STATISTICS IN TRANSITION new series, September 2022
Vol. 23 No. 3, pp. 95–112, DOI 10.2478/stattrans-2022-0032
Received – 31.01.2020; accepted – 31.01.2022

New polynomial exponential distribution: properties and applications

Abdelfateh Beghriche[1], Halim Zeghdoudi[2], Vinoth Raman[3], Sarra Chouia[4]

ABSTRACT

The study describes the general concept of the XLindley distribution. Forms of density and hazard rate functions are investigated. Moreover, precise formulations for several numerical properties of distributions are derived. Extreme order statistics are established using stochastic ordering, the moment method, the maximum likelihood estimation, entropies and the limiting distribution. We demonstrate the new family's adaptability by applying it to a variety of real-world datasets.

Key words: exponential distribution, Xgamma distribution, Lindley distribution, quantile function stochastic ordering, maximum-likelihood estimation, XLindley distribution.

1. Introduction

Statistical models can be used to describe and predict real-world events. In recent years, a variety of distributions have been employed for data modelling in a variety of domains. Recent advances have centred on establishing new families that extend well-known distributions while still allowing for a great deal of flexibility in data modelling in practice. Several distributions have been proposed in the statistical literature to modify lifetime data, including the Lindley, exponential, gamma, Weibull, Zeghdoudi, and Xgamma distributions.

[1] Department of Mathematics, Faculty of Exact Sciences, University the Brothers Mentouri Constantine 1, 25019, Algeria. E-mail: b.abdelfateh1@yahoo.fr.

[2] Laps laboratory, Badji-Mokhtar University, Box 12, Annaba, 23000, Algeria.
E-mail: zeghdoudihalim@yahoo.fr. ORCID: https://orcid.org/0000-0002-4759-5529.

[3] Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Kingdom of Saudi Arabia.
E-mail: vrrangan@iau.edu.sa. ORCID: https://orcid.org/ 0000-0002-3815-2312.

[4] Laps laboratory, Badji-Mokhtar University, Box 12, Annaba, 23000, Algeria.
E-mail: chouiasarra3@gmail.com.

In this paper, we investigate a new polynomial exponential family that includes the distributions of XLindley and Xgamma, as well as Zeghdoudi as special instances, to introduce a new family of single-parameter continuous distributions. The existing literature on modelling survival data, biological sciences, and actuarial sciences will benefit from this new family of distributions.

Assume X is a random variable with values in the range [0, ∞], and the distribution of X depends on an indeterminate parameter θ with values in the range [0, ∞]. The distribution of X can be absolutely continuous or discrete. The distribution of X is a new one-parameter polynomial exponential family and the probability density function is expressed as

$$f_{NPED}(\mathrm{x}, \theta) = \frac{P(x,\theta)e^{-\theta x}}{\sum_{k=0}^{n} a_{k,\theta}\frac{k!}{\theta^{k+1}}}; \quad x, \theta > 0 \tag{1}$$

where $P(x,\theta) = \sum_{k=0}^{n} a_{k,\theta}x^k$, and $a_{k,\theta}$ depend on k and θ.

The following is the format of this research paper:

Section 2 covers the survival and hazard rate functions, moments stochastic orders, mean deviations, extreme domain of attraction, constraint force estimate parameter, the Lorenz curve, and entropies of the new polynomial exponential distribution (NPED). Sections 3 and 4 look at estimating maximum likelihood distribution parameters and inferring a random sample from the XLindley and Xgamma distributions. Finally, various real-world applications demonstrate the superior performance of the XLindley and Xgamma distributions, two special examples of the (NPED) family, as compared to the exponential, Lindley, Zeghdoudi, and exponential distributions.

2. Statistical and reliability measures of some properties of NPED distribution

We present some key statistical and reliability measures, as well as various NPED features, in this section.

2.1. Density and distribution functions

The first derivative of f_{NPED}:

$$\frac{d}{dx}f_{NPED}(\mathrm{x}, \theta) = \frac{[(a_{1,\theta}-\theta a_{0,\theta})+\cdots+(na_{n,\theta}-\theta a_{n-1,\theta})x^{n-1}+a_{n,\theta}x^n]e^{-\theta x}}{\sum_{k=0}^{n} a_{k,\theta}\frac{k!}{\theta^{k+1}}} = 0 \tag{2}$$

gives $x_1, x_2, \ldots, x_n$ solutions.

The NPED cumulative distribution function (CDF) is derived in (3).

$$F_{NPED}(\mathrm{x}) = 1 - \frac{\sum_{k=0}^{n} \frac{a_{k,\theta}\Gamma(k+1,x\theta)}{\theta^{k+1}}}{\sum_{k=0}^{n} a_{k,\theta}\frac{k!}{\theta^{k+1}}}; \; x, \theta > 0 \tag{3}$$

2.2. Survival and hazard rate functions

$$S_{NPED}(\mathrm{x}) = 1 - F_{NPED}(\mathrm{x}) = \frac{\sum_{k=0}^{n} \frac{a_{k,\theta}\Gamma(k+1,x\theta)}{\theta^{k+1}}}{\sum_{k=0}^{n} a_{k,\theta}\frac{k!}{\theta^{k+1}}} \;;\; x, \theta > 0 \tag{4}$$

$$h_{NPED}(\mathrm{x}) = \frac{f_{NPED}(x)}{1-F_{NPED}(x)} = \frac{\sum_{k=0}^{n} a_{k,\theta} x^k e^{-\theta x}}{\sum_{k=0}^{n} \frac{a_{k,\theta}\Gamma(k+1,x\theta)}{\theta^{k+1}}} \;;\; x, \theta > 0 \tag{5}$$

Let equation (4) and (5) be the survival and hazard rate function, respectively.

Proposition 1. Let $h_\theta(\mathrm{x})$be the hazard rate function of X . Then, $h_\theta(\mathrm{x})$ is increasing for:

$$\sum_{k=0}^{n} (k+1)(m-2k)a_{m-k,\theta}a_{k+1,\theta} \geq 0, m = 0, \dots\dots, 2n-1.$$

Proof. According to Glaser (1980) and from the density function (2) we have:

$$\rho(x) = -\frac{f'_{NPED}(x;\theta)}{f_{NPED}(x;\theta)} = -\frac{\sum_{k=1}^{n} k a_{k,\theta} x^{k-1}}{\sum_{k=0}^{n} a_{k,\theta} x^k} + \theta. \tag{6}$$

After simple computations, we obtain:

$$\rho'(x) = \frac{\sum_{m=0}^{2n}\sum_{k=0}^{m}(k+1)(m-2k)a_{m-k,\theta}a_{k+1,\theta}x^{m-1}}{(\sum_{k=0}^{n} a_{k,\theta}x^k)^2} + \theta \tag{7}$$

Which implies that $h_\theta(\mathrm{x})$is increasing for:

$$\sum_{k=0}^{n} (k+1)(m-2k)a_{m-k,\theta}a_{k+1,\theta} \geq 0, m = 0, \dots\dots, 2n-1$$

2.3. Moments and related measures

The k^{th} moment about the origin of $NPED$ is:

$$E(X^i) = \frac{\sum_{k=0}^{n} \frac{a_{k,\theta}(k+i+1)!}{\theta^{k+i+1}}}{\sum_{k=0}^{n} a_{k,\theta}\frac{k!}{\theta^{k+1}}}; i = 1,2, \dots. \tag{8}$$

Corollary 1. Let $X \sim NPED(\theta)$, the mean of X is:

$$E(X) = \frac{\sum_{k=0}^{n} \frac{a_{k,\theta}(k+1)!}{\theta^{k+2}}}{\sum_{k=0}^{n} a_{k,\theta}\frac{k!}{\theta^{k+1}}} \;. \tag{9}$$

Theorem 1. Let $X \sim NPED(\theta)$, $me = median(X)$and $\mu = E(X)$. Then, $me < \mu$.

Proof. According to the increasing of $F(X)$ for all x and θ.

$$F_{NPED}(me) = \frac{1}{2}$$

and

$$F_{NPED}(\mu) = 1 - h(\theta)\sum_{k=0}^{n} \frac{a_{k,\theta}\Gamma\left(k+1, \theta h(\theta)\sum_{k=0}^{n} a_{k,\theta}\frac{(k+1)!}{\theta^{k+2}}\right)}{\theta^{k+1}}.$$

Note that $\frac{1}{2} < F(\mu) < 1$. It is easy to check that $F(me) < F(\mu)$. At the other end we have $me < \mu$.

2.4. Stochastic orders

Definition 1. Consider two random variables X and Y. X is said to be smaller than Y in the:

a) Stochastic order $X \prec_S Y$ if $F_X(t) \geq F_Y(t), \forall t$.

b) Convex order $X \prec_{CX} Y$ Nif for all convex functions Φ and provided expectation exist, $E[\Phi(X)] \leq E[\Phi(Y)]$.

c) Hazard rate order $X \prec_{hr} Y$, if $h_X(t) \geq h_Y(t), \forall t$.

d) Likelihood ratio order $X \prec_{lr} Y$, if $\frac{f_X(t)}{f_Y(t)}$ is decreasing in t.

Remark 1. Likelihood ratio order⇒Hazard rate order⇒Stochastic order.

If E (X) = E (Y), then convex order⇔stochastic order.

Theorem 2. Let $X_i \sim NPED(\theta_i) i = 1,2$ be two random variables. If $\theta_1 \geq \theta_2$, then $X_1 \prec_{lr} X_2$,$X_1 \prec_{hr} X_2$,$X_1 \prec_S X_2$.

Proof. We have:

$$\frac{f_{X_1}(t)}{f_{X_2}(t)} = \frac{\sum_{k=0}^{n} a_{k,\theta}\frac{(k+1)!}{\theta_2{}^{k+2}}}{\sum_{k=0}^{n} a_{k,\theta}\frac{(k+1)!}{\theta_1{}^{k+2}}} e^{-(\theta_1-\theta_2)}. \tag{11}$$

For simplification, we use $\ln(\frac{f_{X_1}(t)}{f_{X_2}(t)})$. Now, we can find

$$\frac{d}{dt}\ln\left(\frac{f_{X_1}(t)}{f_{X_2}(t)}\right) = -(\theta_1 - \theta_2).$$

To this end, if $\theta_1 \geq \theta_2$,, we have $\frac{d}{dt}\ln\left(\frac{f_{X_1}(t)}{f_{X_2}(t)}\right) \leq 0$. This means that $X_1 \prec_{lr} X_2$. Also, according to Remark 1 the theorem is proved.

2.5. Mean deviations

These are two mean deviations: about Mean and Median, defined as: $MD_1 = \int_0^\infty |x-\mu| f(x)dx$ and $MD_2 = \int_0^\infty |x - me| f(x)dx$ respectively, where $\mu = E(X)$ and $me = Median(X)$.

The measures MD_1 and MD_2 can be computed using the following simplified formulas:

$$MD_1 = 2\mu F(\mu) - 2\int_0^{\mu} xf(x)\,dx$$
$$MD_2 = \mu - 2\int_0^{me} xf(x)\,dx$$

2.6. Extreme domain of attraction

As to the extreme value stability, the $F_{NPED}$is in the Gumbel extreme value domain of attraction, that is, there exist two sequences $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$of real numbers such that for any $x\epsilon R$, we have

$$\lim_{x\to+\infty} P\left(\frac{M_n - b_{n,\theta}}{a_{n,\theta}} \leq x\right) = \lim_{x\to+\infty} F_{NPED}\left(a_{n,\theta}x + b_{n,\theta}\right)^n = e^{(-e^{-x})} \quad (12)$$

This follows from Formula 1.2.4 in theorem 1.2.1 (Laurens de Haan, Ana Ferreira (2006)) since we have

$$\lim_{t\to+\infty} \frac{1 - F_{NPED}(t + xf(t))}{1 - F_{NPED}(t)} = \lim_{t\to+\infty} \frac{f_{NPED}(t + xf(t))}{f_{NPED}(t)}$$

$$= \lim_{t\to+\infty} \frac{\sum_{k=0}^{n} a_{k,\theta}(t+xf(t))^{k+1}e^{-\theta(t+xf(t))}}{\sum_{k=0}^{n} a_{k,\theta}t^{k+1}e^{(-\theta t)}} = e^{-x} \quad (13)$$

(Such formula is called Γ -variation). Then, F_{NPED} lies in the Gumbelextreme domain of attraction. In his case, $f(t) = \frac{1}{\theta}$.

So, for (as in the invoked theorem) $a_{n,\theta} = f\left({F^{-1}}_{NPED}\left(1 - \frac{1}{n}\right)\right) = \frac{1}{\theta}$ and $b_{n,\theta} = {F^{-1}}_{NPED}\left(1 - \frac{1}{n}\right)$, we have:

$$\lim_{x\to+\infty} F_{NPED}\left(a_{n,\theta}x + b_{n,\theta}\right)^n = e^{(-e^{-x})}$$

2.7. Estimation of the Stress-Strength Parameter and Lorenz curve

Because it evaluates the system performance, the stress-strength parameter (R) is crucial in the reliability analysis. Furthermore, R indicates the likelihood of a system failure; the system breaks when the applied stress exceeds its strength, i.e.

$R = P(X > Y)$. Here, $X\sim NPED(\theta_1)$, denotes the strength of a system subject to stress Y, and $Y\sim NPED(\theta_2)$,, X and Y are independent of each other. In our case, the stress-strength parameter R is given by:

$$R = P(X > Y) = \int_0^\infty S_X(y)f_Y(y)dy$$

$$= \frac{\int_0^\infty \sum_{k=0}^{n} \frac{a_{k,\theta}\Gamma(k+2, y\theta_1)}{{\theta_1}^{k+2}} \sum_{k=0}^{n} a_{k,\theta}y^{k+1}e^{(-\theta_2 y)}dy}{\left(\sum_{k=0}^{n} a_{k,\theta}\frac{(k+1)!}{{\theta_1}^{k+2}}\right)\left(\sum_{k=0}^{n} a_{k,\theta}\frac{(k+1)!}{{\theta_2}^{k+2}}\right)}$$

The Lorenz curve is a well-known way of describing income and wealth distributions. The graph of the ratio is the Lorenz curve for a positive random variable X. Against $F(x)$with the properties $L(p) \leq p, L(0) = 0$and $L(1) = 1$. If X represents annual income, $L(p)$ is the proportion of total income that accrues to individuals with the $100\%p$ lowest incomes.

If all individuals earn the same income then $L(p) = p$ for all p. The area between the line $L(p) = p$ and the Lorenz curve can be used to calculate income inequality or, more broadly, the variability of X. The Lorenz curve is well known for the exponential distribution and is given by:

$$L(p) = p\{p + (1-p)\log(1-p)\}$$

For the $NPED$ distribution in (3),

$$E(X/X \leq x)F_{NPED}(x)\sum_{k=0}^{n} a_{k,\theta}\frac{(k+2)!}{\theta^{k+3}}(\frac{1-\sum_{k=0}^{n}\frac{a_{k,\theta}\Gamma(k+2,x\theta)}{\theta^{k+2}})}{(\sum_{k=0}^{n} a_{k,\theta}\frac{(k+1)!}{\theta^{k+2}})^{2}} \tag{14}$$

2.8. Entropies

It is commonly understood that entropy and information can be used to calculate the degree of uncertainty in a probability distribution. However, many correlations have been created based on the features of entropy.

The entropy of a random variable X is a measure of the uncertainty's variation. The entropy of Rényi is defined as:

$$J(\gamma) = \frac{1}{1-\gamma}\log\{\int_{0}^{\infty} f^{\gamma}(x)\,dx\}$$

where $\gamma > 0$ and $\gamma \neq 1$. For the $NPED$ distribution in (2), note that for γ integer we have:

$$\int f_{NPED}{}^{\gamma}(x)dx = \frac{\int(\sum_{k=0}^{n} a_{k,\theta}x^{k})^{\gamma}\, e^{(-\theta\gamma x)}dx}{(\sum_{k=0}^{n} a_{k,\theta}\,\frac{k!}{\theta^{k+1}})^{\gamma}}$$

$$= \frac{\sum_{k=0}^{n} b_{k,\theta}(\gamma)\int x^{k\gamma}\, e^{(-\theta\gamma x)}dx}{(\sum_{k=0}^{n} a_{k,\theta}\,\frac{k!}{\theta^{k+1}})^{\gamma}}$$

where: $\int x^{k\gamma}\, e^{(-\theta\gamma x)}dx = -\frac{1}{(\theta\gamma)^{k\gamma+1}}\Gamma(k\gamma+1, x\gamma\theta)$ and $b_{k,\theta}(\gamma)$ in function $a_{k,\theta}$ and γ. Now, the Rényi entropy is given by:

$$J(\gamma) = \frac{1}{1-\gamma}\log\left(\frac{\sum_{k=0}^{n} b_{k,\theta}(\gamma)\frac{(k\gamma)!\Gamma(k\gamma+1)}{(\theta\gamma)^{k\gamma+1}}}{\left(\sum_{k=0}^{n} a_{k,\theta}\frac{k!}{\theta^{k+1}}\right)}\right) \tag{15}$$

2.9. Estimation and inference

Let $X_1, \dots \dots X_n$ be a random sample of $NPED$. The ln-likelihood function $lnl(x_i;\theta)$ is given by:

$$lnl(x_i;\theta) = nlnh(\theta) + \sum_{i=1}^{n} \ln(\sum_{k=0}^{m} a_{k,\theta} {x_i}^k) - \theta \sum_{i=1}^{n} x_i \quad (16)$$

The derivative of $lnl(x_i;\theta)$ with respect to θ is:

$$\frac{lnl(x_i;\theta)}{d\theta} = \frac{n\dot{h}(\theta)}{h(\theta)} + \sum_{i=1}^{n} \frac{\dot{p}(x_i,\theta)}{p(x_i,\theta)} - \sum_{i=1}^{n} x_i$$

The Method of Moments (MoM) and Maximum Likelihood (ML) estimators of the parameter are the same after using $\boldsymbol{NPED}$ (16), and they may be found by solving the following non-linear equation:

$$\frac{\dot{h}(\theta)}{h(\theta)} + \frac{1}{n}\sum_{i=1}^{n} \frac{\dot{p}(x_i,\theta)}{p(x_i,\theta)} - \bar{x} = 0$$

where:

$$\dot{h}(\theta) = \frac{dh(\theta)}{d\theta} \; and \; \dot{p}(\theta) = \frac{dp(\theta)}{d\theta}$$

$$h(\theta)[\sum_{k=0}^{m} \frac{k!}{\theta^{k+2}} \left(a_{k,\theta}(k+1) - \dot{a_{k,\theta}}\theta\right] + \frac{1}{n}\sum_{i=1}^{n} \frac{\dot{p}(x_i,\theta)}{p(x_i,\theta)} - \bar{x} = 0 \quad (17)$$

Although this equation is difficult to answer, we can consider a specific scenario in which, $p(x_i,\theta) = (2+\theta+x_i) \; and \; h(\theta) = \frac{\theta^2}{(1+\theta)^2}$. This case will be studied in Section 3.

3. XLindley distribution and some properties

In this section, we present the XLindley (XL) distribution, which belongs to the new polynomial exponential family of distributions.

A random variable *X* is said to possess an XL distribution if it has the following form:

$$f_{XL}(x;\theta) = \frac{\theta^2(2+\theta+x)}{(1+\theta)^2} e^{-\theta x} \qquad x, \theta > 0 \quad (18)$$

Note that the XL distribution is a member of the new polynomial exponential family where $n = 1, a_{0,\theta} = 2+\theta, a_{1,\theta} = 1$using formula (1). Therefore, the mode of XL is given by

$$mode(X) = -\frac{\theta^2+2\theta-1}{\theta} \; for \;\; x, 0 < \theta < \sqrt{2} - 1 \quad (19)$$

We can find easily the CDF of the XL distribution

$$F_{XL}(x;\theta) = 1 - \left(1 + \frac{\theta x}{(1+\theta)^2}\right) e^{-\theta x} \qquad x,\theta > 0 \tag{20}$$

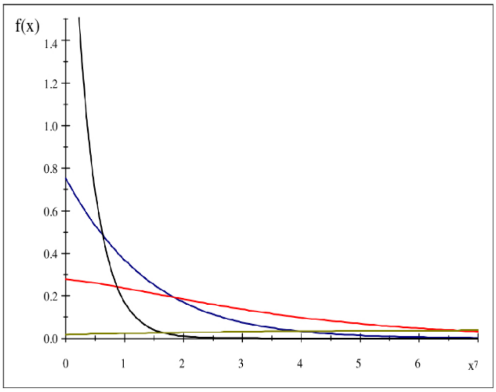

Figure.1. Plots of the density function for some parameter values of θ

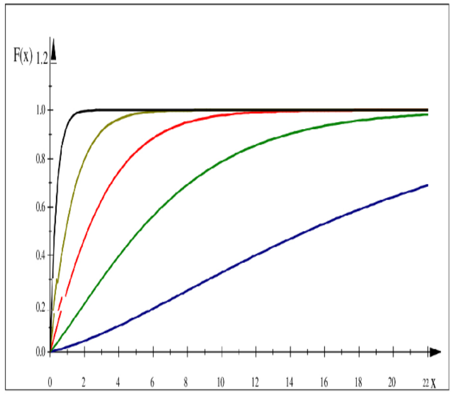

Figure.2. Plots of the cumulative function for some parameters values of θ

3.1. Survival and hazard rate function

For a continuous distribution, the survival function and the failure rate (hazard rate) functions are defined as:

$$S_{XL}(x;\theta) = 1 - F_{IXL}(x;\theta) = (1 + \frac{\theta x}{(1+\theta)^2})e^{-\theta x} \qquad x,\theta > 0 \tag{21}$$

$$h_{XL}(x;\theta) = \frac{f'_{XL}(x;\theta)}{1-F_{XL}(x;\theta)} = \frac{\theta^2(x+\theta+2)}{(1+\theta)^2+\theta x} \qquad x,\theta > 0 \tag{22}$$

Let equation (21) and (22) be the survival and hazard rate function, respectively.

Proposition 2. Let h_{XL} be the hazard rate function of *X*. Then, h_{XL} is increasing.

Proof. According to Glaser (1980) and from the density function (18):

$$\rho(x) = -\frac{f'_{XL}(x)}{f_{XL}(x;\theta)} = \frac{x\theta + \theta^2 - 2\theta - 1}{x + \theta + 2}$$

It follows that:

$$\rho'(x) = \frac{1}{(x + \theta + 2)^2}$$

Imply that h_{XL} is increasing.

3.2. Moments and related measures

The r^{th} moment about the origin of the XLindley distribution can be obtained as:

$$\mu_r' = E(X^r) = \int_0^\infty x^r f_{XL}(x)dx$$

$$= \int_0^\infty x^r \frac{\theta^2(2+\theta+x)}{(1+\theta)^2} e^{-\theta x} dx$$

$$\frac{\theta^2}{(1+\theta)^2}\int_0^\infty x^r (2+\theta+x)e^{-\theta x}dx$$

Finally, using gamma integral and little algebraic simplification, we get a general expression for the r^{th} factorial moment of XL distribution as:

$$\mu_r' = \frac{(\theta^2+2\theta+r+1)r!}{(1+\theta)^2\theta^r} \tag{23}$$

The first four moments can be derived by substituting $r = 1; 2; 3$ and 4 in (23), and then using the relationship between moments about origin and moments about mean, the first four moments about origin of the XL distribution may be obtained as follows:

$$\mu_1' = \frac{(\theta^2+2\theta+2)}{(1+\theta)^2\theta} = \frac{(1+\theta)^2+1}{(1+\theta)^2\theta} = \frac{1}{\theta} + \frac{1}{(1+\theta)^2\theta}$$

$$\mu_2' = \frac{2(\theta^2+2\theta+3)}{(1+\theta)^2\theta^2}$$

$$\mu_3' = \frac{6(\theta^2+2\theta+4)}{(1+\theta)^2\theta^3}$$

$$\mu_4' = \frac{24(\theta^2+2\theta+5)}{(1+\theta)^2\theta^4}$$

Let $X \sim XL(\theta)$, the mean, variance for X be:

$$\mu_1' = E(X) = \frac{(1+\theta)^2+1}{(1+\theta)^2\theta} \tag{24}$$

$$E(X^2) = \frac{2(\theta^2+2\theta+3)}{(1+\theta)^2\theta^2}$$

$$\mu_2 = Var(X) = \frac{(1+\theta)^4+4\theta^2+6\theta+1}{(1+\theta)^4\theta^2}$$

3.3. Estimation of parameter

3.3.1. Maximum Likelihood Estimation (MLE)

Let $X_i \sim XL(\theta)$, $i = 1, \ldots\ldots, n$ be n random variables, the ln-likelihood function, $lnl(x_i; \theta)$ is:

$$L(\theta) = (\tfrac{\theta^2}{(1+\theta)^2})^n \prod_{i=1}^{n}(2+\theta+x_i)\, e^{-\theta \sum_{i=1}^{n} x_i} \tag{25}$$

The logarithm of the likelihood function is:

$$lnl(x_i;\theta) = 2nlog\theta - 2nlog(\theta+1) + \sum_{i=1}^{n} log(2+\theta+x_i) - \theta \sum_{i=1}^{n} x_i$$

$$lnl(x_i;\theta) = 2n[log\theta - log(\theta+1)] + \sum_{i=1}^{n} log(2+\theta+x_i) - \theta \sum_{i=1}^{n} x_i \tag{26}$$

The derivatives of $lnl(x_i;\theta)$ with respect to θ are:

$$\frac{lnl(x_i;\theta)}{\delta\theta} = 0$$

$$\frac{\delta lnl(x_i;\theta)}{\delta\theta} = \frac{2n}{\theta} - \frac{2n}{1+\theta} + \sum_{i=1}^{n} \frac{1}{2+\theta+x_i} - \sum_{i=1}^{n} x_i$$

$$\frac{\delta lnl(x_i;\theta)}{\delta\theta} = \frac{2}{\theta} - \frac{2}{1+\theta} + \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2+\theta+x_i} - \bar{X}$$

$$\frac{\delta lnl(x_i;\theta)}{\delta\theta} = \frac{2}{\theta(1+\theta)} + \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2+\theta+x_i} - \bar{X} \tag{27}$$

To obtain the MLE of θ:$\hat{\theta}_{MLE}$ we can maximize equation (27) directly with respect toθ, or we can solve the non-linear equation$\frac{\delta lnl(x_i;\theta)}{\delta\theta} = 0$. Note that $\widehat{\theta}_{MLE}$ cannot be solved analytically; numerical iteration techniques, such as the Newton-Raphson algorithm, are thus adopted to solve the logarithm of the likelihood equation for which (27) is maximized.

3.3.2. Method of Moments Estimation (MME)

Let $\bar{X}$ be the sample mean, equating sample mean and population mean $E(X)$,

$$E(X) = \sum_{i=1}^{n} \frac{x_i}{n} \tag{28}$$

When we plug in the expression of $E(X)$ from equation (24) and solve the equation for θ, we get

$$\bar{X} = \frac{(1+\theta)^2 + 1}{(1+\theta)^2\theta} = \frac{\theta^2 + 2\theta + 2}{\theta^3 + 2\theta^2 + \theta}$$

We obtain equation of 3rd degree $\bar{X}\theta^3 + \theta^2(2\bar{X} - 1) + \theta(\bar{X} - 2) - 2 = 0$. We take the real part for the solution

$$\hat{\theta}_{MLE} = -\frac{1}{3\bar{X}}(2\bar{X} - 1) + \frac{\frac{2}{9\bar{X}} + \frac{1}{9\bar{X}^2} + \frac{1}{9}}{\sqrt[3]{\sqrt{\frac{1}{27\bar{X}} + \frac{13}{36\bar{X}^2} + \frac{1}{9\bar{X}^3} + \frac{1}{27\bar{X}^4} + \frac{11}{18\bar{X}} + \frac{1}{9\bar{X}^2} + \frac{1}{27\bar{X}^3} + \frac{1}{27}}}} +$$

$$\sqrt[3]{\sqrt{\frac{1}{27\bar{X}} + \frac{13}{36\bar{X}^2} + \frac{1}{9\bar{X}^3} + \frac{1}{27\bar{X}^4} + \frac{11}{18\bar{X}} + \frac{1}{9\bar{X}^2} + \frac{1}{27\bar{X}^3} + \frac{1}{27}}} \quad (29)$$

3.4. Simulation

The behaviour of the estimators for a finite sample size (n) is investigated in this subsection. A simulation study consisting of the following steps is being carried out N=10000 times for selected values of (θ, n), where $\theta = 0.05; 0.25; 1; 2; 5$ and $n = 20; 50; 100$.

– Generate U_i Uniform (0; 1), $i = 1, \dots .., n$.

– Generate Y_i Exponential$(\theta), i = 1, \dots .., n$.

– Generate Z_i Lindley$(\theta), i = 1, \dots .., n$.

– If $U_i \leq p(\theta)$, then set $X_i = Y_i$ otherwise, set $X_i = Z_i$, $i = 1, \dots .., n$

$$verage\ bias(\theta) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta).$$

And the average square error:

$$MSE(\theta) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2$$

Table 1. Average bias of the estimator $\hat{\theta}$

Bias	$\theta = 0.05$	$\theta = 0.25$	$\theta = 1$	$\theta = 2$	$\theta = 5$
$n = 20$	0.00131	0.01002	0.0456	0.2451	0.7512
$n = 50$	0.00095	0.0124	0.0106	0.1162	0.1421
$n = 100$	0.00011	0.00251	0.0122	0.0423	0.0506

Table 2. The average square error of the estimator $\hat{\theta}$

MSE	$\theta = 0.05$	$\theta = 0.25$	$\theta = 1$	$\theta = 2$	$\theta = 5$
$n = 20$	$1{,}03.10^{-6}$	0.000113	0.00236	0.0654	0.6177
$n = 50$	$2, 55.10^{-7}$	0.000214	0.000162	0.01233	0.03135
$n = 100$	$1{,}04.10^{-8}$	$1.34.10^{-5}$	0.000216	0.00184	0.00301

Table 1 and 2 show the outcomes of the simulation. The simulation analysis yielded the following conclusions:

- for some given value of θ, the average of bias of θ and the mean square error of θ decrease as the sample size n increases,
- the mean square error (MSE) gets higher and following a similar way for larger value of θ as we mentioned before.

3.5. Application and goodness of fit

Data set 1: Survival times (in months) of 94 Sierra Leone individuals infected with Ebola virus. It is available at *https://apps.who.int/gho/data/node.ebola-sitrep*. In table 3, we compare the Lindley (LD), Zeghdoudi, exponential, XGamma, and XL distributions using data set 1.

Table 3. Comparison between LD, XG, ZD, Exp and XL distributions.

Survival time m=3.17 , s=2.095	**Obsfreq**	**LD** $\hat{\theta}=0.522$	**Xgamma** $\hat{\theta}=0.689$	**ZD** $\hat{\theta}=0.852$	**Exp** $\hat{\theta}=0.315$	**XL** $\hat{\theta}=0.467$
[0,2]	45	38. 262	37. 652	30. 339	43. 937	41. 028
[2,4]	22	28. 164	27. 197	37.27	23. 4	25. 855
[4,6]	17	15. 075	16. 342	17.743	12. 463	13. 984
[6,8]	7	7. 1187	7. 7769	6.1658	6. 6375	6. 9986
[8,10]	3	3. 1423	3. 2015	1.828	3. 5351	3. 3409
Total	94	94	94	94	94	94
χ^2	-	2. 7899	3. 2040	14.236	1. 8619	**1. 6446**

4. Exponential-gamma $(3, \boldsymbol{\theta})$ (X gamma) distribution and its applications

In this section, we give an overview on Exponential-gamma Eg ($\boldsymbol{\theta}$) (X gamma) distribution (see Subhradev (2016)), which is a member of the NPED. A random variable X is said to possess Eg($\boldsymbol{\theta}$) distribution if it has the following form:

$$f_{EG}(x;\theta)=\frac{\theta^2}{(1+\theta)}(1+\frac{\theta}{2}x^2)\,e^{-\theta x} \qquad x,\theta>0 \tag{30}$$

Note that the *Eg* distribution is a member of the NPED family where:
$n=2, a_{0,\theta}=1, a_{1,\theta}=0, a_{2,\theta}=\frac{\theta}{2}$, using formula (1).

Therefore, the mode of *Eg* ($\boldsymbol{\theta}$) distribution is given by:

$$mode(X)=\frac{1+\sqrt{1-2\theta}}{\theta}\ for\ \ 0<\theta<\frac{1}{2} \tag{31}$$

We can find easily the CDF of the *Eg* ($\boldsymbol{\theta}$) distribution:

$$F_{Eg}(x;\theta)=1-\frac{(1+\theta+\theta x+\frac{\theta^2x^2}{2})}{(1+\theta)}e^{-\theta x} \qquad x,\theta>0 \tag{32}$$

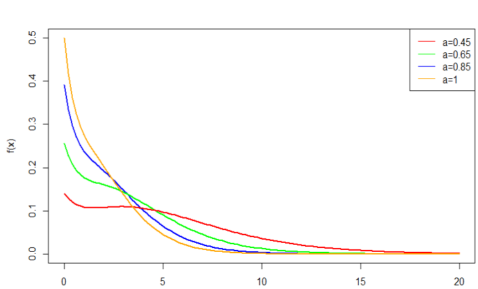

Figure 3. Plots of the density function for some parameters values of θ

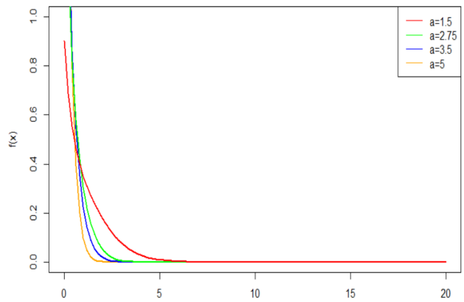

Figure4. Plots of the cumulative function for some parameters values of θ

4.1. Survival and hazard rate function

For a continuous distribution, the survival function and failure rate (hazard rate) functions are defined as:

$$S_{Eg}(x;\theta) = 1 - F_{Eg}(x;\theta) = \frac{(1+\theta+\theta x+\frac{\theta^2x^2}{2})}{(1+\theta)}e^{-\theta x} \qquad x,\theta > 0 \qquad (33)$$

4.2. Moments and related measures

The r^{th} moment about the origin of the *Eg* ($\boldsymbol{\theta}$) distribution can be obtained as:

$$\mu_r' = E(X^r) = \frac{r!(\theta+r+a_r)}{\theta^r(1+\theta)} \qquad (34)$$

where $a_r = a_{r-1} + r$ for $r = 1,2,3, \ldots ..$with $a_0 = 0$ and $a_1 = 2$. In particular,

$$\mu_1' = \frac{(\theta+3)}{\theta(\theta+1)} = Mean(X) = \mu$$

$$\mu_2' = \frac{2(\theta+6)}{\theta^2(\theta+1)}, \mu_3' = \frac{6(\theta+10)}{\theta^3(\theta+1)}, \mu_4' = \frac{24(\theta+15)}{\theta^4(\theta+1)}$$

It is to be noted that, for the exponential distribution with parameter θ, the r^{th} order moment about origin is

$$\mu_r' = \frac{r!}{\theta^r}$$

The j^{th} order central moment of the *Eg* ($\boldsymbol{\theta}$) is

$$\mu_j = E\left[(X-\mu)^j\right] = \sum_{r=0}^{j}\binom{j}{r}\mu_r'(-\mu)^{j-r}.$$ In particular,

$$\mu_2 = \frac{(\theta^2+8\theta+3)}{\theta^2(1+\theta)^2} = \text{var(X)} = \sigma^2$$

$$\mu_3 = \frac{2(\theta^3 + 15\theta^2 + 9\theta + 3)}{\theta^3(1+\theta)^3}$$

$$\mu_4 = \frac{3(5\theta^4 + 88\theta^3 + 310\theta^2 + 288\theta + 177)}{\theta^4(1+\theta)^4}$$

4.3. Estimation of parameter

Let $X_i \sim Eg(\theta)$distribution, $i = 1, \ldots\ldots, n$ be n random variables. The ln-likelihood function, $lnl(x_i; \theta)$ is:

$$L(\theta) = \prod_{i=1}^{n} f(x_i{:}\,\theta) = \prod_{i=1}^{n} \frac{\theta^2(1+\frac{\theta}{2}x_i^2)e^{-\theta x_i}}{1+\theta}$$

The logarithm of the likelihood function is:

$$\log L\,(x_i; \theta) = 2nlog\theta - nlog(1+\theta) + \sum_{i=1}^{n}[\log\left(1+\frac{\theta}{2}x_i^2\right) - \theta x_i] \quad (35)$$

The derivatives of $\log L\,(x_i; \theta)$with respect to θ are:

$$\frac{\delta L}{\delta\theta} = \frac{2n}{\theta} - \frac{n}{1+\theta} + \sum_{i=1}^{n}\left(\frac{x_i^2}{2(1+\frac{\theta}{2}x_i^2)} - x_i\right)$$

We get the likelihood equation as a system of nonlinear equations in θ by setting the left side of the above equation to zero. The MLE of θ in this system is obtained by solving it in θ. It is simple to calculate numerically using a statistical software tool such as the nlmpackage in R programming with arbitrary initial values.

The Fisher information about θ,$I(\theta)$, is

$$I(\theta) = E\left\{-\frac{\partial^2}{\partial^2\theta^2} lnf(X,\theta)\right\} = E\left\{\frac{2}{\theta^2} - \frac{1}{(1+\theta)^2} + \frac{x^4}{4}\frac{1}{\left(1+\frac{\theta}{2}x^2\right)^2}\right\}$$

$$= \frac{2}{\theta^2} - \frac{1}{(1+\theta)^2} + E\left\{\frac{x^4}{4}\frac{1}{(1+\frac{\theta}{2}x^2)^2}\right\} \quad (36)$$

Then the asymptotic $100(1-\alpha)\%$ confidence interval for θ is given by $\hat{\theta} \pm z\alpha/_2 \frac{I^{-1/2}}{\sqrt{n}}$.

4.4. Simulation

Table 4. Average bias and MSE of the estimator $\hat{\theta}$

θ	n	**Bias**	**MSE**
1	50	-0.00086	$3.65\,e^{-05}$
	100	0.00040	$1.56\,e^{-05}$
	500	$1.32\,e^{-05}$	$8.56\,e^{-08}$
1.5	50	-0.000061	$2.64\,e^{-05}$
	100	-0.00063	$3.34\,e^{-05}$
	500	$-3.92\,e^{-06}$	$7.63\,e^{-09}$
1.85	50	0.00174	0.000153
	100	0.00090	$8.61\,e^{-05}$
	500	0.000168	$1.4097\,e^{-05}$

4.5. Data analysis and applications

Application of the Eg distribution is illustrated in two examples.

Data set 2: The data set is taken from Klein and Berger. It shows the survival data on the death times of 26 psychiatric inpatients admitted to the University of Iowa hospitals during the years 1935-1948.

Table. 5. The survival data on the death times of psychiatric inpatients.

1	1	2	22	30	28	32	11	14	36	31	33	33
37	35	25	31	22	26	24	35	34	30	35	40	39

To evaluate the data, we used three different distributions: ED, EED, and Eg distributions. Table 6 shows the estimated unknown parameters, as well as the accompanying Kolmogorov-Sminrov (*K-S*) test statistic and *LogL* values for three alternative models.

Table 6. The estimates, K-S test statistic and $log-likelihood$ for the data set 2

Model	**Estimates**	***K-S***	***LogL***
ED	$\hat{\theta}=0.0378$	0.377	-112.321
EED	$\hat{a}=1.797, \hat{b}=0.052$	0.318	-109.998
Eg	$\hat{\theta}=0.0105$	**0.3146**	**-104.611**

We present the *p-value*, corresponding Akaikes Information Criterion (AIC) (see Akaike, H. (1974) and Bayesian Information Criterion (*BIC*) in the following table 7.

Table 7. The *p-value, AIC* and *BIC* of the models on the base data set 1

Model	*p-value*	*AIC*	*BIC*
ED	0.001	224.264	225.518
EED	0.011	221.974	224.490
Eg	**0.057**	**211.171**	**212.429**

Table 6 provides the fitted distributions' parameter MLEs and log likelihood values, while table 7 shows the AIC, BIC, and p-value values. Tables 6 and 7 show that the Eg (θ) distribution is a strong rival to the other distributions chosen to suit the dataset here.

Data set 3: Chen (Gupta R. D. and Kundu D. (1999)) gave type-II censoring data of samples with complete unit failures: 0.29, 1.44, 8.38, 8.66, 10.20, 11.04, 13.44, 14.37, 17.05, 17.13, and 18.35. Table 8 shows the estimated unknown parameters, as well as the accompanying Kolmogorov-Smirnov (*K-S*) test statistic and *Log L* values for three alternative models.

Table 8. The estimates, *K-S* test statistic and log-likelihood for the data set 2

Model	Estimates	*K-S*	*LogL*
ED	$\hat{\theta} = 0.091$	0.3622	-40.432
EED	$\hat{a} = 1.355, \hat{b} = 0.109$	0.3183	-38.523
Eg	$\hat{\theta} = 0.237$	**0.251**	**-35.642**

We present the *p-value*, corresponding *AIC* and *BIC* for the data set in 2 in Table 9.

Table 9. The $p-values$, (AIC) and (BIC) of the models based on the data set 3

Model	*p-value*	*AIC*	*BIC*
ED	0.098	76.635	77.033
EED	0.172	78.093	78.889
Eg	**0.462**	**72.504**	**72.902**

The parameter MLEs and log-likelihood values of the fitted distributions are shown in table 8, and the values of *AIC*, *BIC*, and *p-values* are shown in Table 9. Tables 8 and 9 show that the *Eg (θ)* is a strong rival to the other distributions employed to suit the dataset here.

5. Conclusions

We have suggested a family of distributions with only one parameter in this paper. Moments, distribution function, characteristic function, failure rate, stochastic order, maximum likelihood approach, and method of moments were among the properties studied.

The Lindley and Zeghdoudi distributions lack the flexibility needed to examine and model many forms of data related to lifetime data and survival analysis. The $NPED$ distribution, on the other hand, is adaptable, straightforward, and simple to use. The novel distribution was used to evaluate two real data sets and was compared to existing distributions (Lindley, exponential, Zeghdoudi, Exponential Exponential and Xgamma). The comparison's findings support the $NPED$ distribution's quality adjustment. We anticipate that our new distribution family will entice many additional life data, reliability analysis, and actuarial science applications.

We can employ a more general distribution with two parameters in future experiments, and

$$f_{New}(x,\theta) = h(\theta)p(x,\theta)cos\theta\exp(-\theta x)$$

where $h(\theta)$ is real-valued functions on $[0,\infty]$, and where $p(x,\theta) = b(\theta) + x^k$.

Acknowledgement

The authors acknowledge the Editor-in-chief, *Prof. Dr. Patryk Barszcz* and the referees of this journal for constant encouragement to finalize the paper.

References

Akaike, H., (1974). A new look at the statistical model identification, IEEE Transactions on Automatic Control, AC-19, pp. 716–723.

Asgharzadeh, A., Hassan, S. and Bakouch, L. E., (2013). Pareto Poisson-Lindley distribution and its application. *Journal of Applied Statistics*, Vol. 40, No. 8, pp. 1−18.

Bashir, S., Rasul, M., (2015). Some properties of the weighted Lindley distribution. *International Journal of Economic and Business Review*, 3 (8), pp. 11−17.

Beghriche, A. F., Zeghdoudi, H., (2019). A Size Biased Gamma Lindley Distribution. *Thailand Statistician*, Vol.17, No 2, pp. 179–189.

Fisher, R. A., (1934). The effects of methods of ascertainment upon the estimation of frequencies. *Ann. Eugenics*, , 6, pp. 13–25.

Ghitany, M. E, Atieh, B and Nadarajah, S., (2008b). Lindley distribution and its applications. Mathematics and Computers in Simulation, 78, pp. 493–506.

Glaser, R. E., (1980). Bathtub and related failure rate characterizations. *J. Amer.Statist. Assoc.*, 75, pp. 667–672.

Gupta R. D., Kundu D., (1999). Generalized Exponential distribution. *Australian and New Zealand Journal of Statistics*, Vol. 41, No. 2, pp.173–188.

Laurens de Haan, Ferreira A., (2006). Extreme value theory: An introduction. *Springer*.

Lawless, J. F., (2003). Statistical models and methods for lifetime data. Wiley, New Y [16] Sen, Subhradev; Maiti, Sudhansu S.; and Chandra, N., (2016). The Xgamma Distribution: Statistical Properties and Application. *Journal of Modern Applied Statistical Methods*, Vol. 15, Iss. 1, Article 38.

Lindley, D. V., (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Society*, series B, 20, pp. 102–107.

Patil, G. P., Rao, C. R., (1977). Weighted distribution survey of their applications, In P. R. Krishnaiah, (E ds.), Applications of statistics, pp. 383–405, Amsterdam, North Holland.

Patil, G. P., Rao, C. R., (1978). Weighted distributions and size biased sampling with applications to wild life populations and human families. *Biometrics*, 34, pp. 179–189.

Sen Subhradev, Maiti Sudhansu S. and Chandra, N., (2016). The xgamma Distribution: Statistical Properties and Application. *Journal of Modern Applied Statistical Methods*, Vol. 15, Iss. 1, Article 38. DOI: 10.22237/jmasm/1462077420.

Zeghdoudi, H., Bouchahed, L., (2018). A new and unified approach in generalizing the Lindley's distribution with applications. *Statistics in Transition new series,* Vol. 19, No. 1, pp. 61–74,

Zeghdoudi, H., Messadia, M., (2018). Zeghdoudi Distribution and its Applications. *International Journal of Computing Science and Mathematics***,** Vol. 9, No.1, pp. 58–65.

Zeghdoudi, H.,. Nedjar, S., (2016c). A pseudo Lindley distribution and its application. *Afr. Stat.*, Vol. 11, No. 1, pp. 923–932.