Asian Journal of Probability and Statistics

Volume 26, Issue 12, Page 93-101, 2024; Article no.AJPAS.127675
ISSN: 2582-0230

Modeling COVID-19 Pandemic Data with New Pareto Model

Zakeia A. Al-Saiary [a*], Rana A. Bakoban [a] and Afnan S. Alamoudi [a]

[a] Department of Mathematics and Statistics, College of Science, University of Jeddah, Jeddah 22252, Saudi Arabia.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/ajpas/2024/v26i12687

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/127675

Original Research Article

Received: 03/10/2024
Accepted: 05/12/2024
Published: 09/12/2024

Abstract

This paper aims to find a statistical model for modeling the COVID-19 data. We deduced a robust and effective model for fitting the COVID 19 mortality. This model is a new Extended-Pareto distribution (NE-P). The maximum likelihood method is utilized to obtain the estimator of the parameters. A simulation was carried out using different sample sizes and different values of the parameters. In addition, the goodness of fit test statistics was calculated for proposed model compared with the baseline model to find out that our new model is the best for modeling data COVID-19.

Keywords: A new extended-pareto distribution; COVID 19 mortality; the maximum likelihood method; goodness of fit.

**Corresponding author: Email: zaalseare@uj.edu.sa;*

Cite as: *Al-Saiary, Zakeia A., Rana A. Bakoban, and Afnan S. Alamoudi. 2024. "Modeling COVID-19 Pandemic Data With New Pareto Model". Asian Journal of Probability and Statistics 26 (12):93-101. https://doi.org/10.9734/ajpas/2024/v26i12687.*

1 Introduction

Many important distributions are widely used in various statistical applications to model several life-time data in applied fields such as engineering, insurance, economics, medicine, and life testing, among others. The limitation of the standard distributions arouses the interest of finding new distributions by extending existing ones. One of the most important of these distributions is Pareto distribution (PD) that is well known in literature for its capability in modelling the heavy-tailed distributions that are mostly common in data on income distribution, economics, survival analysis. So, many authors have been interested in proposing new generalized forms of Pareto distribution which expanded the applications of this distribution to include medicine and health. For example, (Aldahlan et al. 2023) introduced the Marshall–Olkin Pareto Type-I (MOPTI) distribution. They studied the statistical properties of the MOPTI distribution. Also, they presented a simulation study and application on a real data set on breast cancer. (Caeiro and Mateus 2023) developed a new class of estimators for the parameters of Pareto type I distribution named the log-generalized probability-weighted moment (LGPWM). They found the proposed LGPWM estimators were capable of competing with the most commonly used estimation methods. (Jayakumar et al. 2018) presented a new four parameter distribution called New Generalized Pareto distribution, which was a generalization of the classical Pareto distribution. (Boumaraf et al. 2020) used nonlinear optimization methods to find the estimators of beta Pareto distribution.

The motivation behind this research is to expand the application areas of Pareto distribution to include medical data modeling. By introducing a new model as a generalization for Pareto distribution and demonstrate its flexibility in modeling COVID-19 data compared to the original Pareto distribution.

The probability density function (PDF) of Pareto distribution is given by:

$$g(x) = \frac{\beta}{x^{\beta+1}} \;;\; x \geq 1 \;; \beta > 0 \,. \tag{1}$$

and the cumulative distribution function (CDF) is:

$$G(x) = 1 - x^{-\beta} ; \quad x \geq 1 , \ \beta > 0. \tag{2}$$

where β is the scale parameter.

Moreover, many researchers have focused on finding generators for new distributions by finding new families, for example: (Sule et al. in 2020) and (Bantan et al. in 2020) In (Zichuan et al. 2020) the authors studied a new extended (NE-X) family of distributions which is the generator of our new model. The PDF of this NE-X distribution is given by:

$$f(x) = \frac{2\theta^2 g(x)G(x)\left(1-G^2(x)\right)^{\theta-1}}{\left(1-(1-\theta)G^2(x)\right)^{\theta+1}} ; \ \theta > 0 \ ; x \in R \,. \tag{3}$$

The CDF of NE-X distribution is:

$$F(x) = 1 - \left(\frac{1-G^2(x)}{1-(1-\theta)G^2(x)}\right)^{\theta} \ ; \theta > 0; x \in R. \tag{4}$$

Depending on Equations (1), (2) and the family in (Zichuan et al. 2020) we deduced a new distribution called it a New Extended-Pareto Distribution (NE-P). The PDF and CDF of NE-P distribution with two parameters(θ, β) is obtained respectively as:

$$f(x) = \frac{2\theta^2(\beta x^{-(\beta+1)})(1-x^{-\beta})\left(1-(1-x^{-\beta})^2\right)^{\theta-1}}{\left(1-(1-\theta)(1-x^{-\beta})^2\right)^{\theta+1}} ; \ \theta > 0, \ \beta > 0 ; \quad x \in R. \tag{5}$$

And

$$\text{F}(x) = 1 - \left(\frac{1-(1-x^{-\beta})^2}{1-(1-\theta)(1-x^{-\beta})^2}\right)^{\theta} \; ; \; \theta > 0, \; \beta > 0 \; ; x \in R \tag{6}$$

We can rewrite the PDF & CDF of NE-P distribution, using the series representation as follows

$$f(x) = 2\theta^2\beta \sum_{i=0}^{\infty} \sum_{j=0}^{\theta-1} \sum_{k=0}^{2(i+j)+1} \binom{i+\theta}{\theta} \binom{\theta-1}{j} \binom{2(i+j)+1}{k} \times (1-\theta)^i (-1)^{j+k} \left(x^{-\beta(k+1)-1}\right), \tag{7}$$

And

$$F(x) = 1 - [\sum_{i=0}^{\infty} \sum_{j=0}^{\theta} \sum_{k=0}^{2(i+j)} \binom{i+\theta-1}{\theta-1} \binom{\theta}{j} \binom{2(i+j)}{k} \times (1-\theta)^i (-1)^{j+k} x^{-\beta k}]. \tag{8}$$

In the article, we estimate the NE-P distribution parameters using the maximum likelihood estimation and carry out by different complete samples size of NE-P distribution. In addition, the goodness of fit test statistics calculate for proposed models to find out the best of it for data of Coronavirus disease (COVID-19).

2 Maximum Likelihood Estimation Method

This section presents the maximum likelihood estimator of the NE-P distribution parameters (θ, β). If $x_1, x_2, \ldots, x_n$ is a random sample from NE-P distribution, the Likelihood function is $L(\underline{x})$ can be obtained as:

$$L(\underline{x}) = (2\theta^2\beta)^n \prod_{i=1}^{n} x_i^{-(\beta+1)} \prod_{i=1}^{n} \left(1 - x_i^{-\beta}\right) \prod_{i=1}^{n} \left(1 - \left(1 - x_i^{-\beta}\right)^2\right)^{\theta-1} \times \prod_{i=1}^{n} \left(1 - (1-\theta)(1 - x_i^{-\beta})^2\right)^{-(\theta+1)}. \tag{9}$$

And the log-likelihood function is given as follows

$$l = \log(L) = n log(2) + n\log(\theta^2) + n\log(\beta) - (\beta+1)\sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} \log\left(1 - x_i{}^{-\beta}\right) + (\theta - 1)\sum_{i=1}^{n} \log(1 - \left(1 - x_i{}^{-\beta}\right)^2) - (\theta+1)\sum_{i=1}^{n} \log\left(1 - (1-\theta)\left(1 - x_i{}^{-\beta}\right)^2\right). \tag{10}$$

Differentiating (10) with respect to each of the parameters θ and β gives

$$\frac{\partial \, log \, L}{\partial\theta} = \frac{2n}{\theta} + \sum_{i=1}^{n} \log\left(1 - \left(1 - x_i{}^{-\beta}\right)^2\right) - (\theta+1) \times \sum_{i=1}^{n} \frac{\left(1-x_i{}^{-\beta}\right)^2}{\left(1-(1-\theta)\left(1-x_i{}^{-\beta}\right)^2\right)} - \sum_{i=1}^{n} \log\left(1 - (1-\theta)\left(1 - x_i{}^{-\beta}\right)^2\right), \tag{11}$$

$$\frac{\partial \, log \, L}{\partial\beta} = \frac{n}{\beta} - \sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} \frac{x_i{}^{-\beta} \ln x_i}{1-x_i{}^{-\beta}} - 2(\theta - 1) \times \sum_{i=0}^{n} \frac{x_i^{-\beta} ln(x_i)(1-x_i^{-\beta})}{1-\left(1-x_i^{-\beta}\right)^2} + 2(1-\theta^2) \sum_{i=0}^{n} \frac{x_i^{-\beta} ln(x_i)(1-x_i^{-\beta})}{\left(1-(1-\theta)\left(1-x_i{}^{-\beta}\right)^2\right)}. \tag{12}$$

There isn't a closed form to solve this equations for (θ, β). As a result, the equations can be solved numerically using the Newton-Raphson method and Mathematica program V. 11.0 to determine the Maximum Likelihood Estimate $\hat{\theta}_{MLE}$ and $\hat{\beta}_{MLE}$.

3 Simulation Study

In this section, the simulation result for the ML method is given when two parameters are unknown based on complete samples for various sample sizes and proposed initial values for parameters. The parameter values are selected as β = 2.8, θ=1.5 and n = 25, 100, 250, 450, and 1000. This process is repeated N = 500 times.

Furthermore, performance of different estimators is considered in terms of their biases and mean square errors (MSEs) that given, respectively, by

$$Bias(\hat{\lambda}) = E(\hat{\lambda}) - \lambda \quad \text{and} \quad \text{MSE}(\hat{\lambda}) = \text{E}(\hat{\lambda} - \lambda)^2.$$

Where λ any parameter.

Table 1. Mean, MSEs and Bias for the parameter estimates when β_0 = 2.8, θ_0=1.5.

N		MLE	MSE	BIAS
25	$\hat{\beta}$	3.37488	0.76493	0.574879
	$\hat{\theta}$	1.67533	0.627229	0.17533
100	$\hat{\beta}$	3.00401	0.638085	0.204005
	$\hat{\theta}$	1.67228	0.622801	0.17228
250	$\hat{\beta}$	3.00453	0.27286	0.204525
	$\hat{\theta}$	1.579	0.185657	0.0790023
450	$\hat{\beta}$	2.94519	0.339871	0.145187
	$\hat{\theta}$	1.55915	0.151468	0.059146
1000	$\hat{\beta}$	2.89381	0.069541	0.0938141
	$\hat{\theta}$	1.53192	0.05152	0.0319226

From Table 1 MSE and Bias are displayed. It can be illustrated clearly that these estimates are reasonably consistent and approaches to the true values of parameters as sample size increases. Furthermore, with increasing sample size the MSEs and Bias decrease for all parameter combinations. Therefore, it has been concluded that MLE process performs well in estimating the parameters of NE-P distribution.

4 Real Data Applications

In this section, we provide the application with real data sets to assess the flexibility of NE-P distribution comparing with the base line Pareto distribution. The parameters are estimated using maximum likelihood method. Mathematica (V.11.0) is used for computation. Moreover, we consider the model selection criteria, including Akaike information criterion (AIC), Bayesian information criterion (BIC), consistent Akaike information criterion (CAIC), and Hannan-Quinn information criterion (HQIC). They are defined as follows:

$$AIC = -2l(\hat{\theta}) + 2k$$
$$CAIC = AIC + \frac{2k(k+1)}{n-k-1}$$
$$HQIC = -2logl(\hat{\theta}) + 2k\log(\log(n))$$

(See Whittaker and Furlow (Whittaker and Furlow 2009).

5 Daily Mortality Cases of COVID-19

In this section, we will study the data number of daily mortality COVID-19 cases will be compared with (PD).

5.1 Data set 1

The first data represents a COVID-19 mortality rates data belongs to Italy of 59 days, that is recorded from 27 February to 27 April 2020. The data is taken from (Almongy et al. 2021) as follows:

4.571, 7.201, 3.606, 8.479, 11.410 ,8.961, 10.919, 10.908, 6.503 ,18.474, 11.010 ,17.337, 16.561, 13.226, 15.137 ,8.697 ,15.787 ,13.333 ,11.822 ,14.242 ,11.273, 14.330, 16.046, 11.950, 10.282, 11.775 ,10.138 ,9.037 ,12.396 ,10.644, 8.646 ,8.905, 8.906, 7.407, 7.445, 7.214, 6.194, 4.640 ,5.452 ,5.073, 4.416, 4.859 ,4.408 ,4.639 ,3.148 ,4.040 ,4.253 ,4.011 ,3.564, 3.827, 3.134, 2.780, 2.881, 3.341, 2.686, 2.814, 2.508, 2.450 ,1.518.

Table 2. Descriptive statistics for data set 1

n	Min	Q_1	Median	Mean	Q_3	Max	Skewness	Kurtosis
59	1.518	4.04	7.44	8.15	11.41	18.474	0.45	2.12

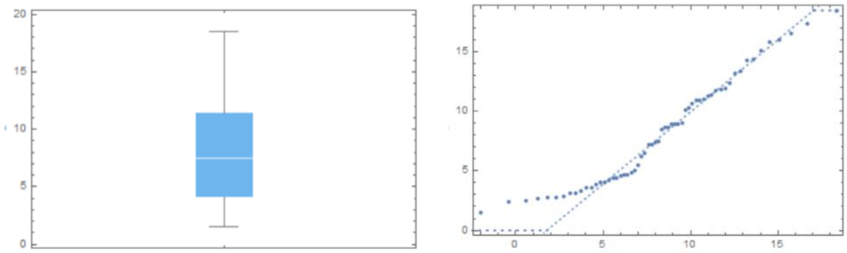

Fig. 1. PP plot of the NE-P distribution and the box plot for data set 1

Table 3. Parameter estimation for various distributions depending on data set 1

Model	Parameters		LL	AIC	CAIC	HQIC
	$\hat{\theta}$	$\hat{\beta}$				
NE-P	0.31	2.31	-190.47	384.94	385.15	386.56
PD		0.52	-211.25	424.50	424.57	425.31

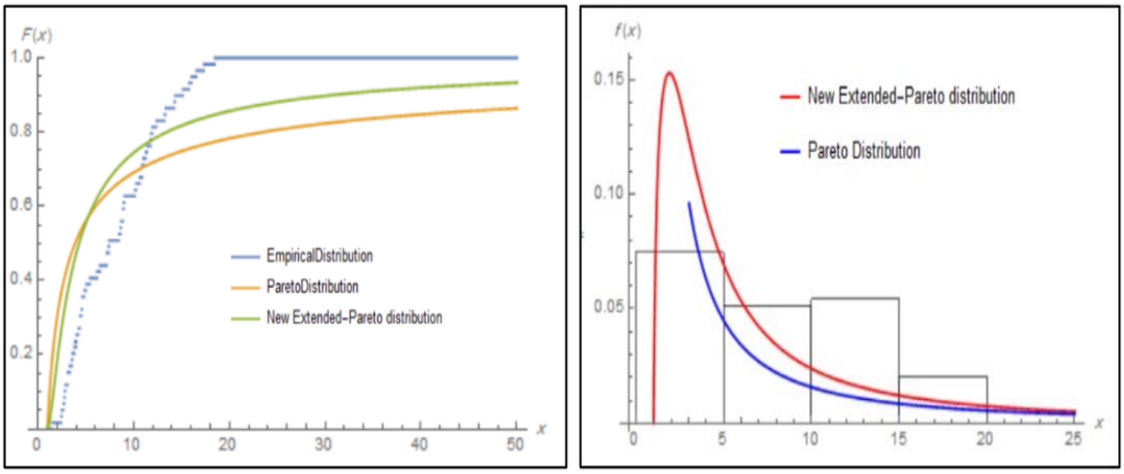

Fig. 2. Plots of the fitted CDF (left) and the histogram with fitted PDF (right) of the NE-P model for data set 1

5.2 Data set 2

The second data set of data of COVID-19 mortality numbers in Mexico of 108 days, that is recorded from 4 March to 20 July 2020. The data is taken from Nagy et al. The data is taken from (Almongy et al. 2021) as follows:

8.826, 6.105, 10.383, 7.267, 13.220, 6.015, 10.855, 6.122, 10.685, 10.035, 5.242, 7.630, 14.604, 7.903, 6.327, 9.391, 14.962, 4.730, 3.215, 16.498, 11.665, 9.284, 12.878, 6.656, 3.440, 5.854, 8.813, 10.043, 7.260, 5.985, 4.424, 4.344, 5.143, 9.935, 7.840, 9.550, 6.968, 6.370, 3.537, 3.286, 10.158, 8.108, 6.697, 7.151, 6.560, 2.988, 3.336, 6.814, 8.325, 7.854, 8.551, 3.228, 3.499, 3.751, 7.486, 6.625, 6.140, 4.909, 4.661, 1.867, 2.838, 5.392, 12.042, 8.696, 6.412, 3.395, 1.815, 3.327, 5.406, 6.182, 4.949, 4.089, 3.359, 2.070, 3.298, 5.317, 5.442, 4.557, 4.292, 2.500, 6.535, 4.648, 4.697, 5.459, 4.120, 3.922, 3.219, 1.402, 2.438, 3.257, 3.632, 3.233, 3.027, 2.352, 1.205, 2.077, 3.778, 3.218, 2.926, 2.601, 2.065, 1.041, 1.800, 3.029, 2.058, 2.326, 2.506, 1.923.

Table 4. Descriptive statistics for data set 2

n	Min	Q_1	Median	Mean	Q_3	Max	Skewness	Kurtosis
108	1.041	3.23	5.19	5.75	7.48	16.498	0.98	3.68

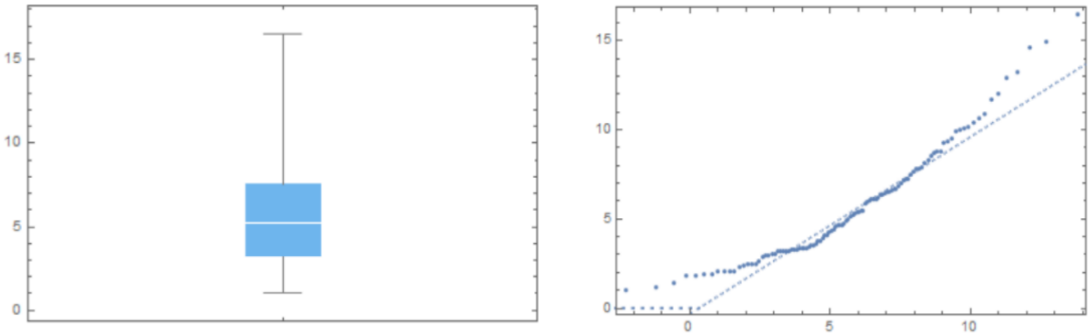

Fig. 3. PP plot of the NE-P distribution and the box plot for data set 2

Table 5. Parameter estimation for various distributions depending on data set 2

Model	Parameters					
	$\hat{\theta}$	$\hat{\beta}$	LL	AIC	CAIC	HQIC
NE-P	0.33	2.65	-296.84	597.69	597.81	599.87
PD		0.63	-329.82	661.64	661.68	662.73

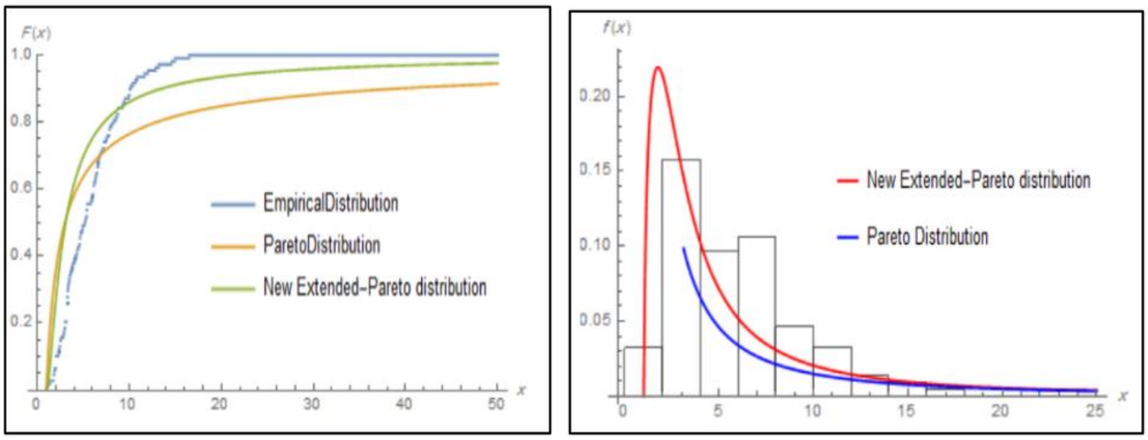

Fig. 4. Plots of the fitted CDF (left) and the histogram with fitted PDF (right) of the NE-P model for data set 2

5.3 Data set 3

The third data set of a COVID-19 data belonging to the Netherlands of 30 days, that is recorded from 31 March to 30 April 2020. This data formed of rough mortality rate. (see Almongy et al. 2021) The data are as follows:

14.918, 10.656, 12.274 ,10.289, 10.832, 7.099, 5.928, 13.211, 7.968, 7.584, 5.555, 6.027, 4.097, 3.611, 4.960, 7.498, 6.940, 5.307, 5.048, 2.857, 2.254, 5.431, 4.462, 3.883, 3.461, 3.647, 1.974, 1.273, 1.416, 4.235.

Table 6. Descriptive statistics for data set 3

n	Min	Q_1	Median	Mean	Q_3	Max	Skewness	Kurtosis
30	1.273	3.64	5.36	6.15	7.58	14.918	0.83	2.95

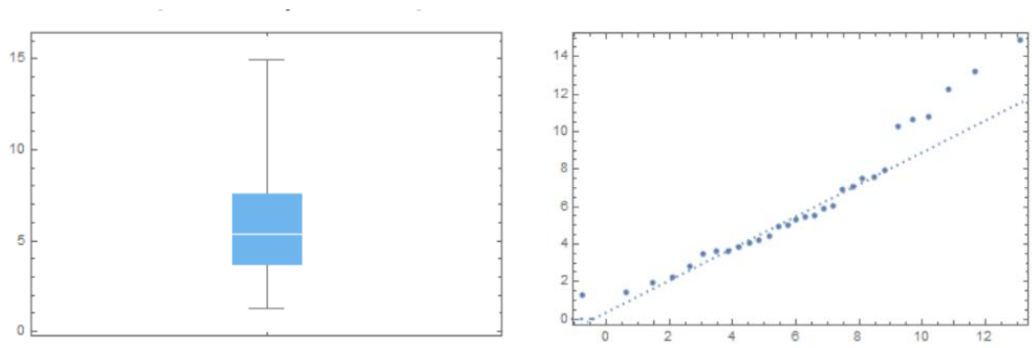

Fig. 5. PP plot of the NE-P distribution and the box plot for data set 3

Table 7. Parameter estimation for various distributions depending on data set 3

Model	**Parameters**					
	$\hat{\theta}$	$\hat{\beta}$	**LL**	**AIC**	**CAIC**	**HQIC**
NE-P	0.37	2.32	-85.46	174.93	175.38	175.83
PD		0.60	-94.38	190.76	190.91	191.21

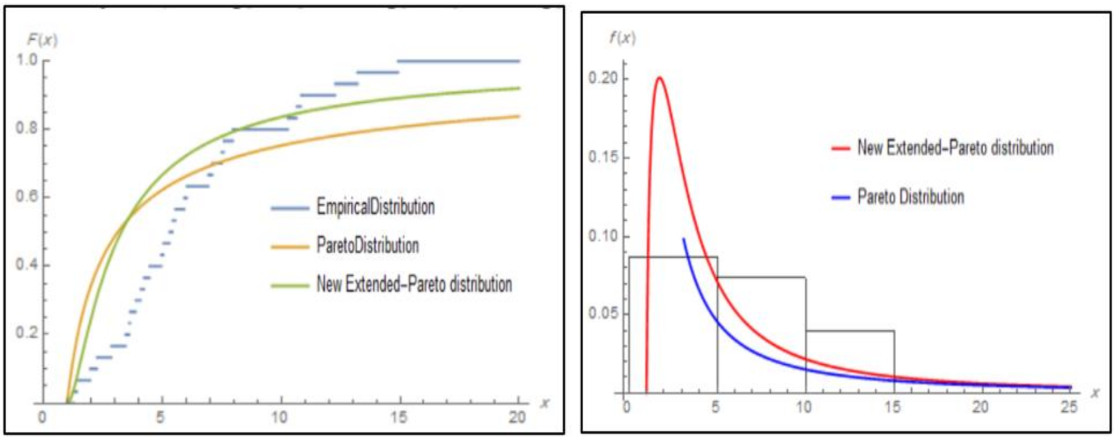

Fig. 6. Plots of the fitted CDF (left) and the histogram with fitted PDF (right) of the NE-P model for data set 3

5.4 Data set 4

The fourth data set represents a COVID-19 data belong to Canada of 36 days, from 10 April to 15 May 2020 (see Almetwally et al. 2021. These data formed of mortality rate. The data are as follows:

3.1091, 3.3825, 3.1444, 3.2135, 2.4946, 3.5146, 4.9274, 3.3769, 6.8686, 3.0914, 4.9378, 3.1091, 3.2823, 3.8594, 4.0480, 4.1685, 3.6426, 3.2110, 2.8636, 3.2218, 2.9078, 3.6346, 2.7957, 4.2781, 4.2202, 1.5157, 2.6029, 3.3592, 2.8349 ,3.1348, 2.5261, 1.5806, 2.7704, 2.1901, 2.4141, 1.9048.

From Tables 3, 5, 7 and 9, the values of log-likelihood (LL), AIC, CAIC and HQIC are minimum and favorable of NE-P distribution compared with PD distribution, which indicate that our new model is the best comparing with the competing model.

Table 8. Descriptive statistics for data set 4

n	**Min**	Q_1	**Median**	**Mean**	Q_3	**Max**	**Skewness**	**Kurtosis**
36	1.5171	2.77	3.17	3.28	3.63	6.8686	1.21	6.15

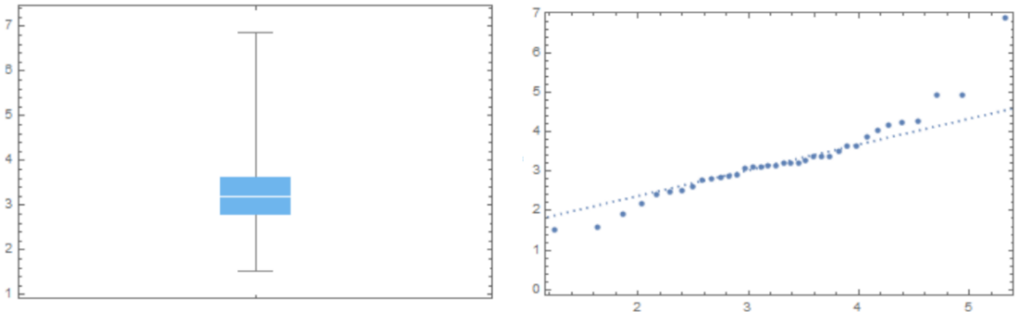

Fig. 7. PP plot of the NE-P distribution and the box plot for data set 4

Table 9. Parameter estimation for various distributions depending on data set 4

Model	**Parameters**					
	$\hat{\theta}$	$\hat{\beta}$	**LL**	**AIC**	**CAIC**	**HQIC**
NE-P	0.32	3.75	-68.20	140.41	140.77	141.51
PD		0.87	-82.13	166.27	166.39	166.82

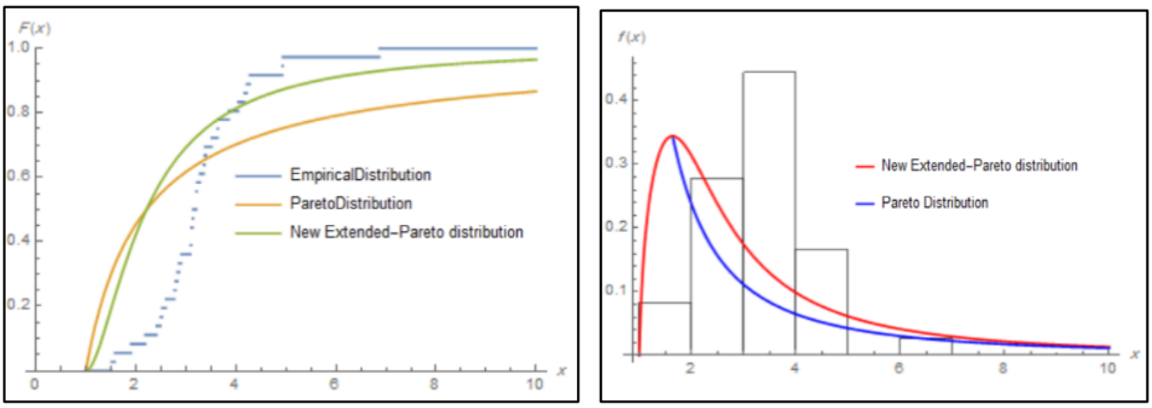

Fig. 8. Plots of the fitted CDF (left) and the histogram with fitted PDF (right) of the NE-P model for data set 4

6 Conclusion

COVID-19 data modeling has gained renewed interest among researchers, particularly in the quest to find new models that are more flexible in modeling this data. In this article We found a New Extended -Pareto distribution as a model for these data. Its parameters were estimated by method of maximum likelihood. Performances of MLE were tested through simulation study. Finally, four real data applications of COVID-19 were analyzed in to assess the flexibility of our new model. We encourage researchers to continue exploring new models for modeling this kind of data sets. Future studies can expand the study of NE-P and apply other types of data as well as compare it to other competing distributions.

Disclaimer (Artificial Intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of this manuscript.

Competing Interests

Authors have declared that no competing interests exist.

References

Aldahlan, M. A., Rabie, A. M., Abdelhamid, M., Ahmed, A. H. N., & Afify, A. Z. (2023). The Marshall–Olkin Pareto Type-I Distribution: Properties, Inference under Complete and Censored Samples with Application to Breast Cancer Data. *Pakistan Journal of Statistics and Operation Research*, 603-622.

Almetwally, E. M., Alharbi, R., Alnagar, D., & Hafez, E. H. (2021). A new inverted topp-leone distribution: applications to the COVID-19 mortality rate in two different countries. *Axioms*, 10(1), 25.

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

Bantan, R. A., Jamal, F., Chesneau, C., & Elgarhy, M. (2020). Type II Power Topp-Leone generated family of distributions with statistical inference and applications. *Symmetry*, 12(1), 75.

Boumaraf, B., Seddik-Ameur, N. and Barbu, V. S. (2020). Estimation of Beta-Pareto distribution based on several optimization methods. *Mathematics*, *8*(7), 1-12.

Caeiro, F., & Mateus, A. (2023). A new class of generalized probability-weighted moment estimators for the Pareto distribution. *Mathematics*, 11(5), 1076.

Jayakumar, K., Krishnan, B. and Hamedani, G. G. (2018). On new generalization of Pareto distribution and its applications. *Communications in Statistics-Simulation and Computation*, 49(5), 1264-1284.

Sule, I., Sani, I. D., Audu, I., & Jibril, H. M. (2020). On the Topp Leone exponentiated-G family of distributions: properties and applications.

Whittaker, T. A., & Furlow, C. F. (2009). The comparison of model selection criteria when selecting among competing hierarchical linear models. *Journal of Modern Applied Statistical Methods*, *8*(1), 15.

Zichuan, M., Hussain, S., Iftikhar, A., Ilyas, M., Ahmad, Z., Khan, D. M. and Manzoor, S. (2020). A new extended-family of distributions: properties and applications. *Computational and mathematical methods in medicine*, 2020, 1-13.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
https://www.sdiarticle5.com/review-history/127675