Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2023;9(4): 484-499

Original Article

On The Modeling of Biomedical Data Sets with a New Generalized Exponentiated Exponential Distribution

Ismail Adekunle Kolawole[1], Ibrahim Sule[2], Olalekan Akanji Bello[3]

[1]Department of Mathematics and Statistics, School of Applied Sciences, Kaduna Polytechnic, Kaduna, Nigeria.
[2]Department of Mathematics and Statistics, Faculty of Sciences, Confluence University of Science and Technology, Osara, Nigeria.
[3]Department of Statistics, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Nigeria.

ARTICLE INFO

Received 08.05.2023
Revised 01.06.2023
Accepted 01.07.2023
Published 15.12.2023

Key words:
Skewness;
Kurtosis;
Biomedical;
Exponentiated exponential

ABSTRACT

Introduction: Many experts in the field of distribution theory have focused on extending probability distributions utilizing extended families of continuous distributions to improve the modeling adaptability of the conventional probability distributions.
Methods: This research employed a continuous family of probability distributions as introduced in the literature using the T-X methodology. Through this, some of statistical and mathematical properties of the model were derived and studied.
Results: This study had introduced a brand-new, five-parameter generalized exponentiated exponential distribution, which is a continuous probability distribution. With the aid of the quantile function, moments, moment generating function, survival function, hazard function, mean, and median, among other mathematical and statistical aspects, the new distribution's shape was deduced and researched. It was also possible to derive the probability density function for the minimum and maximum order statistics for this distribution. The method of maximum likelihood estimate was used to produce a conventional estimation of the unknown parameters. A simulation study was carried out to assess the efficiency and consistency of the estimation method used. To evaluate the fit and adaptability of the new model, it was applied to four real-world datasets in the field of medicine.
Conclusion: The analysis's findings demonstrated that the new model performs better than its counterparts and offers a better fit than the Topp-Leone exponentiated exponential, Topp-Leone Kumaraswamy exponential, exponentiated exponential, and exponential distributions.

Introduction

In the past ten years, researchers have been working to develop more robust and flexible distributions that would better reflect life's realities. To improve lifetime data analysis's capacity to match a variety of lifetime data with a high degree of skewness and kurtosis, numerous methods for creating new continuous distributions have been presented

*.Corresponding Author: ibrahimsule76@gmail.com

in the literature. This is accomplished by using new families of distributions and the inclusion of additional shape parameters to characterize data from a wide range of disciplines, including engineering, economics, biomedical sciences, environmental sciences, and others. Some families of distributions used in adding flexibility to the classical distributions proposed in the literature can be found in, etc.[1-11]

Probability distributions are used in biological statistics to model the uncertainty surrounding measurements, clinical trials, and other research investigations. The field of biomedical statistics is expanding quickly as more data become available and new statistical techniques are created that enable the analysis of more intricate topics relating to human health. Using the family of distributions put forth by[12-13] recently developed a three parameter exponential distribution and applied the model to Engineering and Medical data sets. The model was demonstrated to be very adaptable and to outperform its counterparts. For additional information on this topic, the reader is directed to.[14-22]

The aim of this work is to develop a new distribution called new generalized exponentiated exponential distribution, which is an extension of exponentiated exponential distribution, this research will use the exponentiated exponential model as a baseline to the family of distributions developed by.[3] To assess the applicability and flexibility of the proposed model, it is fitted to four actual data sets derived from the field of biomedical sciences.

Methods

New Generalized Exponentiated Exponential (NGEtEx) Distribution

This section derives the new model using the family of distributions proposed by.[3] The cumulative distribution function (cdf) and the probability density function (pdf) of the family are respectively given as:

$$F(y)=\left[1-\left[1-G(y)\right]^{\gamma\sigma}\right]^{\rho} \quad (1)$$

and

$$f(y)=\gamma\sigma\rho g(y)\left[1-G(y)\right]^{\gamma\sigma-1}\left[1-\left[G(y)\right]^{\gamma\sigma}\right]^{\rho-1} \quad (2)$$

where $g(y)$ and $G(y)$ are the pdf and cdf of the baseline distribution.

$y > 0$ and $\gamma,\sigma,\rho>0$

The cdf and pdf of exponentiated exponential distribution are given respectively as:

$$G(y)=\left[1-e^{-\beta y}\right]^{\theta} \quad (3)$$

and

$$g(y)=\beta\theta e^{-\beta y}\left[1-e^{-\beta y}\right]^{\theta-1} \quad (4)$$

$y > 0$ and $\beta,\theta>0$

The new generalized exponentiated exponential distribution is obtained by inserting equation (3) into equation (1) as

$$F(y)=\left[1-\left[1-\left[1-e^{-\beta y}\right]^{\theta}\right]^{\sigma\gamma}\right]^{\rho} \quad (5)$$

On differentiating equation (5) with respect to y, we have

$$f(y)=\rho\sigma\gamma\theta\beta e^{-\beta y}\left[1-e^{-\beta y}\right]^{\theta-1}\left[1-\left[1-e^{-\sigma y}\right]^{\theta}\right]^{\gamma\sigma-1}\left[1-\left[1-\left[1-e^{-\beta y}\right]^{\theta}\right]^{\sigma\gamma}\right]^{\rho-1} \tag{6}$$

$$x>0,\sigma,\gamma,\rho,\beta,\theta>0$$

Where β is γ, ρ, β, θ the scale parameter and are the shape parameters respectively.

Expansion of Density

The expression for generalized binomial expansion is given as:

$$(1-y)^{\rho-1}=\sum_{i=0}^{\infty}\frac{(-1)^i\Gamma(\rho)}{i!\Gamma(\rho-i)}y^i \tag{7}$$

Using a generalized binomial expansion given in equation (7), we have the expansion for the density of the new model given as:

$$f(y)=\rho\sigma\gamma\theta\beta\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{(-1)^{i+j+k}\Gamma(\rho)\Gamma(\sigma\gamma(i+1))\Gamma(\theta(j+1))}{i!j!k!\Gamma(\rho-i)\Gamma(\sigma\gamma(i+1)-j)\Gamma(\theta(j+1)-k)}\left[e^{-\beta y}\right]^{j+1} \tag{8}$$

Properties of NGEtEx Distribution

In this section, the properties of the new model such as moments, mean, moment generating function, quantile function, median, hazard function, survival function and order statistics are derived and studied.

Moments

$$E(Y^r)=\int_0^{\infty}y^r f(y)dy \tag{9}$$

$$E(Y^r)=\rho\sigma\gamma\theta\beta\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{(-1)^{i+j+k}\Gamma(\rho)\Gamma(\sigma\gamma(i+1))\Gamma(\theta(j+1))\Gamma(r+1)}{i!j!k!\Gamma(\rho-i)\Gamma(\sigma\gamma(i+1)-j)\Gamma(\theta(j+1)-k)(j+1)^{r+1}\beta^r} \tag{10}$$

To obtain the mean, we set in equation (10)

$$E(Y)=\rho\gamma\theta\sigma\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{(-1)^{i+j+k}\Gamma(\rho)\Gamma(\sigma\gamma(i+1))\Gamma(\theta(j+1))\Gamma(2)}{i!j!k!\Gamma(\rho-i)\Gamma(\sigma\gamma(i+1)-j)\Gamma(\theta(j+1)-k)(j+1)^2} \tag{11}$$

Moment generating function (mgf)

$$M_y(t)=\int_0^{\infty}e^{ty}f(y)dy \tag{12}$$

where the expansion of $e^{ty}=\sum_{m=0}^{\infty}\frac{t^m y^m}{m!}$

Therefore, mgf of the new generalized exponentiated exponential distribution is given as

$$E(e^{ty})=\rho\sigma\gamma\theta\beta\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\frac{(-1)^{i+j+k}\Gamma(\rho)\Gamma(\sigma\gamma(i+1))\Gamma(\theta(j+1))t^m\Gamma(m+1)}{i!j!k!\Gamma(\rho-i)\Gamma(\sigma\gamma(i+1)-j)\Gamma(\theta(j+1)-k)(j+1)^{m+1}\beta^m} \tag{13}$$

Quantile function

Quantile function is the inverse of cdf of a distribution. The quantile function is obtained using

$$Q(u)=F^{-1}(u) \tag{14}$$

Applying equation (13) to the cdf of the new model, we have the quantile function given as

$$y=\frac{1}{\beta}\left\{-\log\left(1-\left(1-\left(1-u^{1/\rho}\right)^{1/\sigma\gamma}\right)^{1/\theta}\right)\right\} \tag{15}$$

The median is obtained by setting u = 0.5 in equation (15) given as

$$y_{median}=\frac{1}{\beta}\left\{-\log\left(1-\left(1-\left(1-0.5^{1/\rho}\right)^{1/\gamma\sigma}\right)^{1/\theta}\right)\right\} \tag{16}$$

Hazard function

Hazard function is given as

$$\mathrm{H}(y)=\frac{f(y)}{S(y)} \tag{17}$$

The hazard function of the NGEtEx distribution is given as

$$\mathrm{H}(y)=\frac{\rho\sigma\gamma\theta\beta e^{-\beta y}\left[1-e^{-\beta y}\right]^{\theta-1}\left[1-\left[1-e^{-\sigma y}\right]^{\theta}\right]^{\gamma\sigma-1}\left[1-\left[1-\left[1-e^{-\beta y}\right]^{\theta}\right]^{\sigma\gamma}\right]^{\rho-1}}{1-\left[1-\left[1-\left[1-e^{-\beta y}\right]^{\theta}\right]^{\sigma\gamma}\right]^{\rho}} \tag{18}$$

Survival function

It can be defined as

$$S(y) = 1 - F(y) \tag{19}$$

$$S(y;\sigma,\gamma,\rho)=1-\left[1-\left[1-\left[1-e^{-\beta y}\right]^{\theta}\right]^{\sigma\gamma}\right]^{\rho} \tag{20}$$

Order Statistics

Let $Y_1, Y_2, \ldots Y_n$ be n independent random variable from the NGEtEx distributions and let $Y_{(1)} \le Y_{(2)} \le \ldots \le Y_{(n)}$ be their corresponding order statistic. Let $F_{r:n}(y)$ and $f_{r:n}(y)$, $r=1, 2, 3, \ldots n$ denote the cdf and pdf of the r^{th} order statistics $F_{r:n}$ respectively. The pdf of the r^{th} order statistics of $Y_{r:n}$ is given as

$$f_{r:n}(y)=\frac{1}{B(r,n-r+1)}\sum_{i=0}^{n-r}(-1)^i[F(y)]^{r+i-1}f(y) \tag{21}$$

$$f_{r:n}(y)=\frac{\beta\theta\rho\sigma\gamma}{B(r,n-r+1)}\sum_{i=0}^{n-r}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^{i+j+k+l}\Gamma(\rho(r+i))\Gamma(\gamma\sigma(j+1))\Gamma(\theta(k+1))}{j!k!l!\Gamma(\rho(r+i)-j)\Gamma(\gamma\sigma(j+1)-k)\Gamma(\theta(k+1)-l)}\left[e^{-\beta y}\right]^{l+1} \tag{22}$$

Minimum order statistics

We set in equation (22) to obtain the minimum order statistics of the NGEtEx distribution.

$$f_{1:n}(y)=n\beta\theta\rho\sigma\gamma\sum_{i=0}^{n-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^{i+j+k+l}\Gamma(\rho(1+i))\Gamma(\gamma\sigma(j+1))\Gamma(\theta(k+1))}{j!k!l!\Gamma(\rho(1+i)-j)\Gamma(\gamma\sigma(j+1)-k)\Gamma(\theta(k+1)-l)}\left[e^{-\beta y}\right]^{l+1} \tag{23}$$

Maximum order statistics

We set $r = n$ in equation (22) to obtain the maximum order statistics of the NGEtEx distribution.

$$f_{n:n}(y)=n\beta\theta\rho\sigma\gamma\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^{j+k+l}\Gamma(n+i)\Gamma(\rho(j+1))\Gamma(2\gamma(k+1))}{j!k!l!\Gamma((n+i)-j)\Gamma(\rho(j+1)-k)\Gamma(2\gamma(k+1)-l)}\left[e^{-\beta y}\right]^{l+1} \tag{24}$$

The method of maximum likelihood estimation (MLE) is used in this section to estimate the unknown parameters of the NGEtEx distribution. For a random sample, $Y_1, Y_2, \ldots Y_n$ of size n from the NGEtEx(ρ, σ, γ, β, θ), the log-likelihood function L(ρ, σ, γ, β, θ), of (6) is given as

$$l = n\log\gamma + n\log\rho + n\log\sigma + n\log\theta + n\log\beta - \beta\sum_{i=1}^{n}y_i + (\theta-1)\sum_{i=1}^{n}\log\left[1-e^{-\beta y_i}\right] + (\sigma\gamma-1)\sum_{i=1}^{n}\log\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right] + (\rho-1)\sum_{i=1}^{n}\log\left[1-\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]^{\sigma\gamma}\right] \tag{25}$$

To obtain the estimate of the unknown parameters, the log-likelihood equation is differentiated with respect to each parameter (ρ, σ, γ, β, θ) and equate to zero as:

$$\frac{\partial l}{\partial \gamma}=\frac{n}{\gamma}+(\sigma-1)\sum_{i=1}^{n}\log\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]+(\rho-1)\sum_{i=1}^{n}\frac{\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]^{\sigma\gamma}\log\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]}{1-\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]^{\sigma\gamma}}=0 \quad (26)$$

$$\frac{\partial l}{\partial \sigma}=\frac{n}{\sigma}+(\gamma-1)\sum_{i=1}^{n}\log\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]+(\rho-1)\sum_{i=1}^{n}\frac{\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]^{\sigma\gamma}\log\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]}{1-\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]^{\sigma\gamma}}=0 \quad (27)$$

$$\frac{\partial l}{\partial \rho}=\frac{n}{\rho}+\sum_{i=1}^{n}\log\left[1-\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]^{\sigma\gamma}\right]=0 \quad (28)$$

$$\frac{\partial l}{\partial \beta}=\frac{n}{\beta}-\sum_{i=1}^{n}y_i+(\theta-1)\sum_{i=1}^{n}\left[\frac{\beta e^{\beta y_i}}{1-e^{\beta y_i}}\right]-(\sigma\gamma-1)\sum_{i=1}^{n}\left[\frac{\theta\beta e^{\beta y_i}\left[1-e^{\beta y_i}\right]^{\theta-1}}{1-\left[1-e^{\beta y_i}\right]^{\theta}}\right]+(\rho-1)\sum_{i=1}^{n}\left[\frac{\sigma\gamma\theta\beta e^{\beta y_i}\left[1-e^{\beta y_i}\right]^{\theta-1}\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]^{\sigma\gamma-1}}{1-\left[1-\left[1-e^{\beta y_i}\right]^{\theta}\right]^{\sigma\gamma}}\right]=0 \quad (29)$$

$$\frac{\partial l}{\partial \theta}=\frac{n}{\theta}+\sum_{i=1}^{n}\log\left[1-e^{-\beta y_i}\right]-(\sigma\gamma-1)\sum_{i=1}^{n}\left[\frac{\left[1-e^{\beta y_i}\right]^{\theta}\log\left[1-e^{\beta y_i}\right]}{1-\left[1-e^{\beta y_i}\right]^{\theta}}\right]+(\rho-1)\sum_{i=1}^{n}\left[\frac{\sigma\gamma\left[1-e^{\beta y_i}\right]^{\theta}\left[1-\left[1-e^{-\beta y_i}\right]^{\theta}\right]^{\sigma\gamma-1}\log\left[1-e^{\beta y_i}\right]}{1-\left[1-\left[1-e^{\beta y_i}\right]^{\theta}\right]^{\sigma\gamma}}\right]=0 \quad (30)$$

Equations (26), (27), (28), (29) and (30) do not have a simple analytical form and are therefore not tractable. As a result, we have to resort to non-linear estimation of the parameters using an iterative method.

Results

This section presents some plots of the new model to show the shapes of the model and also, the results obtained from the analysis with respect to the applications are presented.

Simulation and Applications

This section presents a simulation study to assess the efficiency of the estimation method used and also discussed the practical application of the new model to assess its flexibility, robustness and fit in modelling data arising from biomedical science.

Simulation Study

A simulation study is performed to assess the efficiency of the mle. The precision of the mle is studied through bias, and the root mean square error (RMSE) for taking different samples by considering different parameter values. We performed 1000 repetitions from NGEtEx distribution to quantify the bias and RMSE, by taking samples of sizes n = 20, 50, 150, 200, 250, 300 and the initial values of the model parameters (ρ, θ, σ, β, γ) are chosen as (1.0, 1.0, 1.0, 1.0, 2.5).

Using the Akaike information criterion (AIC), the performance of the new model is compared with other existing distributions that are consistent with the baseline distribution in terms of providing good parametric fit to the data sets considered.

$$AIC = -2ll + 2K \quad (31)$$

The model selection is carried out using the AIC. Where *ll* denotes the log-likelihood function evaluated at the maximum likelihood estimates, *K* is the number of parameters in the model.

The model with smallest value of AIC is chosen as the best model to fit the data set better. The comparators presented are: Topp-Leone Kumarasawamy exponentiated exponential (TLKEx), Topp-Leone exponentiated exponential (TLEtEx), Topp-Leone exponential (TLEx), exponentiated exponential (EtEx) and exponential (Ex) distributions.

Data set 1 was given by[23] and represents the lifetime data relating to relief times (in minutes) of patients receiving an analgesic.

1.1, 1.4, 1.3, 1.7, 1.9, 1.8, 1.6, 2.2, 1.7, 2.7, 4.1, 1.8, 1.5, 1.2, 1.4, 3, 1.7, 2.3, 1.6, 2.

Data set 2 has been used by[24] and represents the sum of skin folds in 202 athletes collected at the Australian Institute of Sports as.

28.0, 98, 89.0, 68.9, 69.9, 109.0, 52.3, 52.8, 46.7, 82.7, 42.3, 109.1, 96.8, 98.3, 103.6, 110.2, 98.1, 57.0, 43.1, 71.1, 29.7, 96.3, 102.8, 80.3, 122.1, 71.3, 200.8, 80.6, 65.3, 78.0, 65.9, 38.9, 56.5, 104.6, 74.9, 90.4, 54.6, 131.9, 68.3, 52.0, 40.8, 34.3, 44.8, 105.7, 126.4, 83.0, 106.9, 88.2, 33.8, 47.6, 42.7, 41.5, 34.6, 30.9, 100.7, 80.3, 91.0, 156.6, 95.4, 43.5, 61.9, 35.2, 50.9, 31.8, 44.0, 56.8, 75.2, 76.2,101.1, 47.5, 46.2, 38.2, 49.2, 49.6, 34.5, 37.5, 75.9, 87.2, 52.6, 126.4, 55.6, 73.9, 43.5, 61.8, 88.9, 31.0, 37.6,52.8, 97.9, 111.1, 114.0, 62.9, 36.8, 56.8, 46.5, 48.3, 32.6, 31.7, 47.8, 75.1, 110.7, 70.0, 52.5, 67, 41.6, 34.8, 61.8, 31.5, 36.6, 76.0, 65.1, 74.7, 77.0, 62.6, 41.1, 58.9, 60.2, 43.0, 32.6, 48, 61.2, 171.1, 113.5, 148.9, 49.9, 59.4, 44.5, 48.1, 61.1, 31.0, 41.9, 75.6, 76.8, 99.8, 80.1, 57.9, 48.4, 41.8, 44.5, 43.8, 33.7, 30.9, 43.3, 117.8, 80.3, 156.6, 109.6, 50.0, 33.7, 54.0, 54.2, 30.3, 52.8, 49.5, 90.2, 109.5, 115.9, 98.5, 54.6, 50.9, 44.7, 41.8, 38.0, 43.2, 70.0, 97.2, 123.6, 181.7, 136.3, 42.3, 40.5, 64.9, 34.1, 55.7, 113.5, 75.7, 99.9, 91.2, 71.6, 103.6, 46.1, 51.2, 43.8, 30.5, 37.5, 96.9, 57.7, 125.9, 49.0, 143.5, 102.8, 46.3, 54.4, 58.3, 34.0, 112.5, 49.3, 67.2, 56.5, 47.6, 60.4, 34.9.

Data set 3 was given by[25] and it represents the remission times (in months) of a random sample of one hundred and twenty-eight (128) bladder cancer patients.

0.08, 2.09, 3.48, 4.87, 6.94 , 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50, 2.46 , 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 3.88, 5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93, 11.79, 18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63, 22.69.

Data set 4 was given by[26] and it represents the survival times of one hundred and twenty-one (121) patients with breast cancer obtained from a large hospital in a period from 1929 to 1938.: The data set is as follows:

0.3, 0.3, 4.0, 5.0, 5.6, 6.2, 6.3, 6.6, 6.8, 7.4, 7.5, 8.4, 8.4, 10.3, 11.0, 11.8, 12.2, 12.3, 13.5, 14.4, 14.4, 14.8, 15.5, 15.7, 16.2, 16.3, 16.5, 16.8, 17.2, 17.3, 17.5, 17.9, 19.8, 20.4, 20.9, 21.0, 21.0, 21.1, 23.0, 23.4, 23.6, 24.0, 24.0, 27.9, 28.2, 29.1, 30.0, 31.0, 1.0, 32.0, 35.0, 35.0,

37.0, 37.0, 37.0, 38.0, 38.0, 38.0, 39.0, 39.0, 40.0, 40.0, 40.0, 41.0, 41.0, 41.0, 42.0, 43.0, 43.0, 43.0, 44.0, 45.0, 45.0, 46.0, 46.0, 47.0, 48.0, 49.0, 51.0, 51.0, 51.0, 52.0, 54.0, 55.0, 56.0, 57.0, 58.0, 59.0, 60.0, 60.0, 60.0, 61.0, 62.0, 65.0, 65.0, 67.0, 67.0, 68.0, 69.0, 78.0, 80.0, 83.0, 88.0, 89.0, 90.0, 93.0, 96.0, 103.0, 105.0, 109.0, 109.0, 111.0, 115.0, 117.0, 125.0, 126.0, 127.0, 129.0, 129.0, 139.0, 154.0.

Discussion

From Table 1, it is shown that the estimated RMSEs tend towards zero as sample size increases. Also, the biases decrease as the sample size increases. The numerical results presented revealed the consistency of the mles. Figures 1 and 2 show how the new model is shaped. The new model has symmetric, positive and negative skewness, increasing, decreasing, and constant shapes depending on the value of each parameter, as can be seen from the figures. It can also be seen that the figure 2 portraying the hazard function has an inverted bathtub shape. Due to these characteristics, the model can be used to represent various types of medical research data.

Tables 2, Table 3, Table 4, and Table 5 display the estimated values for each parameter and the models' goodness of fit. The effectiveness metric used to assess the goodness of fit is AIC. When the AIC values are lower, the model performs better. The NGEtEx distribution has the lowest AIC, which makes it more adaptable for modeling the data sets, as can be seen from the tables.

Figures 3, 4, 5, and 6 display the new model's shapes, fit, and adaptability in relation to the data sets under examination. The four data sets under examination are matched by the new model, making it a good fit for the data sets.

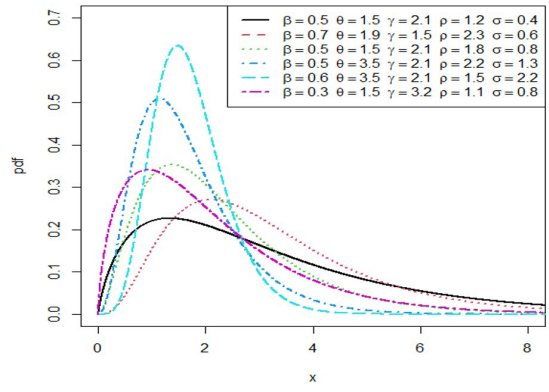

Figure 1. Plots of pdf of the NGEtEx distribution for different parameter values

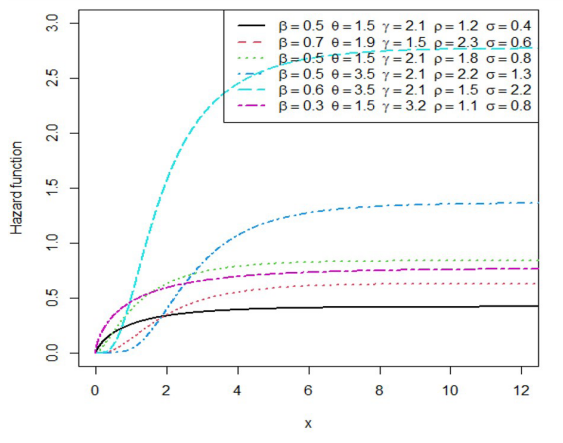

Figure 2. Plots of hazard function of the NGEtEx distribution for different parameter values

Conclusion

The novel generalized exponentiated exponential distribution, which is created in this work, expands the exponentiated exponential distribution. The survival function, hazard rate function, quantile function, mean, median, and order statistics could all be extracted from the new distribution. The maximum likelihood method was used to estimate the model parameters using the R package Adequacy

Table 1. MLEs, biases and RMSE for some values of parameters

N	Actual parameter value	Estimate	Bias	RMSE
20	ρ =1.0	1.3190	0.3190	0.9724
	θ=1.0	1.1092	0.1092	0.4813
	σ=1.0	1.1685	0.1685	1.0346
	β=1.0	1.4097	0.4097	1.2754
	γ=2.5	2.6945	0.1945	0.9377
50	ρ =1.0	1.2022	0.2022	0.7092
	θ=1.0	1.1064	0.1064	0.2851
	σ=1.0	1.0739	0.0739	0.8885
	β=1.0	1.2396	0.2396	0.9263
	γ=2.5	2.6309	0.1309	0.8306
100	α =1.0	1.2017	0.2317	0.6363
	θ=1.0	1.0426	0.0426	0.2158
	σ=1.0	1.0348	0.0348	0.7863
	β=1.0	1.2236	0.2236	0.7342
	γ=2.5	2.6139	0.1139	0.8130
150	ρ =1.0	1.0153	0.0153	0.6185
	θ=1.0	1.0212	0.0212	0.2016
	σ=1.0	1.0227	0.0227	0.6063
	β=1.0	1.2140	0.2140	0.6430
	γ=2.5	2.5669	0.0669	0.7345
200	ρ =1.0	1.0132	0.0132	0.5712
	θ=1.0	1.0122	0.0122	0.1887
	σ=1.0	1.0111	0.1111	0.5422
	β=1.0	1.1141	0.1141	0.6247
	γ=2.5	2.5404	0.0404	0.6354
250	ρ =1.0	1.0122	0.0122	0.4431
	θ=1.0	1.0101	0.0101	0.0459
	σ=1.0	1.0015	0.0015	0.3274
	β=1.0	1.0249	0.0249	0.5285
	γ=2.5	2.5310	0.0310	0.5293
300	ρ =1.0	1.0010	0.0010	0.2213
	θ=1.0	1.0100	0.0100	0.0213
	σ=1.0	1.0011	0.0011	0.1212
	β=1.0	1.0123	0.0123	0.3197
	γ=2.5	2.5101	0.5101	0.3018

Table 2. The models' MLEs and performance requirements based on data set 1

Models	$\hat{\beta}$	$\hat{\theta}$	$\hat{\gamma}$	$\hat{\rho}$	$\hat{\sigma}$	ll	AIC
NGEtEx	8.8613	1.4998	10.0351	28.3239	0.0239	-15.6627	41.3255
TLKEx	3.6229	20.9791	0.3069	-	2.3729	-16.8229	41.6458
TLEtEx	0.9618	0.1970	1.8931	-	-	-17.8662	41.7324
TLEx	0.6435	6.5611	-	-	-	-19.6989	43.3978
EtEx	1.4751	9.4746	-	-	-	-18.3418	40.6836
Ex	0.5263	-	-	-	-	-32.5263	67.3373

Table 3. The models' MLEs and performance requirements based on data set 2

Models	$\hat{\beta}$	$\hat{\theta}$	$\hat{\gamma}$	$\hat{\rho}$	$\hat{\sigma}$	ll	AIC
NGEtEx	0.1809	6.4969	0.4963	4.9049	0.4315	-955.4767	1920.9530
TLKEx	0.1704	12.8171	0.1225	-	0.2887	-958.8534	1925.7070
TLEtEx	0.0194	12.0124	0.7716	-	-	-957.5719	1921.1440
TLEx	0.0203	8.5858	-	-	-	-958.6400	1921.2800
EtEx	0.0203	8.5753	-	-	-	-958.6500	1921.3000
Ex	0.0144	-	-	-	-	-1057.3560	2116.7120

Table 4. The models' MLEs and performance requirements based on data set 3

Models	$\hat{\beta}$	$\hat{\theta}$	$\hat{\gamma}$	$\hat{\rho}$	$\hat{\sigma}$	ll	AIC
NGEtEx	0.4309	2.5357	0.1234	0.6824	1.9293	-411.8115	833.6230
TLKEx	0.3988	2.0262	0.1695	-	0.4392	-414.7476	837.4952
TLEtEx	0.0677	0.6014	1.8081	-	-	-414.5583	835.1167
TLEx	0.0606	1.2181	-	-	-	-415.0776	834.1552
EtEx	0.0606	1.2190	-	-	-	-415.0776	834.1554
Ex	0.1068	-	-	-	-	-416.3419	833.6838

Table 5. The models' MLEs and performance requirements based on data set 4

Models	$\hat{\beta}$	$\hat{\theta}$	$\hat{\gamma}$	$\hat{\rho}$	$\hat{\sigma}$	ll	AIC
NGEtEx	0.2265	0.2369	0.3539	2.2602	0.3771	-578.8936	1167.7870
TLKEx	0.0787	1.9700	0.1807	-	0.5819	-579.9097	1167.8190
TLEtEx	0.0142	1.1241	1.2356	-	-	-580.8998	1167.7996
TLEx	0.0135	1.4254	-	-	-	-582.7919	1169.4182
EtEx	0.0135	1.4248	-	-	-	-582.7091	1169.4182
Ex	0.0217	-	-	-	-	-584.4785	1170.9570

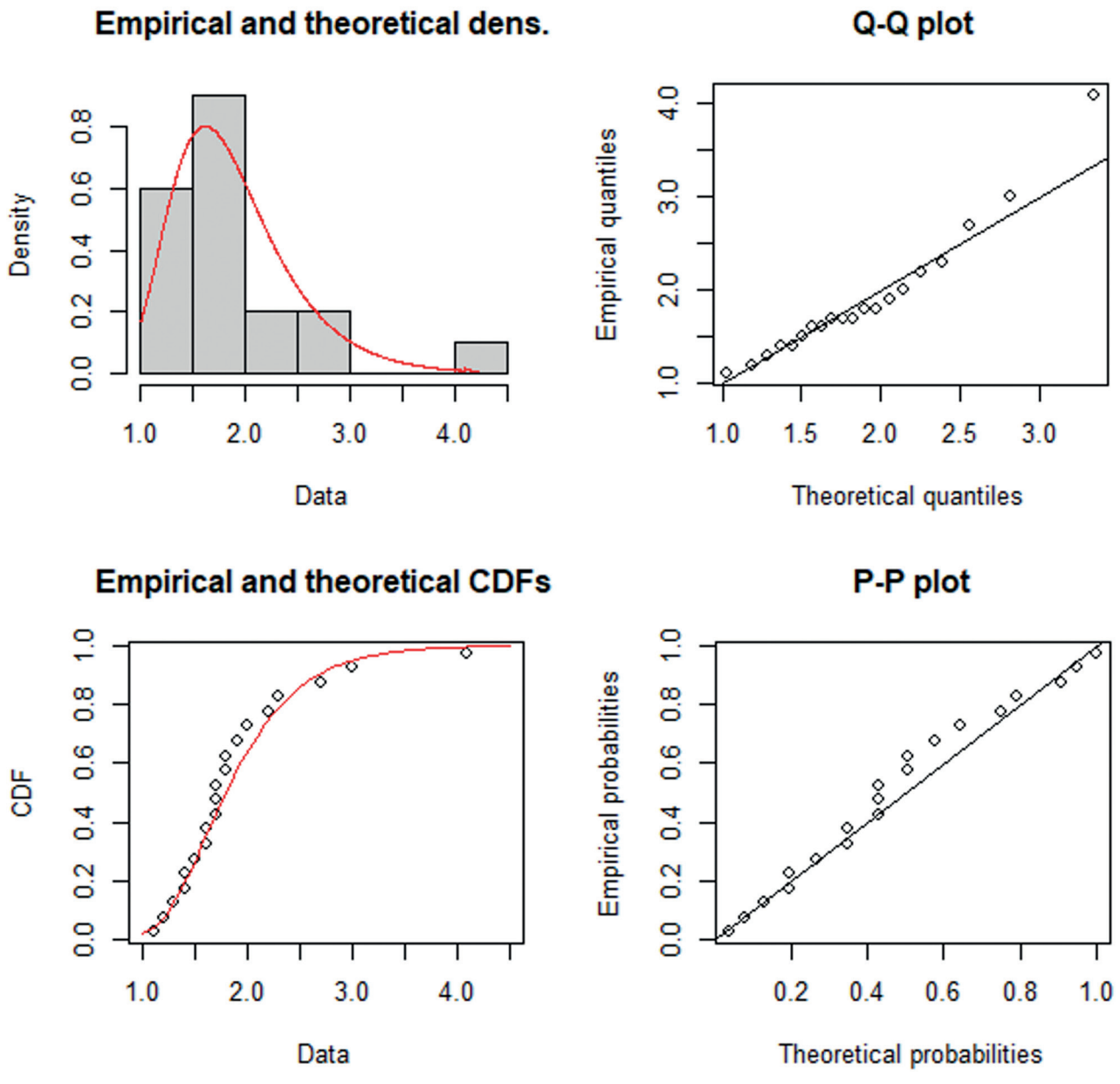

Figure 3. Estimated density plots for data set 1

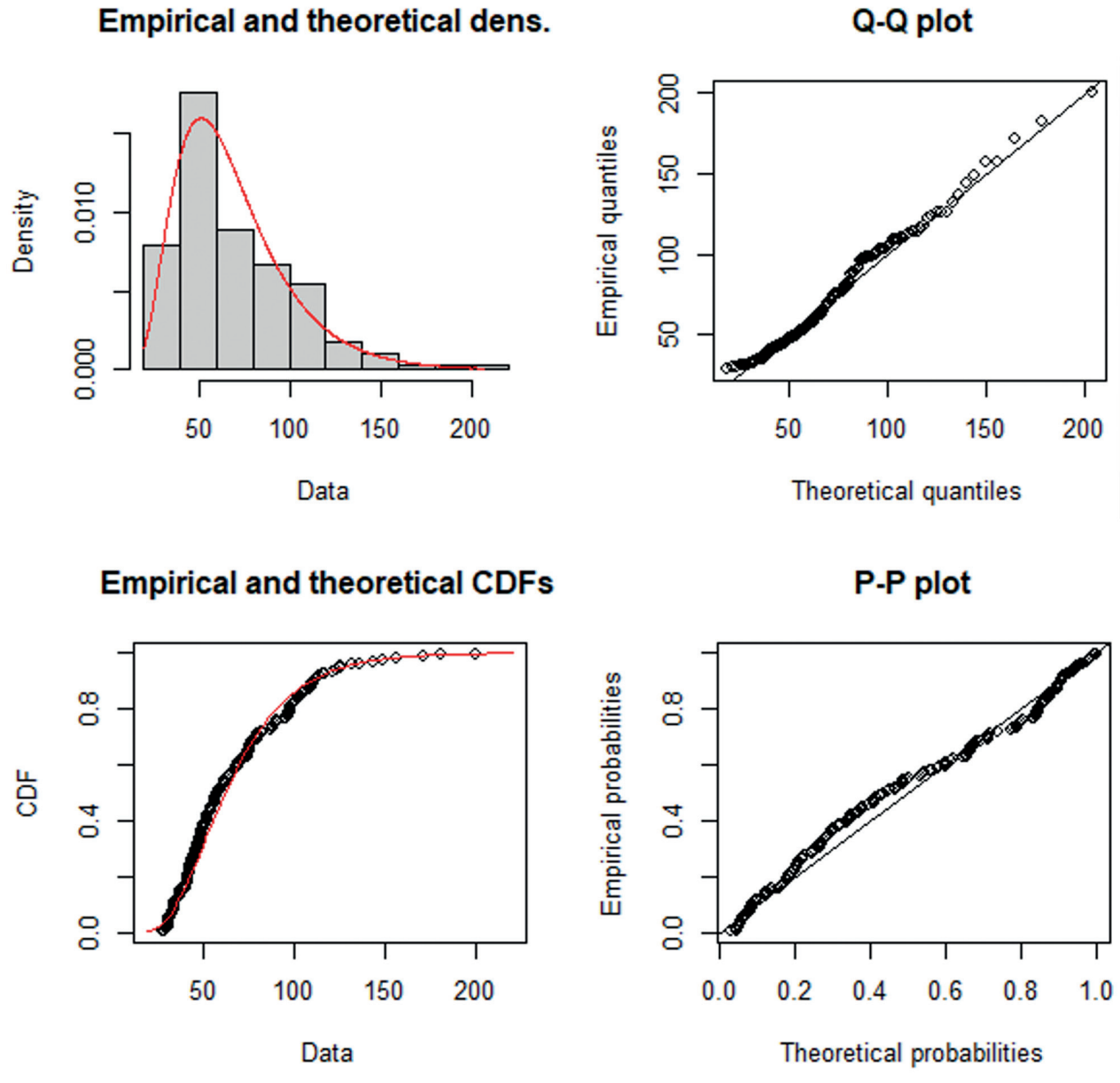

Figure 4. Estimated density plots for data set 2

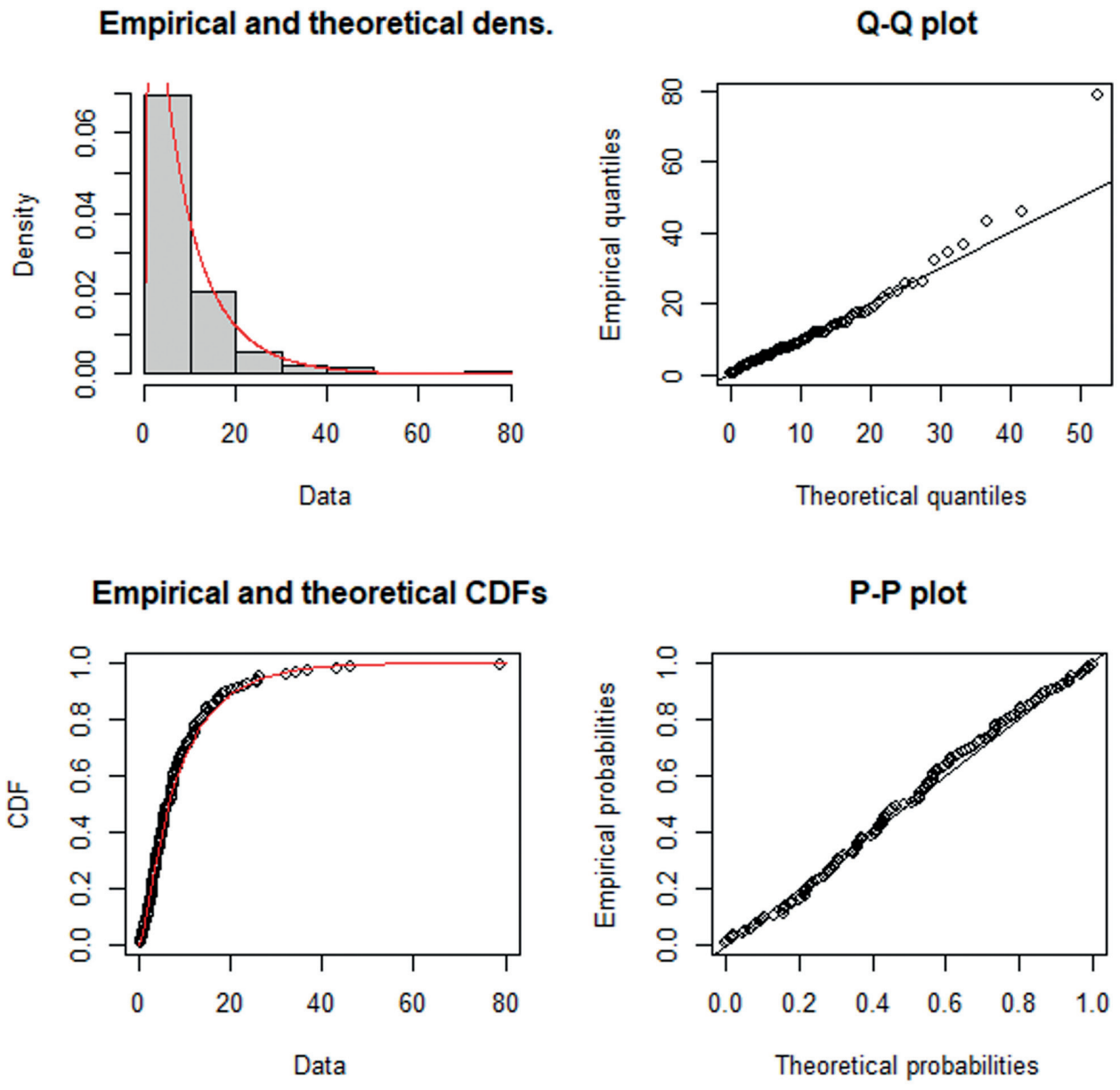

Figure 5. Estimated density plots for data set 3

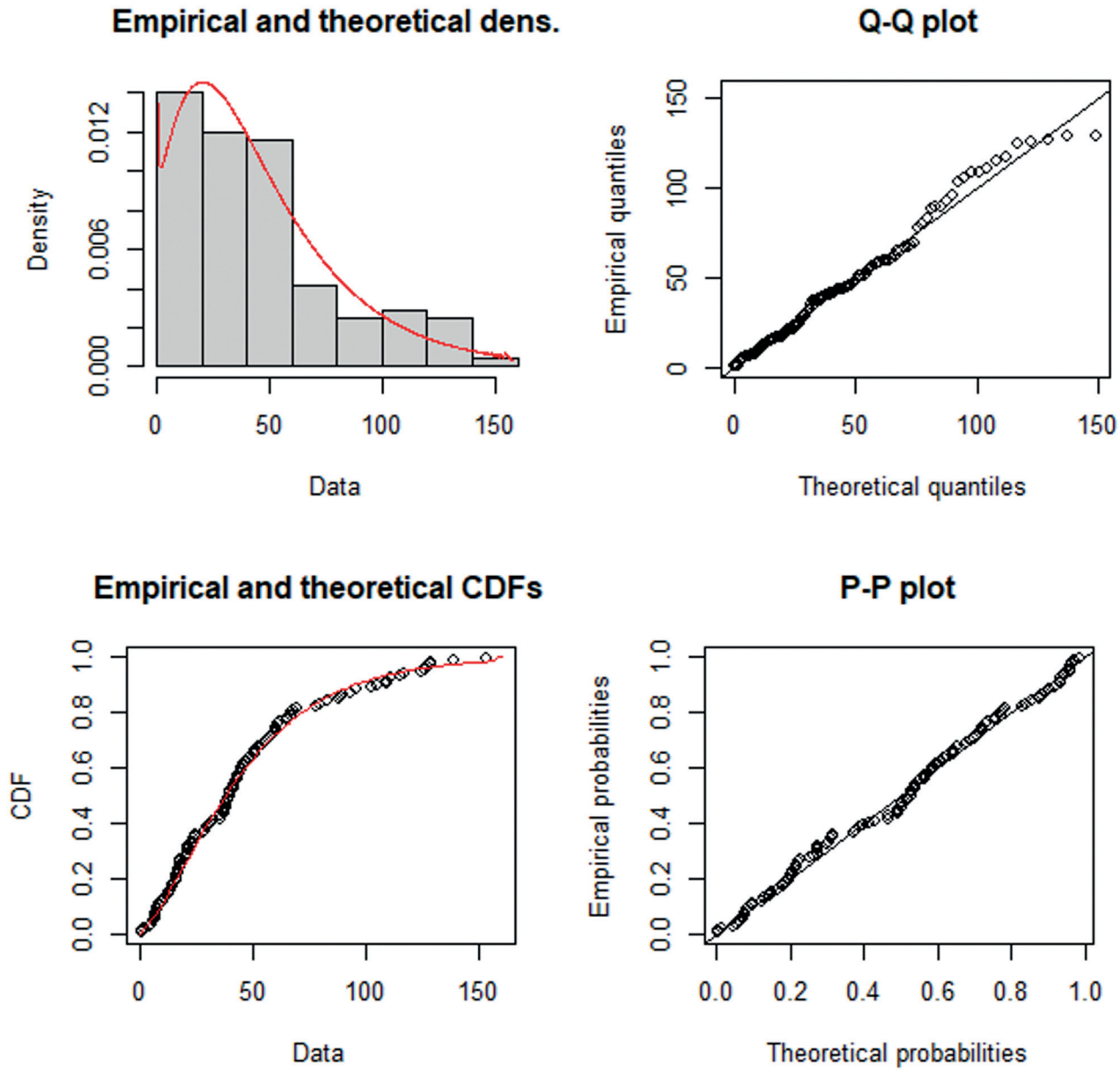

Figure 6. Estimated density plots for data set 4

Model. A simulation study is carried out to assess the consistency of the mle and the results of applying the proposed distribution to four actual data sets derived from the field of medical sciences are displayed in Table 1, Table 2, Table 3, and Table 4. The results showed that the four data sets under examination could be fitted substantially better and with far greater power by the new generalized exponentiated exponential distribution. Figures 3, 4, 5, and 6 for the four data sets' estimated densities, Q-Q plots, and P-P plots further demonstrate how flexible the new model is.

Conflict of interest

The authors have no conflict of interests to declare that are relevant to the content of this article.

References

1. Figueredo, A. J. and Wolf, P. S. A. (2009). Assortative pairing and life history strategy- a cross-cultural study. Human Nature, 20:317–330.

2. Joachims J. Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms, Kluwer, 2002.

3. Cordeiro, G. M., and de Castro, M., (2011). A new family of generalized distributions, Journal of Statistical Computation and Simulation, 81, 7, 883–898.

4. Al-Shomrani, A., Arif, O., Shawky, A., Hanif, S. and Shahbaz, M. Q., (2016). Topp-Leone family of distributions: Some properties and application, Pakistan Journal of Statistics and Operation Research. XII, 3, 443-451.

5. Elgarhy, M., Muhammad, A. H., Gamze, O. and Muhammad, A. N. (2017). A new exponentiated extended family of distributions with applications, Gazi University Journal of Science, 30(3): 101 – 115.

6. Hassan, S. and Nassr, S.G., Power Lindley-G Family of Distributions, Annals of Data Science, 6: 189-210, (2019).

7. Ibrahim, S., Doguwa, S.I., Audu, I. and Jibril, H.M., (2020). On the Topp Leone exponentiated-G Family of Distributions: Properties and Applications, Asian Journal of Probability and Statistics; 7(1): 1-15.

8. Ibrahim, S., Doguwa S. I., Audu, I. and Jibril, H. M., (2020). The Topp Leone Kumaraswamy-G Family of Distributions with Applications to Cancer Disease Data, Journal of Biostatistics and Epidemiology, 6(1):37-48.

9. Anzagra, L., Sarpong, S. and Nasiru, S., (2020). Odd Chen-G family of distributions, Annals of Data Science, doi.org/10.1007/s40745-020-00248-2.

10. Modi, K., Kumar, D. and Singh, Y., (2020). A New Family of Distribution with Application on Two Real Data sets on Survival Problem, Science and Technology Asia, 25(1): 1-10.

11. Rasheed, N. (2020). A new generalized-G class of distributions and its applications with Dagun distribution. Research Journal of Mathematical and Statistical Science,

8(3), 1-13.

12. Bello, O. A., Doguwa, S. I., Yahaya, A. and Jibril, H. M. (2020). A type I half Logistic exponentiated-G family of distributions: Properties and application, Communication in Physical Sciences, 7(3): 147 – 163.

13. Bello, O. A., Doguwa, S. I., Yahaya, A. and Jibril, H. M. (2021). A type II half Logistic exponentiated-G family of distributions with applications to survival analysis, FUDMA Journal of Sciences, 7(3): 147-163.

14. Yousof, H. M., Alizadeh, M., Jahanshahiand, S. M. A., Ramires, T. G., Ghosh, I. and Hamedani, G. G. (2017). The transmuted Topp-Leone G family of distributions: theory, characterizations and applications, Journal of Data Science, 15(4): 723–740.

15. Almutiry, W., Alahmadi, A. A., Elbatal, I., Ragab, I. E., Balogun, O. S. and Elgarhy, M. (2021). Application to Engineering and Medical Data Using Three-Parameter Exponential Model, Mobile Information Systems, vol. 2021, Article ID 9550156, 14 pages. https://doi.org/10.1155/2021/9550156.

16. Aldahlan, M. A. D. and Afify, A. Z. (2020). The odd exponentiated half-logistic exponential distribution: estimation methods and application to engineering data, Mathematics, 8(10), Article ID 1684.

17. Aldahlan, M. A. D. and Afify, A. Z. (2020). A new three-parameter exponential distribution with applications in reliability and engineering, The Journal of Nonlinear Science and Applications, 13(5): 258-269.

18. Ibrahim, M. and Yousof, H. (2020). Transmuted Topp-Leone Weibull lifetime distribution: statistical properties and different method of estimation, Pakistan Journal of Statistics and Operation Research, 16, 501–515.

19. Gupta, R. D. and Kundu, D., (1999). Generalized Exponential Distributions, Australian and New Zealand Journal of Statistics, 41(2): 173-188.

20. Gupta, R. D. and Kundu, D., (2001a). Exponentiated Exponential Family: An Alternative to Gamma and Weibull, Biometrical Journal, 33(1): 117-130.

21. Gupta, R.D. and Kundu, D. (2001b). Generalized Exponential Distributions: Different Methods of Estimation, Journal of Statistical Computation and Simulation, 69(4): 315-338.

22. Gupta, R. D. and Kundu, D. (2003). Discriminating Between the Weibull and the GE Distributions, Computational Statistics and Data Analysis, 43, 179-196.

23. Gupta, R. D. and Kundu, D. (2004). Discriminating Between the Gamma and Generalized Exponential Distributions, Journal of Statistical Computation and Simulation, 74(2): 107-121.

24. Gupta, R. D. and Kundu, D. (2007). Generalized exponential distribution: existing methods and recent developments, Journal of the Statistical Planning and Inference, 137(11): 3537 – 3547.

25. Gross, A. J. and Clark, V. A. (1975). Survival distributions: reliability applications in the biometrical sciences, John Wiley and Sons, Inc., New York.

26. Hosseini, B., Afshari, M. and Alizadeh M. (2018). The generalized odd gamma-G family of distributions: properties and applications. Austrian Journal of Statistics, 47(2):69-89.

27. Lee, E. T. and Wang, J. W: Statistical methods for survival data analysis (3rd Edition), John Wiley and Sons, New York, USA, 535 Pages,(2003) ISBN 0-471-36997-7.

28. Lee, E. T. (1992). Statistical methods for survival data analysis (2nd Edition), John Wiley and Sons Inc., New York, USA, 156 Pages.