DOI https://doi.org/10.5281/zenodo.10049652#58

TWIST, 2023, Vol. 18, Issue 4, pp. 328-337

TWIST

Journal homepage: www.twistjournal.net

The Exponentiated Power Chris-Jerry Distribution:
Properties, Regression, Simulation and Applications to Infant Mortality Rate and Lifetime of COVID-19 Patients

Okechukwu J. Obulezi*
Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria
[*Corresponding author]

Obiorah Emmanuella Ujunwa
Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria

Ifeanyi C. Anabike
Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria

Chinyere P. Igbokwe
Department of Statistics, Abia State Polytechnic, Aba, Nigeria

Abstract
The new three-parameter exponentiated power Chris-Jerry distribution is introduced, and some of its mathematical properties are addressed. Its parameters are estimated by maximum likelihood. A regression model called the log-exponentiated power Chris-Jerry distribution regression model is constructed based on the logarithmic transformation of the proposed distribution. We derived the basic properties of the distribution and showed the flexibility of the proposed model using the plots of the hazard rate function. The new regression model is deployed to fit COVID-19 censored data with the age of patients and diabetic index as the regressors. The usefulness of the proposed model is proved using the infant mortality rate for some selected countries in 2021.

Keywords
COVID-19 data, Exponentiated-G class, Exponentiated Power Chris-Jerry distribution, Infant Mortality Rate, Model Adequacy, Regression model

INTRODUCTION

Survival analysis and reliability engineering are gray areas of application of many emerging distributions due to the central role they play in society. To innovate some existing distributions, various methods exist, and important to this article is the exponentiated-G method. This method introduces an additional shape parameter to a baseline distribution which subsequently creates varying shapes and improves the applicability of the existing distribution. Examples of articles of this class are exponentiated power Ishita by Ferreira and Cordeiro (2023), exponentiated Ishita by Rather and Subramanian (2019), exponentiated power Lindley by Ashour and Eltehiwy (2015), and exponentiated Adya by Ganaie et al. (2023). Other related studies are Ezeilo et al. (2023), Gomaa et al. (2023), Bhat et al. (2023), Baaqeel et al. (2023), Ramadan, A. H. Tolba, and El-Desouky (2022), Musa, Onyeagu, and O. J. Obulezi (2023a), Musa, Onyeagu, and O. J. Obulezi (2023b), A. Tolba et al. (2022), A. H. Tolba, Muse, et al. (2023), A. H. Tolba, Onyekwere, et al. (2023), and Nwankwo et al. (2023).

Exponentiation has dominated the method of extending members of the Lindley class of distributions which include Etaga et al. (2023), Innocent et al. (2023), Anabike et al. (2023), Lindley (1958), O. J. Obulezi, Anabike, Okoye, et al. (2023), Oramulu et al. (2023), O. J. Obulezi, Anabike, Oyo, et al. (2023), O. Obulezi, Igbokwe, and Anabike (2023), O. J. Obulezi, Chidimma, et al. (2023), and Onyekwere and O. J. Obulezi (2022). The resulting distributions lend themselves to a reparametrized regression model which is the main idea conveyed in this article. The log transformation of a distribution to a reparametrized regression model is a development that is very interesting in mathematical statistics because it is an integration of two areas of statistics to create a wider application, see Ferreira and Cordeiro 2023 for a detailed perspective.

For the COVID-19 data with two features namely the age of patients and diabetic history, the proposed will make a good fit and add to existing literature. Essentially, comparing it will members of the same Lindley class of distributions will demonstrate its usefulness.

To this end, the motivation for this article is to innovate a new parametric regression model that will be able to fit some skewed censored data and the rest of the article is in the following arrangement; in section 2, the new model is formulated. In section 3, some of the properties are presented. In section 4, the estimation of the uncensored data procedure is presented. In section 5, the log-transformed regression model equivalent of the proposed distribution is derived together with the estimation. In section 6, an application to the life cycle of COVID-19 patients with a history of diabetic Mellitus with their age disparity is done. In section 7, the second application on the infant mortality rate of some countries in 2021 is also done. The paper is concluded in section 8.

THE PROPOSED MODEL

Ezeilo et al. 2023 introduced the Power Chris-Jerry distribution with probability density function (p.d.f) and cumulative distribution function (c.d.f) given respectively as

$$g(x;\alpha,\theta)=\frac{\alpha\theta^2}{\theta+2}(1+\ \theta x^{2\alpha})x^{\alpha-1}e^{-\theta x^{\alpha}} \tag{1}$$

and

$$G(x;\alpha,\theta)=1-\left(1+\frac{\theta x^{\alpha}\left(\theta x^{\alpha}+2\right)}{\theta+2}\right)e^{-\theta x^{\alpha}} \tag{2}$$

The exponentiated-G family of distributions was developed by Lehmann 2012 and Durrans 1992. A random variable *X* is said to follow Exponentiated-G distribution if its c.d.f and p.d.f are respectively

$$F(x;c,\xi)=G(x;\xi)^c \tag{3}$$

and

$$f(x;c,\xi)=cg(x;\xi)G(x;\xi)^{c-1} \tag{4}$$

where ξ is the parameter vector of $G(.)$. By substituting eq 1 into eq 3 and eq 2 into eq 4 the cdf and pdf of the $X\sim$ EPCJ (c,θ,α) are obtained as follows;

$$F(x;c,\theta,\alpha)=\left\{1-\left[1+\frac{\theta x^{\alpha}\left(\theta x^{\alpha}+2\right)}{\theta+2}\right]e^{-\theta x^{\alpha}}\right\}^{c} \tag{5}$$

and

$$f(x;c,\theta,\alpha)=\frac{c\alpha\theta^2}{\theta+2}(1+\theta x^{2\alpha})x^{\alpha-1}e^{-\theta x^{\alpha}}\left\{1-\left[1+\frac{\theta x^{\alpha}\left(\theta x^{\alpha}+2\right)}{\theta+2}\right]e^{-\theta x^{\alpha}}\right\}^{c-1} \tag{6}$$

The hazard function is given as

$$h(x;c,\theta,\alpha)=\frac{\frac{c\alpha\theta^2}{\theta+2}(1+\theta x^{2\alpha})x^{\alpha-1}e^{-\theta x^{\alpha}}\left\{1-\left[1+\frac{\theta x^{\alpha}(\theta x^{\alpha}+2)}{\theta+2}\right]e^{-\theta x^{\alpha}}\right\}^{c-1}}{1-\left\{1-\left[1+\frac{\theta x^{\alpha}(\theta x^{\alpha}+2)}{\theta+2}\right]e^{-\theta x^{\alpha}}\right\}^{c}} \tag{7}$$

Definition 2.1 (Linear Representation of EPCJ distribution). Using Binomial theorem on $\left\{1-\left[1+\frac{\theta x^{\alpha}(\theta x^{\alpha}+2)}{\theta+2}\right]e^{-\theta x^{\alpha}}\right\}^{c-1}$ yields

$$\left\{1-\left[1+\frac{\theta x^{\alpha}\left(\theta x^{\alpha}+2\right)}{\theta+2}\right]e^{-\theta x^{\alpha}}\right\}^{c-1}=\sum_{i=0}^{\infty}(-1)^i\binom{c-1}{i}\left[1+\frac{\theta x^{\alpha}\left(\theta x^{\alpha}+2\right)}{\theta+2}\right]^{i}e^{-i\theta x^{\alpha}} \tag{8}$$

Again, using binomial on $\left[1+\frac{\theta x^{\alpha}(\theta x^{\alpha}+2)}{\theta+2}\right]^{i}$, one obtains

$$\left[1+\frac{\theta x^{\alpha}\left(\theta x^{\alpha}+2\right)}{\theta+2}\right]^{i}=\sum_{j=0}^{\infty}\binom{i}{j}\frac{\theta^j x^{j\alpha}}{(\theta+2)^j}(\theta x^{\alpha}+2)^j \tag{9}$$

the linear form becomes

$$f(x;c,\alpha,\theta)=\frac{c\alpha\theta^2}{\theta+2}(1+\theta x^{2\alpha})x^{\alpha-1}e^{-\theta x^\alpha}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^i\binom{c-1}{i}\binom{i}{j}\frac{\theta^j x^{j\alpha}}{(\theta+2)^j}(\theta x^\alpha+2)^j e^{-i\theta x^\alpha} \tag{10}$$

Also

$$(\theta+2)^{-(j+1)}=\sum_{k=0}^{\infty}(-1)^k\binom{j+k}{k}\theta^k 2^{-j-k-1};\quad \text{and}\quad (\theta x^\alpha+2)^j=\sum_{l=0}^{j}\binom{j}{l}2^{j-l}\theta^l x^{\alpha l} \tag{11}$$

Therefore; a concise linear representation of the p.d.f of the EPCJ distribution is

$$f(x;c,\alpha,\theta)=c\alpha\sum_{i,j,k=0}^{\infty}\sum_{l=0}^{j}(-1)^{i+k}\binom{c-1}{i}\binom{i}{j}\binom{j+k}{k}\binom{j}{l}\theta^{j+k+l+2}2^{-l-k-1}(1+\theta x^{2\alpha})x^{\alpha(j+l+1)-1}e^{-(i+1)\theta x^\alpha} \tag{12}$$

Fig 1 and fig 2 are the plots of the p.d.f while fig 3 and fig 4 are the plots of the c.d.f. From the plots of the hazard function, the proposed EPCJ distribution has both L-shape, decreasing, and increasing shapes suggesting wider applications.

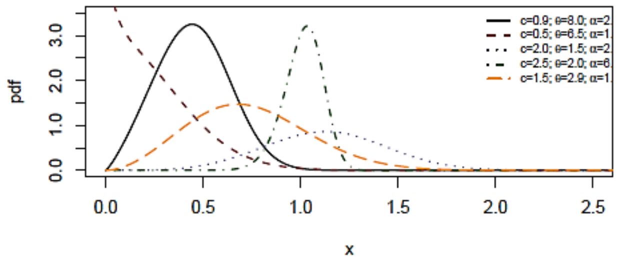

Figure 1: pdf of $X \sim$ EPCJ (c,θ,α)

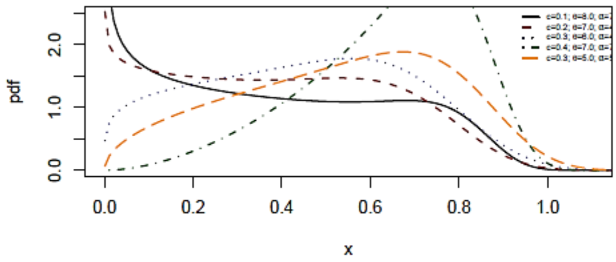

Figure 2: pdf of $X \sim$ EPCJ (c,θ,α)

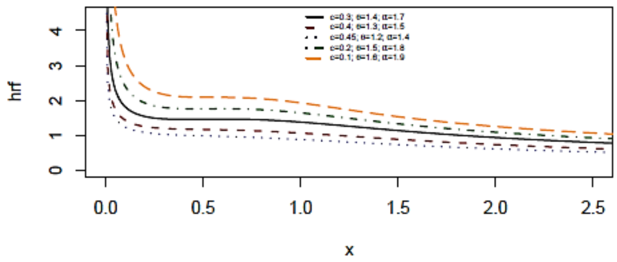

Figure 3: pdf of $X \sim$ EPCJ (c,θ,α)

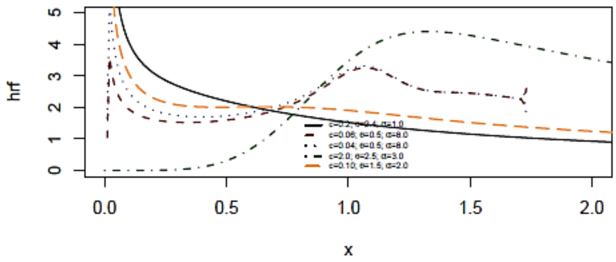

Figure 4: pdf of $X \sim$ EPCJ(c,θ,α)

PROPERTIES

Definition 3.1 Let $X \sim$ EPCJ $(c,\ \theta,\ \alpha)$, then the r^{th} crude moment is given as follows;

$$\begin{aligned}\mu_r' &= \int_0^\infty x^r f(x)dx\\ &= c\alpha\sum_{i,j,k=0}^{\infty}\sum_{l=0}^{j}(-1)^{i+k}\binom{c-1}{i}\binom{i}{j}\binom{j+k}{k}\binom{j}{l}\theta^{j+k+l+2}2^{-l-k-1}\int_0^\infty(1+\theta x^{2\alpha})x^{\alpha(j+l+1)+r-1}e^{-(i+1)\theta x^\alpha}dx\end{aligned} \tag{13}$$

define $\phi=c\alpha\sum_{i,j,k=0}^{\infty}\sum_{l=0}^{j}(-1)^{i+k}\binom{c-1}{i}\binom{i}{j}\binom{j+k}{k}\binom{j}{l}\theta^{j+k+l+2}2^{-l-k-1}$, then

$$\begin{aligned}\mu_r' &= \phi\left\{\int_0^\infty x^{\alpha(j+l-1)+r-1}e^{-(i+1)\theta x^\alpha}dx+\theta\int_0^\infty x^{\alpha(j+l+3)+r-1}e^{-(i+1)\theta x^\alpha}dx\right\}\\ &= c\sum_{i,j,k=0}^{\infty}\sum_{l=0}^{j}(-1)^{i+k}\binom{c-1}{i}\binom{i}{j}\binom{j+k}{k}\binom{j}{l}\theta^{j+k+l+2}2^{-l-k-1}\left\{\frac{\Gamma\left[\frac{r}{\alpha}+j+l+1-1\right]}{(i+1)\,\theta^{\frac{r}{\alpha}+j+l+1}}+\theta\frac{\Gamma\left[\frac{r}{\alpha}+j+l+3-1\right]}{(i+1)\,\theta^{\frac{r}{\alpha}+j+l+3}}\right\}\end{aligned} \tag{14}$$

The mean, 2^{nd}, 3^{rd} and 4^{th} crude moments are obtained by replacing r with 1, 2, 3 and 4 in μ'_r.

Definition 3.2 (Moment generating function). Let $X \sim$ EPCJ (c, θ, α), then the moment generating function $M_X(t)$ can be expressed as

MAXIMUM LIKELIHOOD ESTIMATION UNDER COMPLETE DATA

Let $X \sim$ EPCJ (c, θ, α), then the maximum log-likelihood is given as follows;

$$L(\xi) = \left(\frac{c\alpha\theta^2}{\theta+2}\right)^n e^{-\theta \sum_{i=1}^{n} x^\alpha} \prod_{i=1}^{n}(1+\theta x^{2\alpha})x^{\alpha-1}\left\{1-\left[1+\frac{\theta x^\alpha\left(\theta x^\alpha+2\right)}{\theta+2}\right]e^{-\theta x^\alpha}\right\}^{c-1} \tag{15}$$

The log-likelihood yields,

$$\begin{aligned}\ell = \log L(\xi) &= n\left[\log(c)+\log(\alpha)+2\log(\theta)-\log(\theta+2)\right]-\theta\sum_{i=1}^{n}x^\alpha+\sum_{i=1}^{n}\log\left(1+\theta x_i^{2\alpha}\right)\\ &+(\alpha-1)\sum_{i=1}^{n} log(x_i)+(c-1)\sum_{i=1}^{n}\log\left\{1-\left[1+\frac{\theta x_i^\alpha\left(\theta x_i^\alpha+2\right)}{\theta+2}\right]e^{-\theta x_i^\alpha}\right\}\end{aligned} \tag{16}$$

Taking the partial derivative of ℓ with respect to c, α and θ, we obtain

$$\frac{\partial \ell}{\partial c} = \frac{n}{c}+\sum_{i=1}^{n}\log\left\{1-\left[1+\frac{\theta x_i^\alpha\left(\theta x_i^\alpha+2\right)}{\theta+2}\right]e^{-\theta x_i^\alpha}\right\} \tag{17}$$

Therefore,

$$\hat{c} = -\frac{n}{\sum_{i=1}^{n}\log\left\{1-\left[1+\frac{\theta x_i^\alpha(\theta x_i^\alpha+2)}{\theta+2}\right]e^{-\theta x_i^\alpha}\right\}} \tag{18}$$

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha}-\theta\sum_{i=1}^{n}x^\alpha\log(x_i)+2\theta\sum_{i=1}^{n}\frac{x_i^{2\alpha}\log(x_i)}{1-\theta x_i^{2\alpha}}-\sum_{i=1}^{n}\log(x_i)+(c-1)\sum_{i=1}^{n}\frac{\left\{\theta+4\theta x_i^\alpha+\theta^2 x_i^{2\alpha}\right\}\theta x_i^\alpha\log(x_i)e^{-\theta x_i^\alpha}}{\left\{1-\left[1+\frac{\theta x_i^\alpha(\theta x_i^\alpha+2)}{\theta+2}\right]e^{-\theta x_i^\alpha}\right\}(\theta+2)} \tag{19}$$

and

$$\frac{\partial \ell}{\partial \theta} = \frac{2n}{\theta}-\frac{1}{\theta+2}-\sum_{i=1}^{n}x^\alpha+\sum_{i=1}^{n}\frac{x_i^{2\alpha}}{1+\theta x_i^{2\alpha}}+(c-1)\sum_{i=1}^{n}\frac{\left\{\theta^2-\theta^3 x_i^{2\alpha}-2\theta^2 x_i^{2\alpha}+\theta^2 x_i^\alpha+4\theta\right\}x_i^\alpha e^{-\theta x_i^\alpha}}{\left\{1-\left[1+\frac{\theta x_i^\alpha(\theta x_i^\alpha+2)}{\theta+2}\right]e^{-\theta x_i^\alpha}\right\}(\theta+2)^2} \tag{20}$$

LOG-EXPONENTIATED POWER CHRIS-JERRY DISTRIBUTION REGRESSION MODEL

Let $Y = \log(X)$ where $X \sim$ EPCJ (c, α, θ) defined in eq 6. Define $\alpha = \frac{1}{\sigma}$ and $\theta = e^{-\frac{\mu}{\sigma}}$, the log-Exponentiated Power Chris-Jerry (LEPCJ) density for $y \in \mathbb{R}$ is

$$f(y;c,\sigma,\mu) = \frac{ce^{\left(\frac{y-2\mu}{\sigma}\right)}\left(1+e^{\frac{2y-\mu}{\sigma}}\right)e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}}}{\sigma\left(e^{-\frac{\mu}{\sigma}}+2\right)}\left\{1-\left[1+\frac{e^{\left(\frac{y-\mu}{\sigma}\right)}\left(e^{\left(\frac{y-\mu}{\sigma}\right)}+2\right)}{e^{-\frac{\mu}{\sigma}}+2}\right]e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}}\right\}^{c-1} \tag{21}$$

where $c, \sigma > 0$ and $\mu \in \mathbb{R}$. If $X \sim$ EPCJ (c, θ, α), then $Y = \log(X) \sim$ LEPCJ (c, σ, μ). The survival and density function of $Z = \frac{Y-\mu}{\sigma}$ are

$$s(z;c,\sigma,\mu) = 1-\left\{1-\left[1+\frac{e^z(e^z+2)}{e^{-\frac{\mu}{\sigma}}+2}\right]e^{-e^z}\right\}^c \tag{22}$$

and

$$f(z;c,\sigma,\mu) = \frac{cw(z)e^{-e^z}}{\sigma\left(e^{-\frac{\mu}{\sigma}}+2\right)}\left\{1-\left[1+\frac{e^z(e^z+2)}{e^{-\frac{\mu}{\sigma}}+2}\right]e^{-e^z}\right\}^{c-1}; \quad z \in \mathbb{R} \tag{23}$$

respectively, where $w(z) = e^{\frac{\sigma z-\mu}{\sigma}}\left(1+e^{\frac{2\sigma z+\mu}{\sigma}}\right)$. Using eq 23, we construct a parametric regression model for the response variable Y_i and a vector of explanatory variables $\mathbf{v_i'} = (v_{i1}, v_{i2}, ..., v_{ip})$ as

$$y_i = \mathbf{v}'\beta + \sigma z_i, \quad i = 1, 2, ..., n \tag{24}$$

where $\mu i = v'\beta$, $\beta = (\beta_1, ..., \beta_p)'$ is the vector of unknown regression coefficients and z is the random error with density in equation 12. Define the survival and density functions of $Y_i|v'$ are

$$S(y|\mathbf{v}') = 1 - \left\{1 - \left[1 + \frac{e^{z_i}(e^{z_i}+2)}{e^{-\frac{\mu_i}{\sigma}}+2}\right] e^{-e^{z_i}}\right\}^c \tag{25}$$

and

$$f(y|\mathbf{v}') = \frac{cw(z_i)e^{-e^{z_i}}}{\sigma\left(e^{-\frac{\mu_i}{\sigma}}+2\right)}\left\{1 - \left[1 + \frac{e^{z_i}(e^{z_i}+2)}{e^{-\frac{\mu_i}{\sigma}}+2}\right] e^{-e^{z_i}}\right\}^{c-1} \tag{26}$$

where $w(z_i) = e^{\frac{\sigma z_i-\mu_i}{\sigma}}\left(1+e^{\frac{2\sigma z_i+\mu_i}{\sigma}}\right)$ and $z_i = \frac{y_i-\mu_i}{\sigma}$

MAXIMUM LIKELIHOOD UNDER CENSORED DATA

To estimate the parameters in eq 24 for right-censored data, we define Y_i and C_i as the lifetime and noninformative censoring time (assuming independence) and $y_i = \min(Y_i, C_i)$. Then, the log-likelihood function for $\xi = (c, \sigma, \beta^T)^T$ is

$$\begin{aligned}\ell(\xi) = d\left[\log(c) - \log(\sigma)\right] - \sum_{i\in F}\log\left(e^{-\frac{-\mu 1}{\sigma}}+2\right) + \sum_{i\in F}\log\left[w(z_i)\right] - \sum_{i\in F} e^{z_i} \\ + (c-1)\sum_{i\in F}\log\left\{1 - \left[1 + \frac{e^{z_i}(e^{z_i}+2)}{e^{-\frac{\mu_i}{\sigma}}+2}\right]e^{-e^{z_i}}\right\} + \sum_{i\in C}\left\{1 - \left(1 - \left[1 + \frac{e^{z_i}(e^{z_i}+2)}{e^{-\frac{\mu_i}{\sigma}}+2}\right]e^{-e^{z_i}}\right)^c\right\}\end{aligned} \tag{27}$$

where F and C are the sets of uncensored and censored observations respectively and d is the number of failures. The MLE ξ of the unknown parameter vector can be obtained by maximizing eq 27.

SIMULATIONS

The simulation conditions deployed by Ferreira and Cordeiro 2023 are used in this article due to the compatibility of the two distributions. For the EPCJ distribution under different scenarios, the accuracy of the MLEs is examined. For 1000 repetitions, the acceptance and rejection method is adopted to generate random samples of sizes n = 50, 100, 300, and 600 from the EPCJ distribution. The Average estimates (AEs) of the parameters, Biases, and mean squared error (MSEs) are calculated. The algorithm for generating random samples uses the acceptance-rejection method, see ibid.

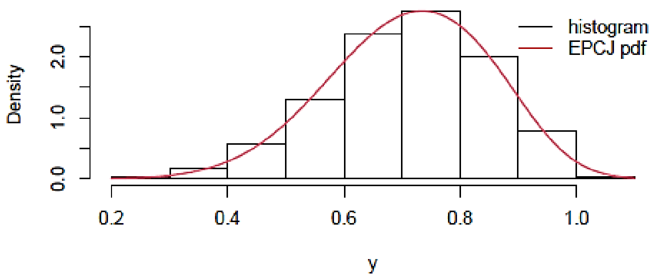

Figure 5: Histogram and estimated pdf for generated samples using the scenario (1.4, 5.0, 7.0)

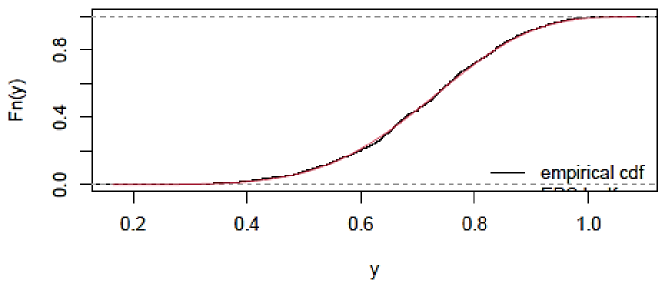

Figure 6: Empirical cdf and estimated cdf for generated samples using the scenario (1.4, 5.0, 7.0)

Figures 5 and 6 show the approximation of the acceptance-rejection method. The estimated pdf and cdf of EPCJ distribution are very close to the histogram and empirical cdf of the generated samples. This indicates a good performance of the method.

The statistics in Table 1 indicate that the AEs converge to the true parameters and that the biases and MSEs tend to zero when n increases, which proves the consistency of the EPCJ estimators. Note that for the (2.0, 0.5, 10) scenario the parameter estimates are less accurate (except for α), while for the (1.4, 5.0, 7.0) scenario, the parameter estimates are more accurate (for c and θ). Overall, the simulation results suggest that larger sample sizes and the appropriate choice of ξ are crucial for accurate parameter estimation of the EPCJ distribution. This is the same conclusion reached by ibid.

Table 1: Simulation measures from the EPCJ distribution

		(2.0, 0.5, 10)			(1.4, 5.0, 7.0)			(3.0, 3.5, 9.0)		
n	ξ	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
50	c	0.4108	-1.5892	2.5421	0.2573	-1.1427	1.3161	0.4346	-2.5654	6.5993
	θ	26.2835	0.9188	478.116	11.7584	4.7584	75.5373	19.8986	10.8986	231.7697
	α	1.4188	1.5892	0.9555	18.0815	13.0815	215.4634	11.5636	8.0636	73.2655
100	c	0.3877	-1.6123	2.6082	0.2564	-1.1436	1.3146	0.4048	-2.5952	6.7454
	θ	22.5304	12.5304	266.0897	9.6005	2.6005	32.1256	17.6725	8.6725	159.325
	α	1.4115	0.9115	0.9032	16.9036	11.9036	167.0537	11.7306	8.2306	74.1056
300	c	0.3968	-1.6032	2.5738	0.2759	-1.1241	1.2661	0.4139	-2.5861	6.6922
	θ	16.8067	6.8067	63.3039	6.9566	-0.0434	2.1214	12.6518	3.6518	24.0334
	α	1.3101	0.8109	0.6817	14.5183	9.5183	96.4223	10.7852	7.2852	55.0501
600	c	0.4019	-1.5981	2.5560	0.2846	-1.1154	1.2455	0.4215	-2.5785	6.6504
	θ	15.3365	5.3365	35.0927	6.4197	-0.5803	1.0891	11.3955	2.3955	9.1690
	α	1.2761	0.7761	0.6149	13.8106	8.8106	80.2566	10.4080	6.9080	48.6338

APPLICATION TO COVID-19 DATA

The dataset comprises the lifetime (in days) of 322 individuals diagnosed with COVID-19 through RT-PCR screening in Campinas, Brazil. These data were previously studied by ibid. The response variable yi represents the time elapsed from the onset of symptoms until death due to COVID-19 (failure). ibid. observed that about 66.45% of the observations are censored. The variables considered (*for* $i = 1, ..., 322$) include: δ_i : censoring indicator (0 = censored, 1 = observed lifetime), v_{i1} : age (in years), and v_{i2} : diabetes mellitus (1 = yes, 0 = no or not informed). The suggested regression model for these COVID-19 data is written as;

$$y_i = \beta_0 + \beta_1 v_{i1} + \beta_2 v_{i2} + \sigma z_i; \quad i = 1, ..., 322, \tag{28}$$

where $z_i \sim$ the pdf in eq 20.

The power Lomax (PLO) distribution by Rady, Hassanein, and Elhaddad 2016, exponentiated power Akash (EPA) distribution (new), power Prakaamy (PP) distribution by Shukla and Shanker 2020, Exponentiated Power Lindley (EPL) distribution by Ashour and Eltehiwy 2015 and power Rama (PR) distribution by Abebe et al. 2019 are used to compare with the proposed Exponentiated Power Chris-Jerry (EPCJ) distribution. Note, that the log- of each distribution is derived following the procedure in section 5 to obtain LPLO, LEPA, LPP, LEPL, and LPR respectively.

The result from table 2 shows that the explanatory variables age and diabetes mellitus are significant at the 5% significance level. The negative signs of β_1 and β_2 mean that older individuals or those with diabetes tend to have shorter failure times. This result is in agreement with that obtained from Ferreira and Cordeiro 2023 earlier study. From Table 3, the LEPCJ regression has the lowest criterion values hence confirming that the LEPCJ model provides a better fit for the COVID-19 data.

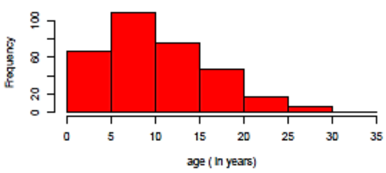

Figure 7: histogram for COVID-19 data

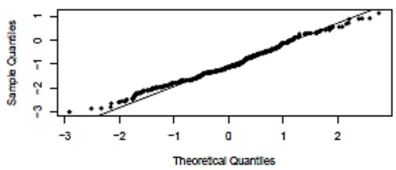

Figure 8: QQ plot for the COVID-19 data

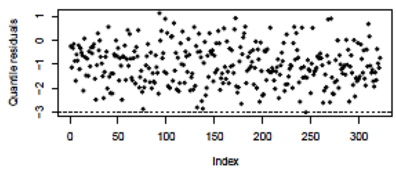

Figure 9: Quantile Residual plot for the COVID-19 data

Table 2: Estimates of the Regression parameters for the COVID-19 data

Distr	c	σ	β_0	β_1	β_2
LEPCJ	0.8065 (0.3917)	0.8666 (0.2435)	3.4401 (0.4757)	-0.0166 (0.0035) $[< 2.9039 \times 10^{-6}]$	-0.2681 (0.1110) $[< 0.01624]$
LPLO	5.3408 (4.0688)	0.5442 (0.0439)	-5.4568 (0.5656)	-0.0193 (0.0041) $[< 3.1327 \times 10^{-6}]$	-0.2963 (0.1232) $[< 0.0194]$
LEPA	0.5843 (0.5845)	0.8380 (0.5255)	3.7801 (0.9222)	-0.0187 (0.0039) $[< 3.0602 \times 10^{-6}]$	-0.2865 (0.1232) $[< 0.0206]$
LPP	1	1.3780 (0.0912)	3.6182 (0.3140)	-0.0285 (0.0042) $[< 4.1498 \times 10^{-11}]$	-0.4074 (0.1340) $[< 0.0026]$
LEPL	0.4863 (0.2957)	0.5429 (0.2367)	4.3019 (0.3900)	-0.0177 (0.0039) $[< 8.8659 \times 10^{-6}]$	-0.2739 (0.1170) $[< 0.0199]$
LPR	1	1.3356 (0.0866)	3.5766 (0.3052)	-0.0273 (0.0040) $[< 6.2066 \times 10^{-11}]$	-0.4021 (0.1308) $[< 0.0023]$

Table 3: Measures of Model Adequacy

Distr	AIC	CAIC	BIC	HQIC
LEPCJ	430.2712	430.6279	449.1439	437.8058
LPLO	431.8449	432.2016	450.7177	439.3795
LEPA	433.6238	433.9805	452.4999	441.1584
LPP	645.113	645.3796	660.2112	651.1407
LEPL	430.2784	430.6351	449.1512	437.8130
LPR	641.4597	641.7264	656.5579	647.4874

APPLICATION TO INFANT MORTALITY RATE DATA

The data on infant mortality rate per 1,000 live births for a few chosen nations in 2021, as reported by https://data.worldbank.org/indicator/SP.DYN.IMRT.IN. The data set is presented in Table 4 below;

Table 4

56	10	22	3	69	6	7	11	4	4	19	13	7	27
12	3	4	11	84	27	25	6	35	14	11	12	6	

The following model adequacy statistics are used; Akaike information criterion (BIC), Corrected Akaike Information criterion (CAIC), Bayesian Information Criterion (BIC), Hannan–Quinn information criterion (HQIC), the model performance is proved since the proposed distribution has minimum value for each of the criteria. The K-S, Cramer von misses W∗, Anderson Darling statistics A∗, and p-value for the proposed distribution show evidence that the new distribution fits the given data more than the competitors.

Table 4: Analytical Measures of Fitness and Adequacy for the infant mortality rate

Distr	NLL	AIC	CAIC	BIC	HQIC	W^*	A^*	K-S	P-value	Rank
EPCJ	102.59	211.035	212.078	214.922	212.191	0.038	0.261	0.095	0.9664	1
EPA	102.58	211.016	212.059	214.903	212.172	0.038	0.261	0.096	0.9658	2
PP	106.07	216.132	216.632	218.724	216.902	0.112	0.731	0.164	0.4596	4
PR	105.8	215.602	216.102	218.193	216.372	0.113	0.741	0.160	0.4933	3

Table 5: MLEs for the parameters (c, θ, α) using the infant mortality rate data

Distr	c	θ	α
EPCJ	5220.3257	8.5601	0.1311
	(0.0005)	(10.5592)	(0.1379)
EPA	7823.7103	7.4998	0.1412
	(0.0001)	(1.7445)	(0.0334)
PP		1.4559	0.5054
		(0.1947)	(0.0463)
PR		0.7371	0.5964
		(0.1383)	(0.0615)

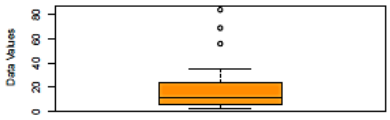

Figure 10: boxplot of the infant mortality rate data

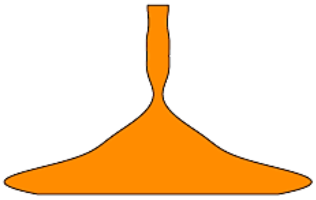

Figure 11: violin plot of the infant mortality rate data

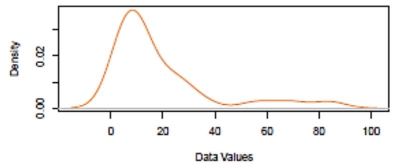

Figure 12: kernel density of the infant mortality rate data

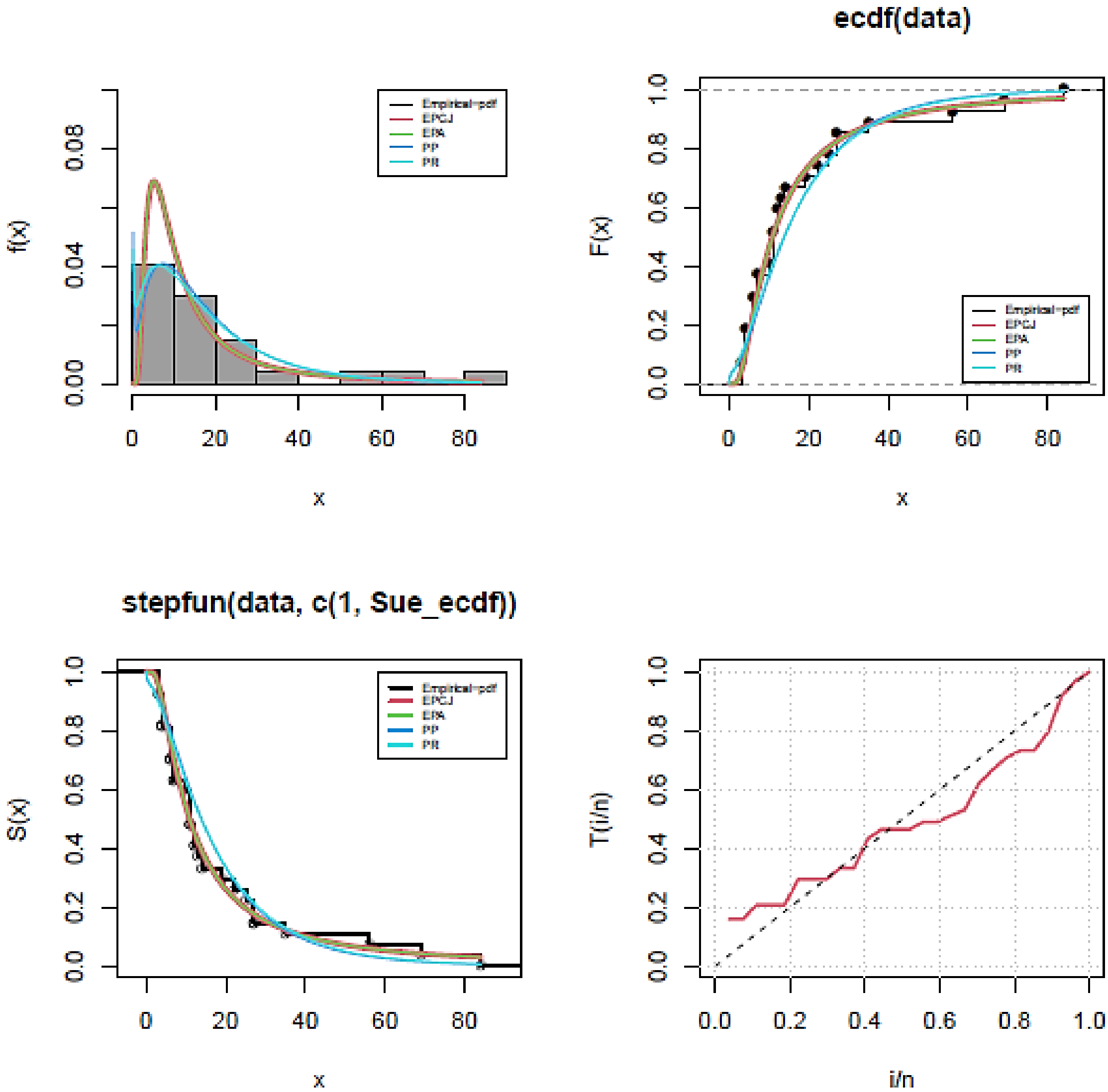

Figure 13: density, cdf, survival function, and TTT plots of the infant mortality rate data

Fig 10 and fig 11 are the boxplot and violin plots of the infant mortality rate data with outlier points very visible. Given that the proposed model best fits the data among competing distributions, it demonstrates the applicability of the EPCJ distribution in modeling skewed data sets. Similarly, fig 13 and fig 14 display graphically how well the proposed model fits the infant mortality data.

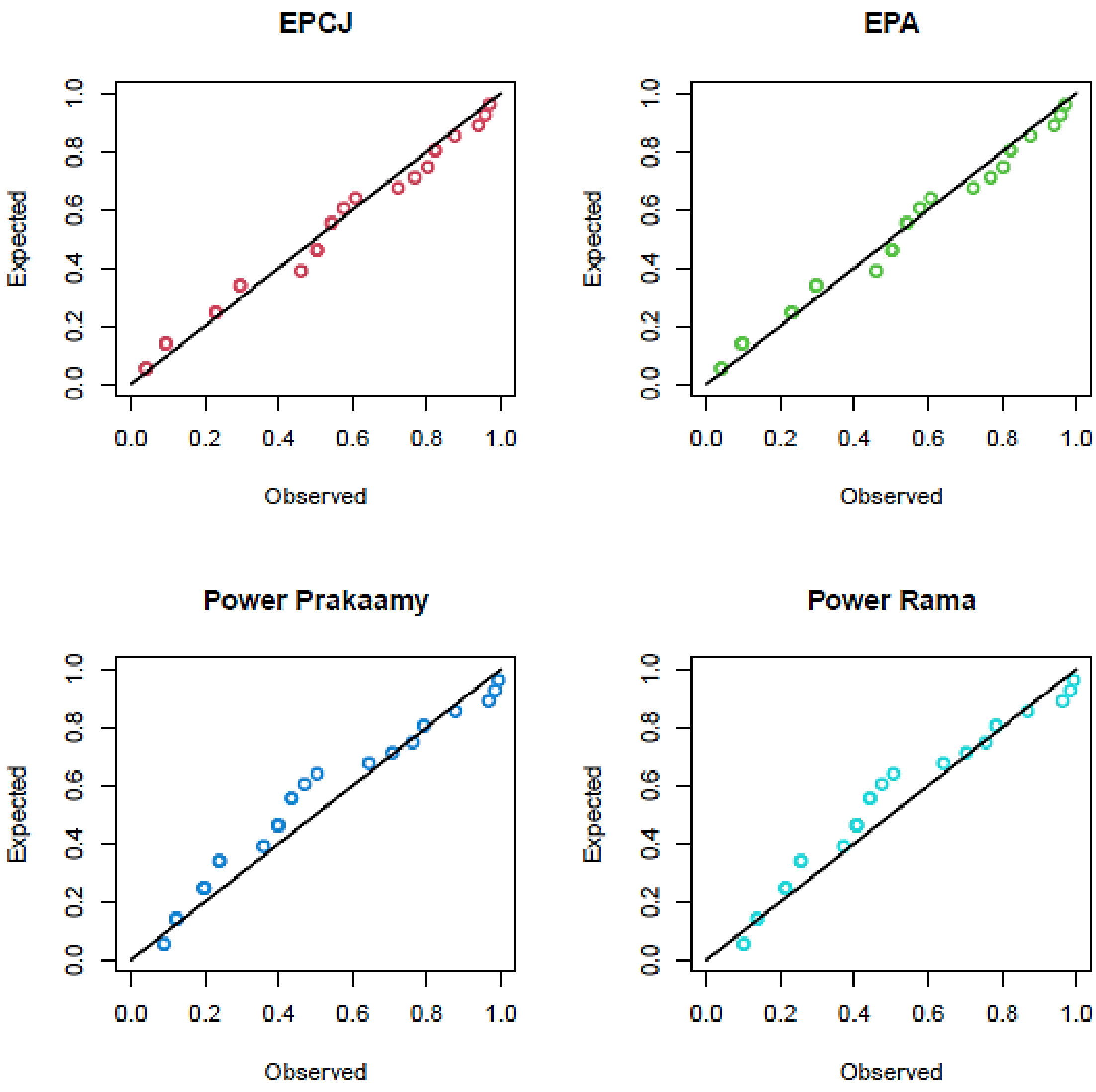

Figure 14: pp plots of the fitted distributions for the infant mortality rate data

CONCLUDING REMARKS

A new lifetime distribution with the potential of modeling skewed data has been proposed and studied in this article. The properties of the proposed distribution were derived and the log-transformation has been used to create a parametric regression model called the log-exponentiated power Chris-Jerry distribution regression model. The maximum likelihood estimation aided the estimation process for uncensored samples while the procedure for the estimation of the unknown parameters when data is censored was also shown. Essentially, the censored COVID-19 data set with the age of patients and diabetic mellitus index was deployed to justify the importance of the distribution. Furthermore, the distribution was fitted to the data on infant mortality rate (below age 5 years) reported for some countries by the World Health Organization in 2021. The distribution performs pretty well in both instances of application.

ACKNOWLEDGEMENT

The authors thank the reviewers and editors for their invaluable comments which enhanced the quality of this article.

FUNDING

This research did not receive any external funding.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Abebe, Berhane et al. (2019). “A two-parameter power Rama distribution with properties and applications”. In: Biometrics and Biostatistics International Journal 8.1, pp. 6–11.
2. Anabike, Ifeanyi C et al. (2023). “Inference on the parameters of Zubair-Exponential distribution with application to survival times of Guinea Pigs”. In: Journal of Advances in Mathematics and Computer Science 38.7, pp. 12–35.
3. Ashour, Samir K and Mahmoud A Eltehiwy (2015). “Exponentiated power Lindley distribution”. In: Journal of advanced research 6.6, pp. 895–905.

4. Baaqeel, Hanan et al. (2023). “Evaluating the System Reliability of the Bridge Structure Using the Unit Half- Logistic Geometric Distribution”. In: Scientific African, e01750.
5. Bhat, AA et al. (2023). “The odd lindley power rayleigh distribution: properties, classical and bayesian estimation with applications”. In: Scientific African 20, e01736.
6. Durrans, S Rocky (1992). “Distributions of fractional order statistics in hydrology”. In:Water Resources Research 28.6, pp. 1649–1655.
7. Etaga, Harrison O et al. (2023). “A New Modification of Shanker Distribution with Applications to Increasing Failure Rate Data”. In: Earthline Journal of Mathematical Sciences 13.2, pp. 509–526.
8. Ezeilo, Christiana I et al. (2023). “Exploring the Impact of Factors Affecting the Lifespan of HIVs/AIDS Patient’s Survival: An Investigation Using Advanced Statistical Techniques”. In: Open Journal of Statistics 13.4, pp. 594–618.
9. Ferreira, Alexsandro A and Gauss M Cordeiro (2023). “The exponentiated power Ishita distribution: Properties, simulations, regression, and applications.” In: Chilean Journal of Statistics (ChJS) 14.1.
10. Ganaie, Rashid A et al. (2023). “Exponentiated Adya Distribution: Properties and Applications”. In: Reliability: Theory & Applications 18.2 (73), pp. 39–49.
11. Gomaa, Rabab S et al. (2023). “The Unit Alpha-Power Kum-Modified Size-Biased Lehmann Type II Distribution: Theory, Simulation, and Applications”. In: Symmetry 15.6, p. 1283.
12. Innocent, Chidera F et al. (2023). “Estimation of the parameters of the power size biased Chris-Jerry distribution”. In: International Journal of Innovative Science and Research Technology 8.5, pp. 423–436.
13. Lehmann, Erich L (2012). The power of rank tests. Springer.
14. Lindley, Dennis V (1958). “Fiducial distributions and Bayes’ theorem”. In: Journal of the Royal Statistical Society. Series B (Methodological), pp. 102–107.
15. Musa, Abuh, Sidney I Onyeagu, and Okechukwu J Obulezi (2023a). “Comparative Study Based on Simulation of Some Methods of Classical Estimation of the Parameters of Exponentiated Lindley–Logarithmic Distribution”. In: Asian Journal of Probability and Statistics 22.4, pp. 14–30.
– (2023b). “Exponentiated Power Lindley-Logarithmic distribution and its applications”. In: Asian Research Journal of Mathematics 19.8, pp. 47–60.
16. Nwankwo, Mmesoma Peace et al. (2023). “Prediction of House Prices in Lagos-Nigeria Using Machine Learning Models”. In: European Journal of Theoretical and Applied Sciences 1.5, pp. 313–326.
17. Obulezi, Okechukwu, Chinyere P Igbokwe, and Ifeanyi C Anabike (2023). “Single acceptance sampling plan based on truncated life tests for zubair-exponential distribution”. In: Earthline Journal of Mathematical Sciences 13.1, pp. 165–181.
18. Obulezi, Okechukwu J, Ifeanyi C Anabike, Grace C Okoye, et al. (2023). “The Kumaraswamy Chris-Jerry Distribution and its Applications”. In: Journal of Xidian University 17.6, pp. 575–591.
19. Obulezi, Okechukwu J, Ifeanyi C Anabike, Orji G Oyo, et al. (2023). “Marshall-Olkin Chris-Jerry distribution and its applications”. In: International Journal of Innovative Science and Research Technology 8.5, pp. 522–533.
20. Obulezi, Okechukwu J, Nwabueze Nneka Chidimma, et al. (2023). “Statistical Analysis on Diagnosed Cases of Malaria and Typhoid Fever in Enugu-Nigeria”. In: GSJ 11.6.
21. Onyekwere, Chrisogonus K and Okechukwu J Obulezi (2022). “Chris-Jerry distribution and its applications”. In: Asian Journal of Probability and Statistics 20.1, pp. 16–30.
22. Oramulu, Dorathy O et al. (2023). “Simulation Study of the Bayesian and Non-Bayesian Estimation of a new Lifetime Distribution Parameters with Increasing Hazard Rate”. In: Asian Research Journal of Mathematics 19.9, pp. 183–211.
23. Rady, El-Houssainy A, WA Hassanein, and TA Elhaddad (2016). “The power Lomax distribution with an application to bladder cancer data”. In: SpringerPlus 5, pp. 1–22.
24. Ramadan, Ahmed T, Ahlam H Tolba, and Beih S El-Desouky (2022). “A unit half-logistic geometric distribution and its application in insurance”. In: Axioms 11.12, p. 676.
25. Rather, Aafaq A and C Subramanian (2019). “Exponentiated Ishita Distribution: Properties and Applications”. In: International Journal of Management, Technology and Engineering 9.5, pp. 2473–2484.
26. Shukla, Kamlesh Kumar and Rama Shanker (2020). “Power Prakaamy distribution and its applications”. In: International Journal of Computational and Theoretical Statistics 7.01.
27. Tolba, Ahlam et al. (2022). “Bayesian and non-Bayesian estimation methods for simulating the parameter of the Akshaya distribution”. In: Computational Journal of Mathematical and Statistical Sciences 1.1, pp. 13–25.
28. Tolba, Ahlam H, Abdisalam Hassan Muse, et al. (2023). “The Gull Alpha Power Lomax distributions: Properties, simulation, and applications to modeling COVID-19 mortality rates”. In: Plos one 18.9, e0283308.
29. Tolba, Ahlam H, Chrisogonus K Onyekwere, et al. (2023). “A New Distribution for Modeling Data with Increasing Hazard Rate: A Case of COVID-19 Pandemic and Vinyl Chloride Data”. In: Sustainability 15.17, p. 12782.