Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v16n3p614

A New Two-parameter Distribution-G Family: Properties and Applications

By AIDI Khaoula,Abd EL-Baset A. Ahmad,Seddik-Ameur

15 December 2023

Electronic Journal of Applied Statistical Analysis
Vol. 16, Issue 3, 2023, 614-638
DOI: 10.1285/i20705948v16n3p614

A New Two-parameter Distribution-G Family: Properties and Applications

AIDI Khaoula*[a], Abd EL-Baset A. Ahmad[b], and Seddik-Ameur N.[a]

[a]*Laboratory of probability and statistics LaPS, University Badji Mokhtar- Annaba- Algeria*
[b]*Department of Mathematics, Assuit University, Assuit 71515, Egypt.*

15 December 2023

In this paper, a new family of lifetime distributions called new two-parameter distribution-G family is introduced. This family includes some new distributions such as new two-parameter distribution -generalized linear exponential family. Some statistical properties of the proposed distribution are obtained, such as the hazard rate function, moments, moment generating function and order statistics. We discuss the estimation of the distribution parameters by method of maximum likelihood in both of complete and right censored cases. A modified criteria test is developed to fit this new model when the parameters are unknown and data are right censored. The performances of the methods used are demonstrated by an intensive simulation study whether the usefulness and the versatility in practical applications of this distribution is illustrated by means of real data sets.

keywords: Censorship, Goodness-of-fit test, Moments, Moment generating function, Maximum likelihood estimation, Markov Chain Monte Carlo.

Although there are many families of distributions in the literature, many practical fields, including lifetime analysis, medicine, engineering, economics, finance, and insurance, clearly call for flexible families of distributions. Recently, numerous extended distributions have been extensively used for modeling data in several areas. Recent extensions focus on defining new families extended some famous distributions and providing great flexibility in modeling data in practice. Hence, several families of new distributions have been proposed in the statistical literature by adding one or more parameters and by compounding different distributions using different techniques. However, a lot of

*khaoula.aidi@yahoo.fr

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

authors have shown a strong interest in providing new distributional families in contemporary literature. For instance, Famoye et al. (2005), Akinsete et al.(2008), Silva et al. (2010), Nadarajah et al. (2011), Alzaatreh et al. (2013), Bagdonavičius et al. (2013), Alizadeh et al. (2015), Elgarhy et al. (2016), Elgarhy et al. (2017), Khan et al. (2018), Elbatala et al. (2019),ldahlan (2020), Almongy et al. (2021), Alshenawy et al. (2022), Alotaibi et al. (2022), Emam and Tashkandy (2023), among others. The obtained models are generally based on classical distributions such as exponential, Weibull,gamma, Burr distributions, among others. These classical models with closed forms cumulative distribution functions (cdf) are proposed in thestatistical literature and can be used to generate new models able to describe more observed data. Doostmoradi (2018) proposed a new two-parameter model and called it new two-parameter distribution (ND), this model can has an increasing, decreasing and unimodal failure rates. By using the approach of Bourguignon et al. (2014), we construct a new family based on the (ND) proposed by Doostmoradi (2018).

In this paper, we introduce a new family of continuous distributions called new two-parameter distribution- G (ND-G) family including some new members of distributions and some of their properties and applications are discussed. Maximum likelihood estimation method is used for determining the unknown parameters in both of complete and right censored cases. A powerful modified criteria goodness-of-fit test is developed for this new model when data are right censored. This test statistic allows us to distinguish between this model from its alternatives. Classical criteria choice models such as AIC, BIC, $CAIC$, $HQIC$, the Kolmogorov-Smirnov statistic (K-S) and the p-value were used to prove the versatility of this distribution family in practice compared to several distributions.
The rest of the article is organized as follows: In Section 2, the ND-G family is composed and described, probability weighted moments and order statistics are obtained in Section 3. Section 4 is devoted to maximum likelihood estimation in complete and right censored data cases. Some members of this family are given in Section 5 where the ND- generalized linear exponential ($ND-GLE$) distribution is developed. We have considered particularly the case where the baseline distribution G is the generalized linear exponential distribution (GLE) because of its flexibility and its several advantages such as this one is a generalization of six models at least. Finally, the usefulness and the versatility in practice applications of the proposed model is illustrated by an extensive simulation study and two applications with real data sets.

1 New Distribution-G Family

A new family of continuous distributions will be proposed making use of the transformed-transformer $T-X$ method suggested by (Alzaatreh et al., 2013) , where X be a random variable (rv) with cdf $G(x)$ and probability density function (pdf) $g(x)$, and T be a rv with cdf $Q(t)$ and pdf $q(t)$. If the function $W(G(x))$, defined in (Alzaatreh et al., 2013), is chosen such that $W(G(x)) = G(x)/\bar{G}(x)$, where $\bar{G}(x) = 1 - G(x)$, then a new family of distributions with cdf $F(x)$, can be obtained from

$$F(x) = \int_0^{\frac{G(x)}{\bar{G}(x)}} q(t)dt.$$

(Doostmoradi, 2018) introduced a new continuous distribution called the new two-parameter distribution (ND), with cdf given by

$$Q(t) = 1 - (1 + \alpha t^\lambda)e^{-\alpha t^\lambda}, \qquad t > 0 \quad (\alpha > 0, \lambda > 0),$$

And pdf $q(t)$, given by

$$q(t) = \lambda \alpha^2 t^{2\lambda-1} e^{-\alpha t^\lambda}.$$

By using the pdf $q(t)$ of the ND and the baseline $G(x, \xi)$ depends on a parameter vector xi, we introduce a new family of continuous distributions called the new two-parameter distribution-G ($ND - G$) family in the form

$$F(x; \alpha, \lambda, \xi) = 1 - \left(1 + \alpha \left[\frac{G(x, \xi)}{\bar{G}(x, \xi)}\right]^\lambda\right) \exp\left\{-\alpha \left[\frac{G(x, \xi)}{\bar{G}(x, \xi)}\right]^\lambda\right\}. \tag{1}$$

The pdf of the $ND - G$ family can be given as

$$f(x; \alpha, \lambda, \xi) = \alpha^2 \lambda g(x, \xi) \frac{[G(x, \xi)]^{2\lambda-1}}{\left[\bar{G}(x, \xi)\right]^{2\lambda+1}} \exp\left\{-\alpha \left[\frac{G(x, \xi)}{\bar{G}(x, \xi)}\right]^\lambda\right\}. \tag{2}$$

Also, the hazard rate function, $h(x) = f(x)/\bar{F}(x)$, of the $ND - G$ family is given by

$$h(x) = \alpha^2 \lambda g(x, \xi) \frac{[G(x, \xi)]^{2\lambda-1}}{\left[\bar{G}(x, \xi)\right]^{2\lambda+1}} \left(1 + \alpha \left[\frac{G(x, \xi)}{\bar{G}(x, \xi)}\right]^\lambda\right)^{-1}.$$

The pdf (2) can be written, making use of the power series for the exponential factor, as

$$f(x; \alpha, \lambda, \xi) = \sum_{j=0}^{\infty} \frac{(-1)^j \lambda \alpha^{j+2}}{j!} g(x, \xi) \frac{[G(x, \xi)]^{(j+2)\lambda-1}}{\left[\bar{G}(x, \xi)\right]^{(j+2)\lambda+1}}, \tag{3}$$

And making use of the generalized binomial expansion for $\left[\bar{G}(x, \xi)\right]^{-[(j+2)\lambda+1]}$, we have

$$\left[\bar{G}(x, \xi)\right]^{-[(j+2)\lambda+1]} = \sum_{k=0}^{\infty} \binom{(j+2)\lambda + k}{k} [G(x, \xi)]^k,$$

Hence (3) reduces to

$$f(x; \alpha, \lambda, \xi) = \sum_{k=0}^{\infty} \epsilon_k g(x, \xi) [G(x, \xi)]^{(j+2)\lambda+k-1}, \tag{4}$$

Where

$$\epsilon_k = \sum_{j=0}^{\infty} \frac{(-1)^j \lambda \alpha^{j+2}}{j!} \binom{(j+2)\lambda + k}{k}. \tag{5}$$

The pdf (4) can be written, as infinite linear combination, as

$$f(x;\alpha,\lambda,\xi)=\sum_{k=0}^{\infty}\omega_k\psi_{(j+2)\lambda+k}(x,\xi),$$

Where

$$\begin{aligned}\omega_k &= \frac{\epsilon_k}{(j+2)\lambda+k}, \\ \psi_{c_{j,k}}(x;\xi) &= c_{j,k}g(x;\xi)\left[G(x,\xi)\right]^{c_{j,k}-1}, \\ c_{j,k} &= (j+2)\lambda+k,\end{aligned} \tag{6}$$

And $\psi_{c_{j,k}}(x;\xi)$ represents the exponentiated-G pdf with the power parameter $c_{j,k}$. Also, we can show that

$$\bar{F}(x;\alpha,\lambda,\xi)=\sum_{j,k=0}^{\infty}\left[\zeta_{1_{j,k}}G^{j\lambda+k}(x;\xi)+\zeta_{2_{j,k}}G^{(j+1)\lambda+k}(x;\xi)\right], \tag{7}$$

Where

$$\begin{aligned}\zeta_{1_{j,k}} &= \frac{(-1)^j\alpha^j}{j!}\binom{j\lambda+k-1}{k}, \\ \zeta_{2_{j,k}} &= \frac{(-1)^j\alpha^{j+1}}{j!}\binom{(j+1)\lambda+k-1}{k}.\end{aligned} \tag{8}$$

In the sequel we need the function $[F(x)]^s$; s is integer, so using the same manner as above and using (1), it can be written by the form

$$[F(x;\alpha,\lambda,\xi)]^s=\sum_{u=0}^{\infty}S_u\left[G(x;\xi)\right]^{(d+l)\lambda+u}, \tag{9}$$

Where

$$S_u=\sum_{i=0}^{s}\sum_{t=0}^{i}\sum_{d=0}^{t}\sum_{l=0}^{\infty}\frac{(-1)^{i+t+l}t^l\alpha^{d+l}}{l!}\binom{s}{i}\binom{i}{t}\binom{t}{d}\binom{(d+l)\lambda+u-1}{u}. \tag{10}$$

2 Statistical Properties

This section is devoted to derive some statistical properties for the $ND-G$ family such as: probability weighted moments,non-central moments, moment generating function and functions of order statistics and their moments.

2.1 Probability weighted moments

Greenwood et al. (1979) defined probability weighted moments (PWMs) for a random variable X, denoted by $M_{r,s,m}$, by the following quantities

$$M_{r,s,m}=E\left[X^rF^s(X)\bar{F}^m(X)\right]=\int_0^{\infty}x^rf(x)F^s(x)\bar{F}^m(x)dx. \tag{11}$$

The quantities $M_{r,0,0}$, $M_{1,s,0}$ and $M_{1,0,m}$, represent rth non-central moments and L-moments, respectively. Greenwood et al.(1979) state that L-moments are more convenient because they are more directly interpretable as measures of the scale and shape of probability distributions. The non-central moments of the ND-G family can be obtained from (4) by

$$\begin{aligned} \mu'_r &= E(X^r) = M_{r,0,0} \\ &= \sum_{k=0}^{\infty} \epsilon_k \int_0^{\infty} x^r g(x,\xi) \left[G(x,\xi)\right]^{(j+2)\lambda+k-1} dx, \end{aligned} \tag{12}$$

Which can be written as

$$\mu'_r = \sum_{k=0}^{\infty} \epsilon_k \Psi_{c_{j,k}}(\xi), \tag{13}$$

where

$$\Psi_{c_{j,k}}(\xi) = \int_0^{\infty} x^r g(x,\xi) \left[G(x,\xi)\right]^{c_{j,k}-1} dx, \tag{14}$$

Where $c_{j,k} = (j+2)\lambda + k$, from (6).
Also, $M_{1,s,0} = E\left[XF^s(X)\right]$, L-moment, can be given, substituting (4), (9) in (11), by

$$E\left[XF^s(X)\right] = \sum_{u,k=0}^{\infty} S_u \epsilon_k \int_0^{\infty} x g(x,\xi) \left[G(x,\xi)\right]^{(d+j+l+2)\lambda+u+k-1} dx, \tag{15}$$

Which can be written as

$$E\left[XF^s(X)\right] = \sum_{u,k=0}^{\infty} S_u \epsilon_k \Omega_{b_{u,k}}(\xi), \tag{16}$$

Where

$$\Omega_{b_{u,k}}(\xi) = \int_0^{\infty} x g(x,\xi) \left[G(x,\xi)\right]^{b_{u,k}-1} dx,$$

And $b_{u,k} = (d+j+l+2)\lambda + u + k$, where ϵ_k and S_u are given in (5) and (10). The moment generating function, $M_X(t)$, using the Maclaurin series expansion of the function $\exp(-tx)$ in terms of μ'_r, can be given by the form

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu'_r,$$

Then, using (12), we obtain

$$M_X(t) = \sum_{r,k=0}^{\infty} \frac{\epsilon_k t^r}{r!} \Psi_{c_{j,k}}(\xi), \tag{17}$$

Where $\Psi_{c_{j,k}}(\xi)$ is given by (13).

2.2 Order statistics

Order statistics and their moments play a fundamental rule in various areas such as data analysis relating to quality control, reliability, hydrological, problems of estimation and hypothesis tests and extreme values. Therefore, we investigate some properties of order statistics for the proposed family of distributions. If $X_1, X_2, \ldots, X_n$ is a random sample from a population with cdf $F(x)$ and pdf $f(x)$, then the pdf of the i^{th} order statistic $X_{i:n}$, is given by

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!}\left[F(x)\right]^{i-1}\left[\bar{F}(x)\right]^{n-i} f(x),$$

Which can be written as

$$f_{i:n}(x) = \frac{n!}{(n-i)!} f(x) \sum_{\ell=0}^{i-1} \frac{(-1)^{\ell}}{\ell!(i-\ell-1)!}\left[\bar{F}(x)\right]^{n+\ell-i}. \tag{18}$$

Substituting from (4) and (7) in (18), we can obtain

$$\begin{aligned} f_{i:n}(x) &= \frac{n!}{(n-i)!} \sum_{\ell=0}^{i-1} \frac{(-1)^{\ell}}{\ell!(i-\ell-1)!} g(x,\xi)[G(x,\xi)]^{(2\lambda-1)(n+\ell-i)} \\ &\quad \times \left[\sum_{j,k=0}^{\infty} \epsilon_k [G(x;\xi)]^{2(j\lambda+k)} \left(\zeta_{1_{j,k}} + \zeta_{2_{j,k}} G^{\lambda}(x;\xi)\right)\right]^{n+\ell-i} \end{aligned} \tag{19}$$

The r^{th} moment of the i^{th} order statistic can be given, from (19) by

$$\begin{aligned} \mu_{i:n}^{(r)} &= \frac{n!}{(n-i)!} \sum_{\ell=0}^{i-1} \frac{(-1)^{\ell}}{\ell!(i-\ell-1)!} \int_0^{\infty} x^r g(x,\xi)[G(x,\xi)]^{(2\lambda-1)(n+\ell-i)} \\ &\quad \times \left[\sum_{j,k=0}^{\infty} \epsilon_k [G(x;\xi)]^{2(j\lambda+r)} \left(\zeta_{1_{j,k}} + \zeta_{2_{j,k}} G^{\lambda}(x;\xi)\right)\right]^{n+\ell-i} dx. \end{aligned} \tag{20}$$

3 Maximum Likelihood Estimation

3.1 Maximum likelihood estimators for complete data

To determine the maximum likelihood estimates (MLEs) of the parameters of the $ND-G$ family of distributions from complete sample, let us consider $x_1, x_2, ..., x_n$, n observed values drawn from the $ND-G$ family with parameters α, λ and ξ. Let $\Theta = (\alpha, \lambda, \xi)^T$ be the $p \times 1$ parameter vector. The total log-likelihood function of Θ is given, from (2), by

$$\begin{aligned}\ell(\Theta) &= \sum_{i=1}^{n} \ln(f(x_i;\Theta)) \\ &= 2n\ln\alpha + n\ln\lambda + \sum_{i=1}^{n}\ln g(x_i,\xi) + (2\lambda-1)\sum_{i=1}^{n}\ln G(x_i,\xi) \\ &\quad -(2\lambda+1)\sum_{i=1}^{n}\ln\bar{G}(x_i,\xi) - \alpha\sum_{i=1}^{n}\left[M(x_i,\xi)\right]^{\lambda},\end{aligned}$$

Where $M(x,\xi) = \frac{G(x,\xi)}{\bar{G}(x,\xi)}$. The components of the score function are

$$\frac{\partial\ell}{\partial\alpha} = \frac{2n}{\alpha} - \sum_{i=1}^{n}\left[M(x_i,\xi)\right]^{\lambda},$$

$$\frac{\partial\ell}{\partial\lambda} = \frac{n}{\lambda} + 2\sum_{i=1}^{n}\ln G(x_i,\xi) - 2\sum_{i=1}^{n}\ln\bar{G}(x_i,\xi) - \alpha\sum_{i=1}^{n}\left[M(x_i,\xi)\right]^{\lambda}\ln M(x_i,\xi)$$

$$\begin{aligned}\frac{\partial\ell}{\partial\xi_k} &= \sum_{i=1}^{n}\frac{g'_k(x_i,\xi)}{g(x_i,\xi)} + (2\lambda-1)\sum_{i=1}^{n}\frac{G'_k(x_i,\xi)}{G(x_i,\xi)} \\ &+ (2\lambda+1)\sum_{i=1}^{n}\frac{G'_k(x_i,\xi)}{\bar{G}(x_i,\xi)} - \alpha\lambda\sum_{i=1}^{n}\left[M(x_i,\xi)\right]^{\lambda-1}M'_k(x_i,\xi),\end{aligned}$$

Where $Z'_k(.,\xi) = \partial Z(.,\xi)/\partial\xi_k$ and $k = 1, 2, \ldots, d$, where d represents the total number of parameters of $G(x;\xi)$.
Setting $\frac{\partial\ell}{\partial\alpha}$, $\frac{\partial\ell}{\partial\lambda}$ and $\frac{\partial\ell}{\partial\xi_k}$ equal to zero and then solving the resulting equations simultaneously yields the MLEs $\hat{\Theta} = \left(\hat{\alpha},\hat{\lambda},\hat{\xi}\right)^T$ of $\Theta = (\alpha,\lambda,\xi)^T$. These equations cannot be solved analytically and statistical software can be used to solve them numerically using iterative methods such as the Newton-Raphson type algorithms.

3.2 Maximum likelihood estimation with right censorship

In reliability studies and survival analysis, data are often censored. If $X_1, X_2, ..., X_n$ is a censored sample from the $ND-G$ distribution, each observation can be written as $x_i = \min(X_i, C_i)$ for $i = 1, ..., n$ where X_i are failure times and C_i censoring times. Censoring is considered to be non-informative, so the likelihood function is reduced to

$$L(x,\Theta) = \Pi_{i=1}^{n} f^{\delta_i}(x_i)\bar{F}^{1-\delta_i}(x_i), \qquad \delta_i = 1_{X_i<C_i}.$$

In this case, the total log-likelihood function of $\Theta = (\alpha, \lambda, \xi)^T$ is obtained as follow

$$\begin{aligned} L_n(\Theta) &= \sum_{i=1}^{n} \delta_i \ln h(x_i, \Theta) + \sum_{i=1}^{n} \ln S(x_i, \Theta) \\ &= \sum_{i=1}^{n} \delta_i \left[\begin{array}{c} 2\ln\alpha + \ln\lambda + \ln g(x_i,\xi) + (2\lambda-1)\ln G(x_i,\xi) \\ -(2\lambda+1)\ln \bar{G}(x_i,\xi) - \ln\left(1+\alpha\left[M(x_i,\xi)\right]^{\lambda}\right) \end{array} \right] \\ &\quad + \sum_{i=1}^{n} \ln\left(1+\alpha\left[M(x_i,\xi)\right]^{\lambda}\right) - \alpha\sum_{i=1}^{n}\left[M(x_i,\xi)\right]^{\lambda} \end{aligned}$$

Where $M(x,\xi) = \frac{G(x,\xi)}{\bar{G}(x,\xi)}$. The components of the score functions are

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{n} \delta_i \left[\frac{1}{\alpha} - \frac{\left[M(x_i,\xi)\right]^{\lambda}}{1+\alpha\left[M(x_i,\xi)\right]^{\lambda}} \right] + \sum_{i=1}^{n} \frac{\left[M(x_i,\xi)\right]^{\lambda}}{1+\alpha\left[M(x_i,\xi)\right]^{\lambda}} - \sum_{i=1}^{n}\left[M(x_i,\xi)\right]^{\lambda}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda} &= \sum_{i=1}^{n} \delta_i \left[\frac{1}{\lambda} + 2\ln M(x_i,\xi) - \frac{\alpha\left[M(x_i,\xi)\right]^{\lambda}\ln M(x_i,\xi)}{1+\alpha\left[M(x_i,\xi)\right]^{\lambda}} \right] \\ &\quad + \sum_{i=1}^{n} \frac{\alpha\left[M(x_i,\xi)\right]^{\lambda}\ln M(x_i,\xi)}{1+\alpha\left[M(x_i,\xi)\right]^{\lambda}} - \alpha\sum_{i=1}^{n}\left[M(x_i,\xi)\right]^{\lambda}\ln M(x_i,\xi) \end{aligned}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \xi_k} &= \sum_{i=1}^{n} \delta_i \left[\frac{g'_k(x_i,\xi)}{g(x_i,\xi)} + (2\lambda-1)\frac{G'_k(x_i,\xi)}{G(x_i,\xi)} + (2\lambda+1)\frac{G'_k(x_i,\xi)}{\bar{G}(x_i,\xi)} - \frac{\alpha\lambda\left[M(x_i,\xi)\right]^{\lambda-1}M'_k(x_i,\xi)}{1+\alpha\left[M(x_i,\xi)\right]^{\lambda}} \right] \\ &\quad + \sum_{i=1}^{n} \frac{\alpha\lambda\left[M(x_i,\xi)\right]^{\lambda-1}M'_k(x_i,\xi)}{1+\alpha\left[M(x_i,\xi)\right]^{\lambda}} - \alpha\lambda\sum_{i=1}^{n}\left[M(x_i,\xi)\right]^{\lambda-1}M'_k(x_i,\xi), \end{aligned}$$

As in complete data case, the maximum likelihood estimators can not be obtained in their explicit form, so numerical methods are required.

4 The ND-generalized Linear Exponential Family

The interest of ND-generalized linear exponential family resides in the fact that this one contains several sub-models which havevarious applications. In this section, maximum likelihood estimation method is used to determine the unknown parameters and a modified chi-square test statistic is constructed to fit the proposed model when data are right censored. If the rv X has the generalized linear exponential (GLE) distribution with cdf given by

$$G(x) = 1 - e^{-\left(\sigma x + \frac{\theta x^2}{2}\right)^{\beta}}. \tag{21}$$

For positive parameters β, σ and θ. The pdf of GLE distribution is given by

$$g(x) = \beta\left(\sigma + \theta x\right)\left(\sigma x + \frac{\theta x^2}{2}\right)^{\beta-1} e^{-\left(\sigma x + \frac{\theta x^2}{2}\right)^{\beta}}, x > 0. \tag{22}$$

Thus, substituting (21) into (1), we obtain

$$F(x;\Theta) = 1 - \left(1 + \alpha\left[e^{\left(\sigma x + \frac{\theta x^2}{2}\right)^{\beta}} - 1\right]^{\lambda}\right)\exp\left\{-\alpha\left[e^{\left(\sigma x + \frac{\theta x^2}{2}\right)^{\beta}} - 1\right]^{\lambda}\right\}, \qquad (23)$$

Where $\Theta = (\alpha, \lambda, \sigma, \theta, \beta)$, is the vector of parameters. The rv X has cdf (23) is called ND-generalized linear exponential $(ND - GLE)$ **family** with five parameters $(\alpha, \lambda, \sigma, \theta, \beta)$. The pdf and the hazard rate functions of $ND - GLE$ can be derived easily. The proposed model have increasing, decreasing and cup shape failure rate functions. Figures (1) and (2) illustrate the graphical behavior of the pdf and hazard rate function of $ND - GLE$ for selected values of the parameters.

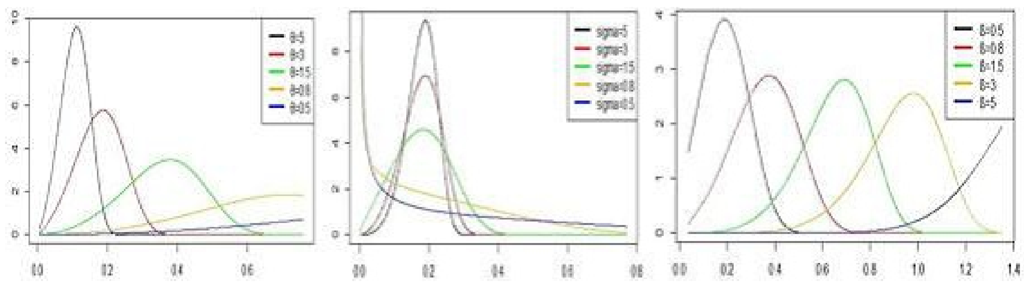

Figure 1: pdf curves of ND-GLE

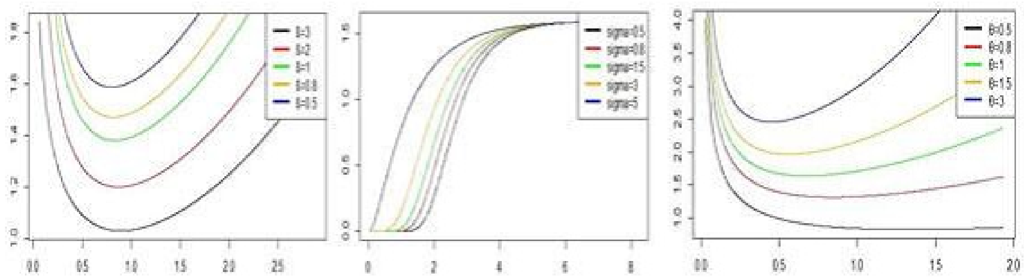

Figure 2: hasard rate function curves of ND-GLE

Several new distributions are deduced from the proposed model. The corresponding cdf are given in Table 1.

Table 1: cdf of sub models of ND-GLE **family**

	Name	α	λ	σ	θ	β	$S(x) = 1 - F(x)$
1	ND-linear exponential	–	–	–	–	1	$\left(1+\alpha\left[e^{\sigma x+\frac{\theta x^2}{2}}-1\right]^{\lambda}\right)e^{-\alpha\left[e^{\sigma x+\frac{\theta x^2}{2}}-1\right]^{\lambda}}$
2	ND-exponential	–	–	–	0	1	$\left(1+\alpha\left[e^{\sigma x}-1\right]^{\lambda}\right)e^{-\alpha[e^{\sigma x}-1]^{\lambda}}$
3	ND-Weibull	–	–	$\frac{1}{\sigma}$	0	–	$\left(1+\alpha\left[e^{(x/\sigma)^{\beta}}-1\right]^{\lambda}\right)e^{-\alpha\left[e^{(x/\sigma)^{\beta}}-1\right]^{\lambda}}$
4	ND-Rayleigh	–	–	0	–	1	$\left(1+\alpha\left[e^{\frac{\theta x^2}{2}}-1\right]^{\lambda}\right)e^{-\alpha\left[e^{\frac{\theta x^2}{2}}-1\right]^{\lambda}}$
5	ND-Modified Weibull	–	–	–	2θ	1	$\left(1+\alpha\left[e^{\sigma x+\theta x^2}-1\right]^{\lambda}\right)e^{-\alpha\left[e^{\sigma x+\theta x^2}-1\right]^{\lambda}}$
6	ND-weighted Weibull	–	–	$\sigma^{\frac{1}{\beta}}$	0	–	$\left(1+\alpha\left[e^{\sigma x^{\beta}}-1\right]^{\lambda}\right)e^{-\alpha\left[e^{\sigma x^{\beta}}-1\right]^{\lambda}}$

5 Maximum Likelihood Estimation for $ND - GLE$

5.1 Maximum likelihood estimators for complete data

As shown in section 4, the log-likelihood function of the $ND - GLE$ distribution can be written as:

$$\ell_n(\Theta) = 2n\ln\alpha + n\ln\lambda\beta + \sum_{i=1}^{n}\ln(\sigma+\theta x_i) + (\beta-1)\sum_{i=1}^{n}\ln u_i + \sum_{i=1}^{n}u_i^{\beta} - \alpha\sum_{i=1}^{n}\varphi_i^{\lambda} + (2\lambda-1)\sum_{i=1}^{n}\ln\varphi_i$$

Where $u_i(\sigma,\theta) \equiv u_i = \sigma x_i + \frac{\theta x_i^2}{2}$, $\varphi_i(\sigma,\theta,\beta) \equiv \varphi_i = e^{u_i^{\beta}} - 1$ And score equations given below are solved numerically

$$\frac{\partial\ell}{\partial\alpha} = \frac{2n}{\alpha} - \sum_{i=1}^{n}\varphi_i^{\lambda}$$

$$\frac{\partial\ell}{\partial\lambda} = \frac{n}{\lambda} - \alpha\sum_{i=1}^{n}\varphi_i^{\lambda}\ln\varphi_i + 2\sum_{i=1}^{n}\ln\varphi_i$$

$$\frac{\partial\ell}{\partial\beta} = \frac{n}{\beta} + \sum_{i=1}^{n}u_i^{\beta}\ln u_i - \alpha\lambda\sum_{i=1}^{n}u_i^{\beta}\ln(u_i)\,e^{u_i^{\beta}}\varphi_i^{\lambda-1} + \sum_{i=1}^{n}\ln u_i + (2\lambda-1)\sum_{i=1}^{n}\frac{u_i^{\beta}\ln(u_i)\,e^{u_i^{\beta}}}{\varphi_i}$$

$$\frac{\partial\ell}{\partial\sigma} = \sum_{i=1}^{n}\frac{1}{\sigma+\theta x_i} + \beta\sum_{i=1}^{n}x_i u_i^{\beta-1} - \alpha\lambda\beta\sum_{i=1}^{n}x_i u_i^{\beta-1}e^{u_i^{\beta}}\varphi_i^{\lambda-1} + (\beta-1)\sum_{i=1}^{n}\frac{x_i}{u_i} + (2\lambda-1)\sum_{i=1}^{n}\frac{x_i u_i^{\beta-1}e^{u_i^{\beta}}}{\varphi_i}$$

$$\frac{\partial\ell}{\partial\theta} = \sum_{i=1}^{n}\frac{x_i}{\sigma+\theta x_i} + \frac{\beta}{2}\sum_{i=1}^{n}x_i^2 u_i^{\beta-1} - \frac{\alpha\lambda\beta}{2}\sum_{i=1}^{n}x_i^2 u_i^{\beta-1}e^{u_i^{\beta}}\varphi_i^{\lambda-1} + (\beta-1)\sum_{i=1}^{n}\frac{x_i^2}{2u_i} + \frac{(2\lambda-1)\beta}{2}\sum_{i=1}^{n}\frac{x_i^2 u_i^{\beta-1}e^{u_i^{\beta}}}{\varphi_i}$$

5.2 Maximum Likelihood Estimation with right censorship

In presence of right censorship, le loglikelihhod function of the $ND-GLE$ distribution is:

$$L_n(\Theta)=\sum_{i=1}^{n}\delta_i\left[\begin{array}{c}2\ln\alpha+\ln\lambda\beta+\ln(\sigma+\theta x_i)+(\beta-1)\ln u_i\\ +u_i^{\beta}+(2\lambda-1)\ln\varphi_i-\ln\left(1+\alpha\varphi_i^{\lambda}\right)\end{array}\right]+\sum_{i=1}^{n}\ln\left(1+\alpha\varphi_i^{\lambda}\right)-\alpha\sum_{i=1}^{n}\varphi_i^{\lambda}.$$

And score equations given below are solved numerically

$$\frac{\partial L}{\partial\alpha}=\sum_{i=1}^{n}\delta_i\left(\frac{2}{\alpha}-\frac{\varphi_i^{\lambda}}{1+\alpha\varphi_i^{\lambda}}\right)+\sum_{i=1}^{n}\frac{\varphi_i^{\lambda}}{1+\alpha\varphi_i^{\lambda}}-\sum_{i=1}^{n}\varphi_i^{\lambda}$$

$$\frac{\partial L}{\partial\lambda}=\sum_{i=1}^{n}\delta_i\left(\frac{1}{\lambda}+\frac{\ln\varphi_i\left(2+\alpha\varphi_i^{\lambda}\right)}{1+\alpha\varphi_i^{\lambda}}\right)+\alpha\sum_{i=1}^{n}\frac{\varphi_i^{\lambda}\ln\varphi_i}{1+\alpha\varphi_i^{\lambda}}-\alpha\sum_{i=1}^{n}\varphi_i^{\lambda}\ln\varphi_i$$

$$\begin{aligned}\frac{\partial L}{\partial\beta}&=\sum_{i=1}^{n}\delta_i\left(\frac{1}{\beta}+\ln u_i+\frac{u_i^{\beta}\ln u_i\left(2\lambda e^{u_i^{\beta}}-1\right)}{e^{u_i^{\beta}}-1}-\frac{\alpha\lambda u_i^{\beta}\ln(u_i)\,e^{u_i^{\beta}}\varphi_i^{\lambda-1}}{1+\alpha\varphi_i^{\lambda}}\right)\\&+\alpha\lambda\sum_{i=1}^{n}\frac{u_i^{\beta}\ln(u_i)\,e^{u_i^{\beta}}\varphi_i^{\lambda-1}}{1+\alpha\varphi_i^{\lambda}}-\alpha\lambda\sum_{i=1}^{n}u_i^{\beta}\ln(u_i)\,e^{u_i^{\beta}}\varphi_i^{\lambda-1}\end{aligned}$$

$$\begin{aligned}\frac{\partial L}{\partial\sigma}&=\sum_{i=1}^{n}\delta_i\left(\frac{1}{\sigma+\theta x_i}+\frac{(\beta-1)x_i}{u_i}+\frac{u_i^{\beta-1}\beta x_i\left(2\lambda e^{u_i^{\beta}}-1\right)}{\varphi_i}-\frac{\alpha\lambda\beta x_i u_i^{\beta-1}e^{u_i^{\beta}}\varphi_i^{\lambda-1}}{1+\alpha\varphi_i^{\lambda}}\right)\\&+\alpha\lambda\beta\sum_{i=1}^{n}\frac{x_i u_i^{\beta-1}e^{u_i^{\beta}}\varphi_i^{\lambda-1}}{1+\alpha\varphi_i^{\lambda}}-\alpha\lambda\beta\sum_{i=1}^{n}x_i u_i^{\beta-1}e^{u_i^{\beta}}\varphi_i^{\lambda-1}\end{aligned}$$

$$\begin{aligned}\frac{\partial L}{\partial\theta}&=\sum_{i=1}^{n}\delta_i\left(\frac{x_i}{\sigma+\theta x_i}+\frac{(\beta-1)x_i^2}{2u_i}+\frac{\beta x_i^2 u_i^{\beta-1}\left(2\lambda e^{u_i^{\beta}}-1\right)}{2\varphi_i}-\frac{\alpha\lambda\beta}{2}\frac{x_i^2 u_i^{\beta-1}e^{u_i^{\beta}}\varphi_i^{\lambda-1}}{1+\alpha\varphi_i^{\lambda}}\right)\\&+\frac{\alpha\lambda\beta}{2}\sum_{i=1}^{n}\frac{x_i^2 u_i^{\beta-1}e^{u_i^{\beta}}\varphi_i^{\lambda-1}}{1+\alpha\varphi_i^{\lambda}}-\frac{\alpha\lambda\beta}{2}\sum_{i=1}^{n}x_i^2 u_i^{\beta-1}e^{u_i^{\beta}}\varphi_i^{\lambda-1}\end{aligned}$$

6 Test statistic for right censored data

6.1 Modified chi-square goodness-of-fit test

In complete data case, different methods are used for selection the model used in the analysis which is not the case when data are censored. Based on the approach of Bagdonavicius and Nikulin (2011) a modified chi-square goodness-of-fit test statistic for $ND-GLE$ distribution is proposed in this work when data are right censored. To validate the null hypothesis that $X_1,...,X_n$, n i.i.d. random variables fit a parametric distribution F_0 with unkonwn parameters and right censorship such as:

$$H_0: \quad P(X_i \leq x \mid H_0) = F_0(x;\Theta),\ x \geq 0, \quad \Theta = (\Theta_1, ..., \Theta_s)^T\ , \Theta \subset R^s,$$

the authors proposed to group the sample into r classes I_j where $r \succ s$ and used the distances between the observed U_j and the expected e_j failure times to fall into the grouping intervals $I_j$to compare empirical and theoretical distributions. Based on the maximum likelihood estimators $\widehat{\Theta}$ of the unkonwn parameters on ungrouped data, the statistic test Y^2 is defined by

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

Where the quadratic form Q is given as

$$\begin{aligned}
Q &= W^T \widehat{G}^- W & \widehat{A}_j = U_j/n, \qquad & U_j = \sum_{i:X_i \in I_j} \delta_i, \\
W &= (W_1, ..., W_s)^T, \quad \widehat{G} = [\widehat{g}_{ll'}]_{sxs}, & & \widehat{g}_{ll'} = \widehat{i}_{ll'} - \sum_{j=1}^{r} \widehat{C}_{lj} \widehat{C}_{l'j} \hat{A}_j^{-1}, \\
\widehat{C}_{lj} &= \frac{1}{n} \sum_{i:x_i \in I_j} \delta_i \frac{\partial}{\partial \Theta} \ln h(x_i, \widehat{\Theta}), & & \widehat{i}_{ll'} = \frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{\partial \ln h(x_i, \widehat{\Theta})}{\partial \Theta_l} \frac{\partial \ln h(x_i, \widehat{\Theta})}{\partial \Theta_{l'}}, \\
\widehat{W}_l &= \sum_{j=1}^{r} \widehat{C}_{lj} \widehat{A}_j^{-1} Z_j, \quad l, l' = 1, ..., s & & Z_j = \frac{1}{\sqrt{n}} (U_j - e_j), \quad j = 1, 2, ..., r\ .
\end{aligned}$$

The distribution of the statistic Y^2 follows a chi-square distribution under the null hypothesis H_0. For more details on the description and applications of modified chi-square tests, one can consult Voinov *et al.*(2013) . Notice that this approach was used to provide test statistics criteria to fit several generalized models such the AFT- generalized inverse Weibull distribution Goual and Seddik-Ameur (2014), generalized exponential AFT distributions with censored data Seddik-Ameur and Aidi (2016), Bertholon model with censored dataChouia and Seddik-Ameur (2017), accelerated failure time and proportional hazard Weibull extension modelsSeddik-Ameur and Treidi (2018).

6.2 Criteria test for the $ND - GLE$ distribution

Suppose that a right censored sample $X_1, X_2, ..., X_n$ follows the $ND-GLE$ distribution. Each observation can be written as $x_i = \min(X_i, C_i)$ for $i = 1, ..., n$ where X_i are failure times and C_i censoring times. Let us consider τ is a finite time, and let's group observed data into $r > s$ sub-intervals $I_j = (a_{j-1}, a_j]$ of $[0, \tau]$. To validate the null hypothesis H_0 that this sample belongs to the $ND - GLE$ model, first we have to determine the estimated limit intervals a_j which are given by

$$\hat{a}_j = H^{-1}\left(\frac{E_j - \sum_{l=1}^{i-1} H\left(x_l, \widehat{\Theta}\right)}{n - i + 1}, \widehat{\Theta}\right), \qquad \hat{a}_r = \max\left(X_{(n)}, \tau\right),$$

Where $H\left(x_l,\widehat{\Theta}\right)$ reprents the cumulative hazard rate function of $ND-GLE$. We then obtain the numbers of observed and expected failure times U_j and e_j such as $e_j=\frac{E_r}{r}$ for any j, with $E_r=\sum_{i=1}^n H\left(x_i,\widehat{\Theta}\right)$ and

$$E_j=\frac{-j}{r-1}\sum_{i=1}^n \ln\left(\left(1+\alpha\left[e^{\left(\sigma x+\frac{\theta x^2}{2}\right)^\beta}-1\right]^\lambda\right)\exp\left\{-\alpha\left[e^{\left(\sigma x+\frac{\theta x^2}{2}\right)^\beta}-1\right]^\lambda\right\}\right),\qquad j=1,..r-1$$

The components of the quadratic form Q are obtained from the estimated information matrix $\hat{\imath}_{ll'}$ and the estimated matrix $\hat{C}$ with the following elements:

$$\hat{C}_{1j}=\sum_{i:x_i\in I_j}^n \delta_i\left(\frac{2}{\alpha}-\frac{\varphi_i^\lambda}{1+\alpha a_i^\lambda}\right)$$

$$\hat{C}_{2j}=\sum_{i:x_i\in I_j}^n \delta_i\left(\frac{1}{\lambda}+\frac{\ln\varphi_i\left(2+\alpha\varphi_i^\lambda\right)}{1+\alpha\varphi_i^\lambda}\right)$$

$$\hat{C}_{3j}=\sum_{i:x_i\in I_j}^n \delta_i\left(\frac{1}{\beta}+\ln u_i+\frac{u_i^\beta\ln u_i\left(2\lambda e^{u_i^\beta}-1\right)}{e^{u_i^\beta}-1}-\frac{\alpha\lambda u_i^\beta\ln\left(u_i\right)e^{u_i^\beta}\varphi_i^{\lambda-1}}{1+\alpha\varphi_i^\lambda}\right)$$

$$\hat{C}_{4j}=\sum_{i:x_i\in I_j}^n \delta_i\left(\frac{1}{\sigma+\theta x_i}+\frac{(\beta-1)x_i}{u_i}+\frac{u_i^{\beta-1}\beta x_i\left(2\lambda e^{u_i^\beta}-1\right)}{\varphi_i}-\frac{\alpha\lambda\beta x_i u_i^{\beta-1}e^{u_i^\beta}\varphi_i^{\lambda-1}}{1+\alpha\varphi_i^\lambda}\right)$$

$$\hat{C}_{5j}=\sum_{i:x_i\in I_j}^n \delta_i\left(\frac{x_i}{\sigma+\theta x_i}+\frac{(\beta-1)x_i^2}{2u_i}+\frac{\beta x_i^2 u_i^{\beta-1}\left(2\lambda e^{u_i^\beta}-1\right)}{2\varphi_i}-\frac{\alpha\lambda\beta}{2}\frac{x_i^2u_i^{\beta-1}e^{u_i^\beta}\varphi_i^{\lambda-1}}{1+\alpha\varphi_i^\lambda}\right)$$

So we deduced the statistic Y^2 for the $ND-GLE$ distribution when parameters are unknown and estimated by maximum likelihood method and data are right censored. This statistic follows a chi-squared distribution with r degrees of freedom.

$$Y_n^2\left(\widehat{\Theta}\right)=\sum_{j=1}^r\frac{\left(U_j-e_j\right)^2}{U_j}+\hat{W}^T\left[\hat{\imath}_{ll'}-\sum_{j=1}^r\hat{C}_{lj}\hat{C}_{l'j}\hat{A}_j^{-1}\right]^{-1}\hat{W}$$

7 Simulations

An extensive simulation study is carried out to show the performance of the techniques used and the feasability of the goodness-of-fit test developed in this work. At this end, we generated $N=10,000$ right censored samples with different sizes ($n=15,25,50,130,350,500$) from the $ND-GLE$ model with parameter values ($\alpha=1.5$, $\lambda=2$, $\beta=1$, $\sigma=1.2$,$\theta=0.9$). Firstly, we compute the MLEs of the unkown parameters, their bias and square mean errors, then we provide the criteria Y^2 of the corresponding samples.

7.1 Maximum likelihood estimation for ND-GLE

Using R statistical software and the Barzilai-Borwein (BB) algorithm (Varadhan and Gilbert, 2010), we calculate the maximum likelihood estimators of the unknown parameters, the corresponding bias and mean square errors (MSE) for complete samples (Table 2) and right censored samples (Table 3).

Table 2: MLEs of$\alpha, \lambda, \beta, \sigma, \theta$,bias and MSE for complete data

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\hat{\alpha} = 1.5$	1.5296	1.5246	1.5201	1.5137	1.5091	1.5033
$S.M.E$	0.0078	0.0067	0.0061	0.0052	0.0045	0.0037
$Bias$	0.1036	0.0925	0.0769	0.0593	0.0480	0.0356
$\hat{\lambda} = 2$	1.9723	1.9766	1.9804	1.9804	1.9929	1.9977
$S.M.E$	0.0064	0.0058	0.0047	0.0040	0.0031	0.0023
$Bias$	0.0914	0.0768	0.0657	0.0525	0.0382	0.02377
$\hat{\beta} = 1$	0.9699	0.9737	0.9793	0.9851	0.9909	0.9956
$S.M.E$	0.0077	0.0072	0.0064	0.0055	0.0048	0.0042
$Bias$	0.1198	0.1052	0.0924	0.0790	0.0675	0.0566
$\hat{\sigma} = 1.2$	1.2283	1.2224	1.2185	1.2121	1.2079	1.2017
$S.M.E$	0.0062	0.0051	0.0045	0.0036	0.0029	0.0021
$Bias$	0.1013	0.0867	0.0756	0.0624	0.0481	0.0376
$\hat{\theta} = 0.9$	0.9277	0.09234	0.9196	0.9129	0.9071	0.9023
$S.M.E$	0.0069	0.0063	0.0052	0.0045	0.0036	0.0028
$Bias$	0.1054	0.0908	0.0780	0.0646	0.0531	0.0422

The results of Table 2 and Table 3 show that the maximum likelihood estimators obtained are very close to the theoretical ones and their square mean errors are all less than $\mathbf{10^{-2}}$ even for little sample sizes which confirm the properties of maximum likelihood estimation method and the effectiveness of the procedure used confirm the consistency of the maximum likelihood estimators.

7.2 Test statistic Y^2

For testing the null hypothesis H_0 with respect to $ND-GLE$ distribution, we calculate the values of Y^2 of all the samples as shown above.
Then we compute the number of cases of rejection of the null hypothesis H_0 and we give a comparison between the different theoretical values of significance level α = 1%,5% and 10% and their empirical values. The observed results show that the empirical levels of significance, within simulation errors, are very close to those corresponding to the

Table 3: MLEs of $\alpha, \lambda, \beta, \sigma, \theta$,bias and MSE for censored

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\hat{\alpha}$	1.9456	1.9567	1.9657	1.9754	1.9876	1.9899
$S.M.E$	0.0046	0.0038	0.0029	0.0022	0.0018	0.00010
$Bias$	0.1077	0.0862	0.0632	0.0432	0.0312	0.0212
$\hat{\lambda}$	1.6254	1.6135	1.5987	1.5876	1.5678	1.5345
$S.M.E$	0.0056	0.0045	0.0038	0.0031	0.0028	0.0020
$Bias$	0.2382	0.1314	0.0921	0.0754	0.0523	0.0352
$\hat{\beta}$	3.5523	3.5415	3.5369	3.5221	3.5147	3.5012
$S.M.E$	0.0040	0.0032	0.0027	0.0019	0.0010	0.0005
$Bias$	0.0533	0.0450	0.0314	0.0294	0.0212	0.0122
$\hat{\sigma}$	0.8547	0.8496	0.8320	0.8298	0.8196	0.8007
$S.M.E$	0.0031	0.0029	0.0025	0.0019	0.0015	0.0004
$Bias$	0.0835	0.0619	0.0512	0.0478	0.0314	0.0201
$\hat{\theta}$	0.5430	0.5369	0.5322	0.5296	0.5201	0.5102
$S.M.E$	0.0021	0.0019	0.0015	0.0012	0.0008	0.0007
$Bias$	0.1231	0.0787	0.0612	0.0513	0.0274	0.0213

theoretical ones (Tables 4-11).
This study also allows us to calculate the values of Y^2 of the other members of this distribution family namely $ND-LE$, $ND-E$, $ND-W$, $ND-R$, $ND-MW$, $ND-WW$ The results are given in Tables.

Table 4: Simulated levels of significance for Y^2 against their theoretical values for ND-GLE distribution

$M = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\alpha = 1\%$	0.0042	0.0053	0.0075	0.0084	0.0089	0.0095
$\alpha = 5\%$	0.0210	0.0310	0.0298	0.0312	0.0475	0.0496
$\alpha = 10\%$	0.0812	0.0890	0.0901	0.0952	0.0991	0.1008

Case of ND linear exponential distribution $ND - LE$

Table 5: Simulated levels of significance for Y^2 against their theoretical values for $ND-LE$ distribution

$M = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\alpha = 1\%$	0.0045	0.0059	0.0087	0.0094	0.0119	0.0108
$\alpha = 5\%$	0.0245	0.0279	0.0399	0.0402	0.0421	0.0488
$\alpha = 10\%$	0.0741	0.0809	0.0821	0.0958	0.0991	0.1012

Case of ND exponential $ND - E$

Table 6: Simulated levels of significance for Y^2 against their theoretical values for $ND-E$ distribution

$M = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\alpha = 1\%$	0.0052	0.0065	0.0080	0.0105	0.0135	0.0107
$\alpha = 5\%$	0.0315	0.0387	0.0402	0.0441	0.0462	0.0474
$\alpha = 10\%$	0.0749	0.0835	0.0865	0.0898	0.0937	0.0995

Case of ND Weibull $ND - W$

Table 7: Simulated levels of significance for Y^2 against their theoretical values for ND-W distribution

$M = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\alpha = 1\%$	0.0046	0.0062	0.0079	0.088	0.0091	0.0102
$\alpha = 5\%$	0.0363	0.0373	0.0443	0.0450	0.0476	0.0493
$\alpha = 10\%$	0.0689	0.0752	0.0775	0.0845	0.0886	0.0978

Case of ND Rayleigh distribution $ND - R$

Table 8: Simulated levels of significance forY^2 against their theoretical values for ND-R distribution

$M = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\alpha = 1\%$	0.0039	0.0078	0.0092	0.0121	0.0104	0.0098
$\alpha = 5\%$	0.0351	0.0382	0.0398	0.0409	0.0434	0.0478
$\alpha = 10\%$	0.0678	0.0698	0.0745	0.0824	0.0886	0.0928

Case of Rayleigh distribution RD

Table 9: Simulated levels of significance for Y^2 against their theoretical values for ND-RD distribution

$M = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\alpha = 1\%$	0.0054	0.0068	0.0072	0.0085	0.0098	0.0109
$\alpha = 5\%$	0.0425	0.0450	0.0412	0.0470	0.0510	0.0550
$\alpha = 10\%$	0.0736	0.0735	0.0769	0.0810	0.0932	0.0996

Case of ND Modified Weibull $ND - MW$

Table 10: Simulated levels of significance for Y^2 against their theoretical values for ND-MW distribution

$M = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\alpha = 1\%$	0.0074	0.0082	0.0086	0.0094	0.0099	0.0112
$\alpha = 5\%$	0.0314	0.0357	0.0389	0.0412	0.0424	0.0496
$\alpha = 10\%$	0.0854	0.0874	0.0889	0.0945	0.0989	0.1025

Case of ND weighted Weibull $ND - WW$

Table 11: Simulated levels of significance for Y^2 against their theoretical values for ND-WW distribution

$M = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\alpha = 1\%$	0.0042	0.0032	0.0047	0.0071	0.085	0.0097
$\alpha = 5\%$	0.0241	0.0327	0.0389	0.0478	0.0498	0.0533
$\alpha = 10\%$	0.0714	0.0774	0.0889	0.0845	0.0939	0.1089

As expected the results confirm that the statistic test developed in this work can be used to check suitably the validity of the proposed model $ND-GLE$ and its sub-models.

7.3 Power study

To evaluate the powerful of the test statistic Y^2 proposed whether data fits well to the $ND - GLE$, we have considered some alternative distributions such as $ND - LE$, $ND - MW$, $ND - R$, $ND - WW$. The results are given in Tables 12, 13,14,15. Based on these results, It is observed that at level of significance 0.10, all test power values for Y^2 are higher than 80, then we can distinguish $ND-GLE$ from the competing distributions for all sample sizes.

Table 12: Power of $Y^2$for NDGLE against NDLE

$N=10,000$	$n_1=25$	$n_2=50$	$n_3=130$	$n_4=350$	$n_5=500$
$\alpha=1\%$	0.4685	0.5412	0.6714	0.7748	0.7594
$\alpha=5$	0.6857	0.6549	0.7584	0.8251	0.8529
$\alpha=10\%$	0.7849	0.8201	0.8235	0.9255	0.9238

Table 13: Power of Y^2 for NDGLE against NDMW

$N=10,000$	$n_1=25$	$n_2=50$	$n_3=130$	$n_4=350$	$n_5=500$
$\alpha=1\%$	0.4326	0.5493	0.6514	0.7241	0.7563
$\alpha=5\%$	0.5241	0.7245	0.7231	0.8235	0.8648
$\alpha=10\%$	0.6225	0.8102	0.8514	0.9543	0.9457

Table 14: Power of Y^2 for NDGLE against NDR

$N=10,000$	$n_1=25$	$n_2=50$	$n_3=130$	$n_4=350$	$n_5=500$
$\alpha=1\%$	0.4314	0.5456	0.6844	0.7451	0.8563
$\alpha=5\%$	0.5784	0.7745	0.7895	0.8745	0.9213
$\alpha=10\%$	0.78345	0.8151	0.8597	0.9523	0.9822

Table 15: Power of Y^2 for NDGLE against NDWW

$N=10,000$	$n_1=25$	$n_2=50$	$n_3=130$	$n_4=350$	$n_5=500$
$\alpha=1\%$	0.4632	0.6347	0.7741	0.7864	0.8201
$\alpha=5\%$	0.6487	0.7985	0.8245	0.8745	0.9132
$\alpha=10\%$	0.7965	0.8222	0.9351	0.9712	0.9999

8 Applications

The interest of the proposed family is illustrated by the analysis of a right skewed distribution from economic studies and a right censored data from survival analysis.

8.1 Complete data analysis

This data set is the actual taxes data set used by Mead (2016). The data consists of the monthly actual taxes revenue in Egypt from January 2006 to November 2010. The distribution is highly skewed to the right. The actual taxes revenue data (in 1000 million Egyptian pounds) are: 5.9, 20.4, 14.9, 16.2, 17.2, 7.8, 6.1, 9.2, 10.2, 9.6, 13.3, 8.5, 21.6, 18.5, 5.1, 6.7, 17, 8.6, 9.7, 39.2, 35.7, 15.7, 9.7, 10, 4.1, 36, 8.5, 8, 9.2, 26.2,21.9,16.7,21.3, 35.4, 14.3, 8.5, 10.6, 19.1, 20.5, 7.1, 7.7, 18.1, 16.5, 11.9, 7,8.6,12.5, 10.3, 11.2, 6.1, 8.4, 11, 11.6, 11.9, 5.2, 6.8, 8.9, 7.1, 10.8.
Using R statistical software and BB-solve algorithms, the MLEs of the different alternative distribution parameters are obtained for these data as follows

$$\begin{aligned} ND-GLE(\alpha &= 1.5962, \lambda = 1.7495, \beta = 0.9365, \sigma = 0.2513, \theta = 0.82365) \\ ND-LE(\alpha &= 1.4230, \lambda = 1.0352, \sigma = 0.4153, \theta = 0.6365) \\ ND-R(\alpha &= 2.0361, \lambda = 1.2395, \theta = 1.8236) \\ ND-MW(\alpha &= 0.9365, \lambda = 1.2846, \sigma = 0.7456, \theta = 1.6352) \\ ND-WW(\alpha &= 1.2748, \lambda = 0.9365, \beta = 0.8265, \sigma = 0.1423) \end{aligned}$$

In order to choose the model that best describes these data, we propose to calculate the values of the classical goodness-of-fit statistics on the one hand and the modified chi-square test proposed in this work on the other hand see Table 16.
From Table 16, one can see that the values of classical selection model criteria tests are the smallest ones for the $\mathbf{ND-GLE}$ distribution compared to the alternatives, although they are close to those of $\mathbf{ND-LE}$ distribution. However the value of the proposed goodness-of-fit test $\mathbf{Y}^2$ of the $\mathbf{ND-GLE}$ is much inferior to the ND-LE one which shows that the $\mathbf{ND-GLE}$ model is the best suited to describe these data.
To further validate the results obtained, the histogram plot of the dataset with the distributions compared is presented in Figure 3. The corresponding empirical cdf plots are presented in Figure 4.
On the basis of the obtained test statistics values, and the pdf and cdf curves Figure 4, we deduce that the $ND-GLE$ distribution can adequately describe these data.

8.2 Censored data analysis

We are considering sample data from 50 patients with acute myeloid leukemia,reported to the International Register of Bone Marrow Transplants. These patients had an allogeneic bone marrow transplant where the HLA (Histocompatibility Leukocyte Antigen) homolog marrow was used to rebuild their immune systems. The ordered data of this study are presented in Table 17, where (*)represents censorship.

Table 16: values of criteria statistics for model selection

$Distributions$	$-NLL$	AIC	$CAIC$	BIC	$HQIC$
$ND-GLE$	224.776	324.681	324.963	327.124	325.137
$ND-LE$	224.928	324.739	325.120	327.578	325.418
$ND-R$	225.837	325.246	326.584	328.976	327.826
$ND-MW$	225.613	324.982	325.767	328.212	326.775
$ND-WW$	226.934	327.461	327.134	330.537	229.224

$Distributions$	W	A	$K-S$	$p-value$	Y_n^2
$ND-GLE$	0.0778	0.5463	0.1634	0.5646	6.8569
$ND-LE$	0.0867	0.5786	0.1786	0.5424	7.6235
$ND-R$	0.1093	0.7896	0.2567	0.5120	9.8525
$ND-MW$	0.0946	0.7612	0.2077	0.5268	8.6235
$ND-WW$	0.1167	0.8467	0.3088	0.5012	10.0213

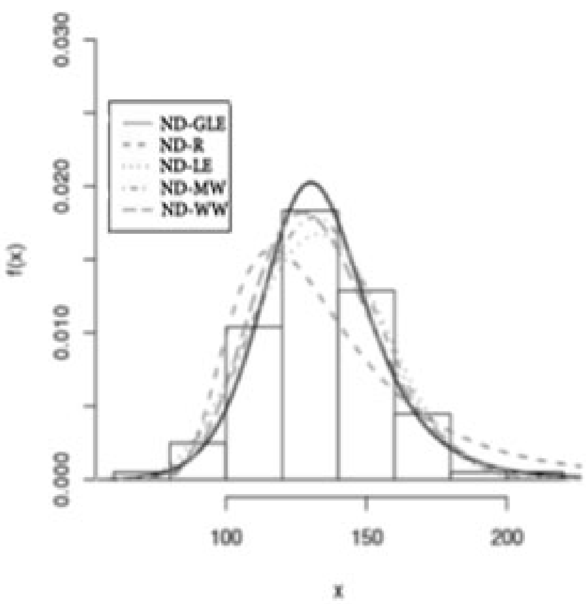

Figure 3: Histogram plot of the dataset with the compared distribution

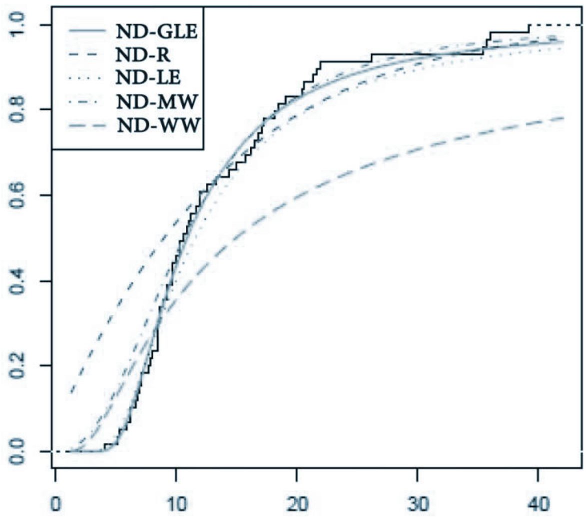

Figure 4: Empirical cdf of the dataset with the compared distributions

Table 17: Leukemia survival rate (in months) for autologous and allogenic transplants.

0.030	0.493	0.855	1.184	1.283	1.480	1.776	2.138	2.500
2.763	2.993	3.224	3.421	4.178	4.441*	5.691	5.855*	6.941*
6.941	7.993*	8.882	8.882	9.145*	11.480	11.513	12.105*	12.796
12.993*	13.849*	16.612*	17.138*	20.066	20.329*	22.368*	26.776*	28.717*
28.717*	32.928	33.783*	34.211	34.770*	39.539	41.118*	45.033	46.053*
46.941	48.289*	57.401*	58.322	60.625*				

Using BB solve software, we compute the maximum likelihood estimators of the unknown parameters

$$\widehat{\Theta} = \left(\hat{\alpha}, \hat{\lambda}, \hat{\beta}, \hat{\sigma}, \hat{\theta}\right)^T = (1.3956, 0.2463, 0.3748, 0.6395, 2.4628)^T.$$

Then, we grouped the observations into $r = 5$ intervals I_j. To calculate the criteria test Y_n^2, we must provide the values of all its components. The intermediaire results are summarized in Table 8.

So, we obtain the value of Y_n^2

$$Y_n^2 = X^2 + Q = 5.6138 + 2.9465 = 8.5603$$

As the value of $Y_n^2 = 8.5603$ is less than the critical value $\chi_5^2 = 11.0705$ (for significance level $\alpha = 0.05$), so we can say that these data can be fitted by the $ND{-}GLE$ distribution.

Table 18: values of $\widehat{a}_j, e_j, U_j, \hat{C}_{1j}, \hat{C}_{2j}, \hat{C}_{3j}, \hat{C}_{4j}, \hat{C}_{5j}$

$\hat{a}_j$	2.712	7.649	16.974	31.456	60.625
U_J	9	10	11	8	13
$\hat{C}_{1j}$	0.0415	0.0395	0.04126	0.0351	0.0512
$\hat{C}_{2j}$	−1.526	−1.4523	−2.8463	−3.4152	−1.9467
$\hat{C}_{3j}$	−4.1266	−3.8249	−2.5137	−4.6232	−5.8468
$\hat{C}_{4j}$	3.5267	3.8469	4.1253	3.4129	4.5396
$\hat{C}_{5j}$	−2.1632	−1.2863	−2.6439	−2.9465	−1.0954
e_j	4.2836	4.2836	4.2836	4.2836	4.2836

We also calculated the test statistics Y_n^2 to fit these data to the competing models. The results are given in Table 19 . As we can see from Table 19 and in Figure 5 5, the proposed

Table 19: Values of the test statistics $Y_n^2$for Leukemia data to fit different alternatives

Modeling distribution	Y_n^2
ND GLE	8.5603
ND linear exponential	11.523
ND Rayleigh	12.964
ND Modified Weibull	12.509
ND weighted Weibull	13.468

model $ND - GLE$ fits Leukemia data of Klein and Moeschberger (2003) for patients suffering from acute lymphoblastic leukemia better than all the competing models.

From Figure 5 one can see that leukemia data can be modeled by $\mathbf{ND - GLE}$ or $\mathbf{ND - MW}$ distributions, but the main purpose of the proposed test is to choose between close alternatives. In this case the test statistic value for $\mathbf{ND - GLE}$ model is $\mathbf{8, 5603}$ and that of $\mathbf{ND - MW}$ is $\mathbf{12, 509}$ which confirm that the suited model for leukemia data is the $\mathbf{ND - GLE}$.

The obtained empirical cumulative density function (cdf) of this dataset is very close to that of the $\mathbf{ND - GLE}$ distribution compared to its competitors.

9 Conclusion

A new lifetime distribution family which includes several generalized models is proposed. Its statistical properties and characteristics are studied. Maximum likelihood estimation

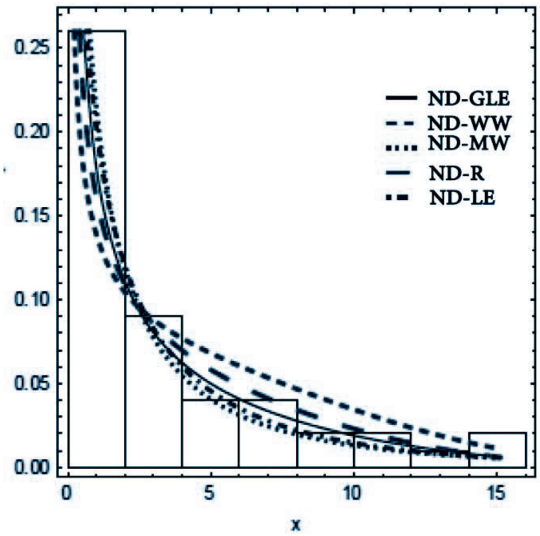

Figure 5: Histogram of Leukemia data versus different pdf curves

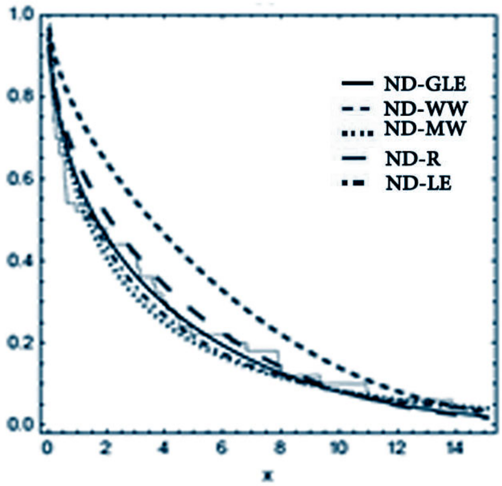

Figure 6: Empirical cdf of the dataset with the compared distributions.

method is used for determining the unknown parameters in complete and right censored cases. A new test statistic which able to fit data to the $\mathbf{ND-G}$ model is constructed when data are right censored. A power study showed that the proposed test can be used to distinguish between the new model and its alternatives. Real data sets are used to illustrate the usefulness of this distribution and the practicability of the proposed test.

Aknowledgements

The authors would like to thank the editors and referees for their suggestions which helped improve this work.

References

Alizadeh, M., Emadi, M., Doostparast, M., Cordeiro, G., Ortega, E., and Pescim, R. (2015). Kumaraswamy odd log-logistic family of distributions: Properties and applications. *Hacettepe Journal of Mathematics and Statistics*, 44(6):1491–1512.

Almongy, H. M., Almetwally, E. M., Aljohani, H. M., Alghamdi, A. S., and Hafez, E. (2021). A new extended rayleigh distribution with applications of covid-19 data. *Results in Physics*, 23:104012.

Alshenawy, R., Haj Ahmad, H., and Al-Alwan, A. (2022). Progressive censoring schemes for marshall-olkin pareto distribution with applications: Estimation and prediction. *Plos one*, 17(7):e0270750.

Alzaatreh, A., Lee, C., and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71(1):63–79.

Bagdonavicius, V. and Nikulin, M. (2011). Chi-squared goodness-of-fit test for right censored data. *International Journal of Applied Mathematics and Statistics*, 24(1):1–11.

Bagdonavičius, V. B., Levuliene, R. J., and Nikulin, M. S. (2013). Chi-squared goodness-of-fit tests for parametric accelerated failure time models. *Communications in Statistics-Theory and Methods*, 42(15):2768–2785.

Chouia, S. and Seddik-Ameur, N. (2017). A modified chi-square test for bertholon model with censored data. *Communications in Statistics-Simulation and Computation*, 46(1):593–602.

Doostmoradi, A. (2018). A new distribution with two parameters to lifetime data. *Biostatistics and Biometrics Open Access Journal*, 8(2):30–35.

Elgarhy, M., Haq, M., and Ozel, G. (2017). A new exponentiated extended family of distributions with applications. *Gazi University Journal of Science*, 30(3):101–115.

Elgarhy, M., Hassan, A. S., and Rashed, M. (2016). Garhy-generated family of distributions with application. *Mathematical Theory and Modeling*, 6(2):1–15.

Emam, W. and Tashkandy, Y. (2023). Modeling the amount of carbon dioxide emissions application: New modified alpha power weibull-x family of distributions. *Symmetry*, 15(2):366.

Famoye, F., Lee, C., and Olumolade, O. (2005). The beta-weibull distribution. *Journal of Statistical Theory and Applications*, 4(2):121–136.

Goual, H. and Seddik-Ameur, N. (2014). Chi-squared type test for the aft-generalized inverse weibull distribution. *Communications in statistics-Theory and Methods*, 43(13):2605–2617.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water resources research*, 15(5):1049–1054.

Khan, M. S., King, R., and Hudson, I. L. (2018). Kumaraswamy exponentiated chen distribution for modelling lifetime data. *Appl. Math*, 12(3):617–623.

Mead, M. (2016). On five-parameter lomax distribution: properties and applications. *Pakistan Journal of Statistics and Operation Research*, pages 185–199.

Nadarajah, S., Cordeiro, G. M., and Ortega, E. M. (2011). General results for the beta-modified weibull distribution. *Journal of Statistical Computation and Simulation*, 81(10):1211–1232.

Seddik-Ameur, N. and Aidi, K. (2016). Chi-square tests for generalized exponential distributions with censored data. *Electronic Journal of Applied Statistical Analysis*, 9(2):371–384.

Seddik-Ameur, N. and Treidi, W. (2018). On testing the fit of accelerated failure time and proportional hazard weibull extension models. *Journal of statistical theory and practice*, 12:397–411.

Silva, G. O., Ortega, E. M., and Cordeiro, G. M. (2010). The beta modified weibull distribution. *Lifetime data analysis*, 16:409–430.

Varadhan, R. and Gilbert, P. (2010). Bb: An r package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *Journal of statistical software*, 32:1–26.