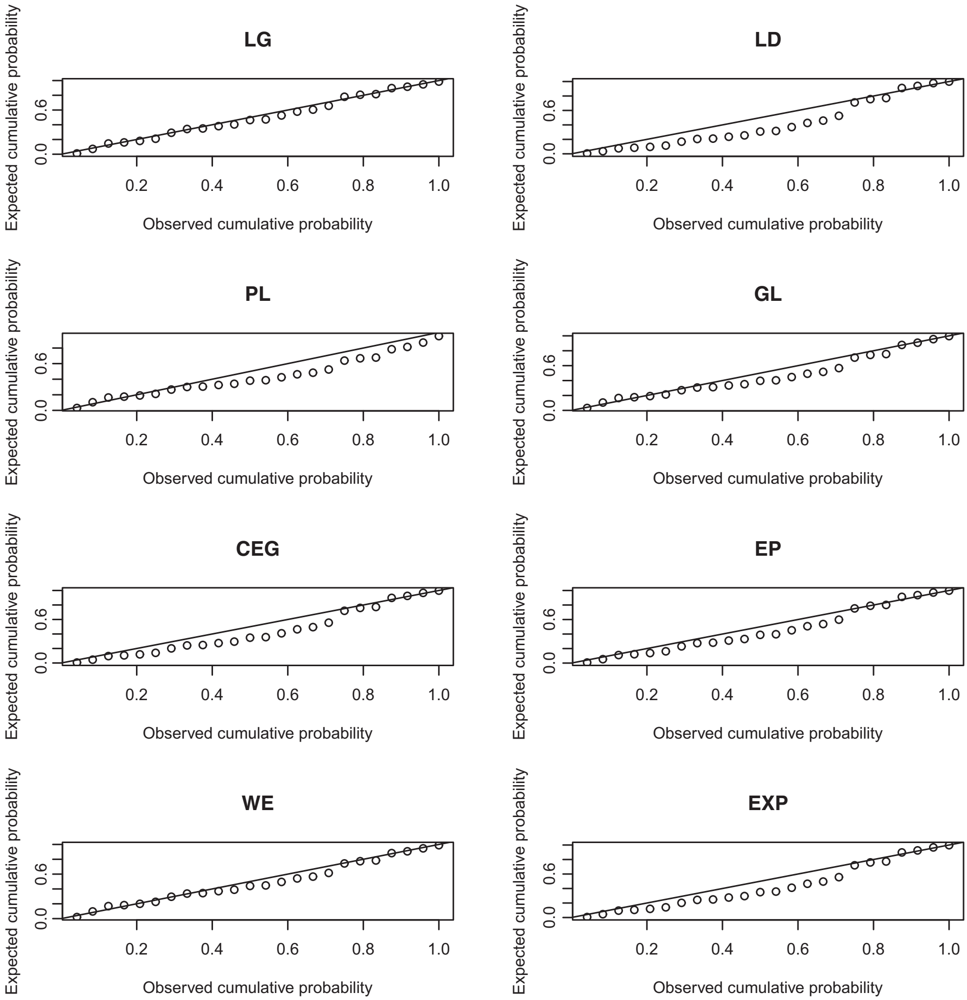


Figure 3: P-P plots for the earthquake dataset

6 Concluding Remarks

In this article, we have introduced a continuous Lindley-geometric distribution by compounding the Lindley distribution and geometric 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.

Acknowledgements. Gui's work was partially supported by the program for the Fundamental Research Funds for the Central Universities (NO: 2014 RC042; 2015JBM109). The authors would like to thank the referees and Editor for their helpful suggestions.

References

ADAMIDIS, K. and LOUKAS, S. (1998). A lifetime distribution with decreasing failure rate. *Stat. Probab. Lett.* **39**, 1, 35–42.

ARNOLD, B., BALAKRISHNAN, N. and NAGARAJA, H. (2008). *A first course in order statistics*, Vol. 54. Society for Industrial Mathematics.

BAKOUCH, H., AL-ZAHRANI, B., AL-SHOMRANI, A., MARCHI, V. and LOUZADA F (2012). An extended Lindley distribution. *J. Korean Stat. Soc.* **41**, 1, 75–85.

BASU, A. P. and KLEIN, J. P. (1982). Some recent results in competing risks theory. *Lect. Notes-Monogr. Series* **2**, 216–229.

COX, D. and HINKLEY D (1979). *Theoretical statistics.* Chapman & Hall/CRC, London.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Socy. Series B (Methodological)* **39**, 1–38.

FELLER, W. (1948). On the kolmogorov-smirnov limit theorems for empirical distributions. *Ann. Math. Stat.* **19**, 2, 177–189.

GHITANY, M., AL-MUTAIRI, D. and NADARAJAH, S (2008). Zero-truncated poisson–lindley distribution and its application. *Math. Comput. Simul.* **79**, 3, 279–287.

GHITANY, M., ATIEH, B. and NADARAJAH, S (2008). Lindley distribution and its application. *Math. Comput. Simul.* **78**, 4, 493–506.

GHITANY, M., AL-MUTAIRI, D., BALAKRISHNAN, N and AL-ENEZI, L (2013). Power Lindley distributionand associated inference. *Comput. Stat. Data Anal.* **64**, 20–33.

JODRÁ, P. (2010). Computer generation of random variables with Lindley or poisson–Lindley distribution via the Lambert W function. *Math. Comput. Simul.* **81**, 4, 851–859.

KUṠ, C. (2007). A new lifetime distribution. *Comput. Stat. Data Anal.* **51**, 9, 4497–4509.

LINDLEY, D. (1958). Fiducial distributions and Bayes' theorem. *J. R. Stat. Soc. Series B (Methodological)* **20**, 1, 102–107.

NADARAJAH, S., BAKOUCH, H. S. and TAHMASBI, R (2011). A generalized Lindley distribution. *Sankhya B* **73**, 2, 331–359.

RAMESH, C. and KIRMANI, S (1987). On order relations between reliability measures. *Stoch. Models* **3**, 1, 149–156.

SANKARAN, M. (1970). The discrete poisson-Lindley distribution. *Biometrics* **26**, 1, 145–149.

SHANKER, R., SHARMA, S. and SHANKER, R (2013). A two-parameter Lindley distribution for modeling waiting and survival times data. *Appl. Math.* **4**, 363–368.

Wenhao Gui
Lei Guo
Department of Mathematics,
Beijing Jiaotong University,
Beijing 100044, China
E-mail: whgui@bjtu.edu.cn

Huainian Zhang
Department of Mathematics &
Physics,
Beijing Institute of Petro-chemical
Technology,
Beijing 102617, China

Paper received: 24 January 2016.