Pareto type-II-Exponential Odd Distribution and Estimation of Its parameters

Ban M.Tuij [1,b) and Ahmed AL-Adilee [2,a)

[1]Department of Mathematics, Faculty of Education for girls, University of kufa, Iraq
[2]Department of Mathematics, Faculty of Education, University of kufa, Iraq
[a)]ahmeda.aladilee@uokufa.edu.iq
[b)] banm.tuij@student.uokufa.edu.iq

Abstract. A new distribution namely Pareto type-II-Exponential from Pareto family is introduced which is denoted by (PEOD). We explore the cumulative distribution function (cdf) and its probability density function (pdf), as well as other features such as the survival function, moments, and graphs of such functions. We also use the maximum likelihood technique to discover the approximate values that meet the intended distribution and estimate the parameters model of the created distribution PEOD. Finally, we use the created PEOD to evaluate some data and compare the findings to the goodness-of-fit measurements with regard to the Pareto distribution to determine which distribution better matches the data.

Keywords: Continues distributions, Odd Distributions approaches, Moments, Goodness of Fit Measures, Data Set Analysis.

---

1. Introduction

Statistical distributions are often used to describe real-world data and play an essential part in parametric inference. Standard distributions do not suit all sorts of real data sets well in practice. As a result, statisticians are creating a plethora of novel distributions for the analysis of real data that are more flexible than traditional distributions. New distributions are created by merging two or more existing distributions or adding more parameters to existing distributions. In this paper, a new distribution was created called PEOD and now we will get to know it.

The PEOD is an odd distribution that is generated from combining an Exponential distribution with Pareto type-II distribution to obtain the requested odd distribution. Moreover, we estimate the new parameter model resulting from this combination by substituting the location parameter of the Pareto type-II distribution with Exponential distribution. A comparison of PEOD with Pareto type-II distribution has been discussed via the measures of goodness-of-fit. This method appears to improve the analysis of data sets that have been used to compare the Pareto type-II distribution using statistics genuine data sets. The usage of maximum likelihood method to estimate the parameter for a two parameter

Pareto type-II distribution with estimation of three parameters model of PEOD. We have also utilized the goodness-of-fit measures to find the optimal distribution that matches the data set among others.

Now, some common estimation in Gumbel-Pareto distribution [1], Beta-pareto distribution [2], Weibull-Pareto distribution [3], beta-gamma [4], Exponential-Weibull Exponential [5], Pareto Poisson–Lindley distribution [6], Weibull- generalized exponential [7] and gamma-uniform [8], Rayleigh-uniform distribution [9].

Finally, we refer to sections of this paper: in the following section, we go over some fundamental ideas linked to statistical concepts such as pdf, cdf, and others. The primary themes that this research is concerned with are presented in part three. We give numerous phrases that describe the entire work in section four.

2. Preliminaries and Background

We review some preliminaries related to Exponential distribution and Pareto type-II distribution. The pdf, the cdf, survival function $\mathrm{S}(x)$, and the hazard function $\mathrm{h}(x)$, [10].

$$\boldsymbol{f(x)=\frac{\alpha\beta^{\alpha}}{(x+\beta)^{\alpha+1}}}\ ,\ \boldsymbol{0< x<\infty} \tag{2.1}$$

$$F(x)=1-\frac{\beta^{\alpha}}{(x+\beta)^{\alpha}}\ ,0< x<\infty \tag{2.2}$$

$$s(x)=\frac{\beta^{\alpha}}{(x+\beta)^{\alpha}}\ ,0< x<\infty \tag{2.3}$$

$$h(x)=\frac{f(x)}{s(x)}, \tag{2.4}$$

Also, we need only to review the pdf, the cdf, and the survival function distribution, respectively.[11]

$$\boldsymbol{g(x)=\lambda e^{-\lambda x},0\le x\le\infty\ \ ,\lambda>0} \tag{2.5}$$

where λ is scalar parameter

$$\boldsymbol{G(x)=1-e^{-\lambda x},\lambda>0} \tag{2.6}$$

$$\overline{G}\,(x)=e^{-\lambda x},\lambda>0 \tag{2.7}$$

From (2.6), and (2.7), a new random variable that is used to generate the odd distribution by the odd ratio can be obtained as follows:

$$\boldsymbol{X=\frac{G(X)}{\overline{G}(x)},\qquad 0\le x\le\infty\ \ ,\lambda>0} \tag{2.8}$$

We can also refer to the measures of goodness-of-fit that we will use to compare the best fit of data set, that is, see [12].

$$\boldsymbol{AIC=-2\,\hat{\ell}+2k} \tag{2.9}$$

$$\mathrm{BIC}=-2\,\hat{\ell}+\mathrm{k}\ln(\mathrm{m}) \tag{2.10}$$

$$\mathrm{CAIC} = -2\,\hat{\ell} + \frac{2\mathrm{km}}{\mathrm{m}-\mathrm{k}-1} \tag{2.11}$$

$$\mathrm{HQIC} = -2\,\hat{\ell} + 2\mathrm{k}\ln(\ln(\mathrm{m})) \tag{2.12}$$

where, $\hat{\ell}$ is the maximum log-likelihood function, m is the sample size of the desired data, k is the number of the parameters of each estimated distribution within the maximum likelihood method.

3. Methodology and Results

3.1. The Generating of PEOD

In this part, we present the combination of Pareto type-II distribution with Exponential distribution to generate the PEOD. By using the formula in equation (2.8), we obtain a random variable that has the following form.

$$\boldsymbol{X} = \frac{\boldsymbol{G(X)}}{\boldsymbol{\bar{G}(x)}} = \boldsymbol{e^{\lambda x} - 1}, \qquad \boldsymbol{\lambda > 0} \tag{3.1}$$

By substituting such random variable into the distribution function in (2.2), so this yields that

$$\mathbf{F(x) = p(X \le x) = p\left(X \le e^{\lambda x} - 1\right)} = \int_{\mathbf{0}}^{\mathbf{e^{\lambda x}-1}} \mathbf{f(x)dx = F(e^{\lambda x} - 1)}, \tag{3.2}$$

which means that we have a cdf of PEOD, that is:

$$\boldsymbol{F(x) = \ 1 - \frac{\beta^{\alpha}}{(\,e^{\lambda x} + \beta - 1)^{\alpha}},\ \ 0 < x < \infty\ , \alpha, \beta, \lambda\ >\ 0} \tag{3.3}$$

By differentiate $F(x)$ with respect to x, we obtain pdf of PEOD, that is

$$\boldsymbol{f(x)} = \frac{\boldsymbol{dF\left(\frac{G(X)}{\bar{G}(x)}\right)}}{\boldsymbol{dx}} = \frac{\boldsymbol{dF\left(e^{\lambda x} - 1\right)}}{\boldsymbol{dx}} = \frac{\boldsymbol{f(x)g(x)}}{\boldsymbol{[G(x)]^2}}, \tag{3.4}$$

Then, the pdf has the following form

$$\boldsymbol{f(x) = \alpha\lambda\beta^{\alpha}e^{\lambda x}\ \frac{1}{(\,e^{\lambda x} + \beta - 1)^{\alpha+1}}\ ,\ \ 0 < x < \infty\ , \alpha, \beta, \lambda\ >\ 0} \tag{3.5}$$

From the cdf and pdf in equations (3.3), and (3.5) we can obtain other properties like the survival function S(x), the hazard rate function h(x), reversed hazard rate function r(x), and the cumulative hazard rate function H(x) of the random variable X are, respectively given by:

$$\boldsymbol{s(x : \alpha, \beta, \lambda) = 1 - F(X) = \frac{\beta^{\alpha}}{(\,e^{\lambda x} + \beta - 1)^{\alpha}}}, \tag{3.6}$$

$$h(x : \alpha, \beta, \lambda) = \frac{f(x :\ \alpha, \beta, \lambda)}{\bar{F}(x :\ \alpha, \beta, \lambda)} = \frac{\alpha\lambda e^{\lambda x}}{(\,e^{\lambda x} + \beta - 1)}, \tag{3.7}$$

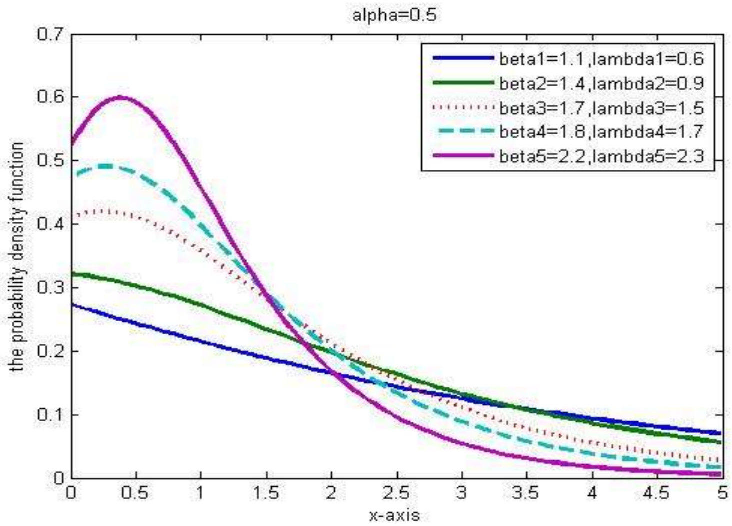
$$r(\,x : \alpha, \beta, \lambda) = \frac{f(x :\ \alpha, \beta, \lambda)}{F(x : \alpha, \beta, \lambda)} = \frac{\alpha\lambda\beta^{\alpha}e^{\lambda x}}{(\,e^{\lambda x} + \beta - 1)^{\alpha+1} - \beta^{\alpha}(\,e^{\lambda x} + \beta - 1)}\ , \tag{3.8}$$

$$H\,(x : \alpha, \beta, \lambda) = \,-Ln\,[1\, -\, F\,(x :\ \alpha, \beta, \lambda)] = \alpha ln\left(\,e^{\lambda x} + \beta - 1\right) - \alpha ln\beta, \tag{3.9}$$

Moreover, we have derived the formulas of moments about the origin, and the mean of the PEOD, which have the forms below:

$$\mu_r' = \frac{\alpha\beta^{\alpha}}{\sum_{i=0}^{\infty}\binom{\alpha+1}{i}(\beta-1)^{(\alpha-i+1)}\ (i-1)^{r+1}\lambda^{r}}\Gamma(r+1), \tag{3.10}$$

The plots of (cdf), (pdf) and survival function of the PEOD for different parameters. Also, the plots of the survival function $s(x)$ are given by the following figures



.

Figure 3.1:The p.d.f of PEOD

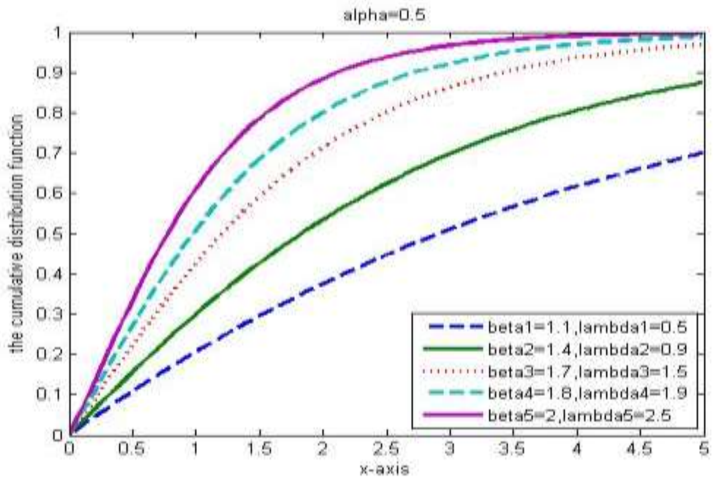


Figure 3.2:The c.d.f of PEOD

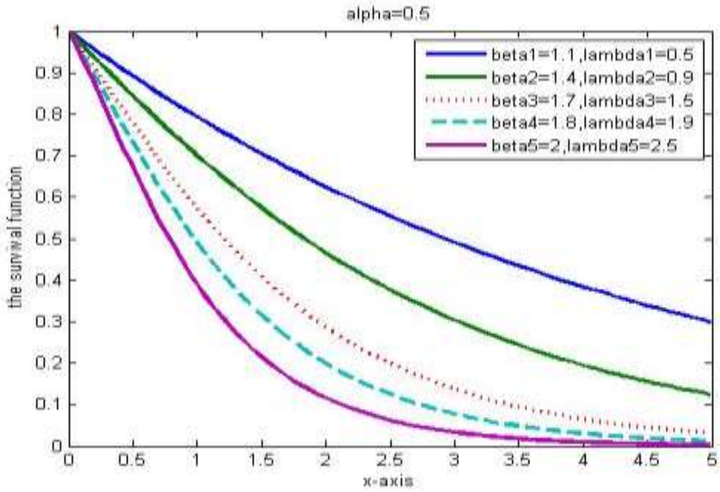


Figure3.3: The survival function of PEOD

3.2. Estimation of PEOD Parameters

We investigate the estimation of the parameters model of the generated distribution PEOD by the MLE. This method is preferred comparing to other methods because it almost yields unbiased estimators. If $X_1, X_2, \dots, X_n$ be a random sample from PEOD, then the likelihood function of those odd distribution is given by:

$$L = \prod_{i=1}^{n} f\,(x{:}\,\alpha, \beta, \lambda), \tag{3.11}$$

By substituting the pdf in (3. 5) into equation (3.11), we obtain the likelihood function, that is

$$L = \prod_{i=1}^{n} \alpha\lambda\beta^{\alpha} e^{\lambda x} \frac{1}{(e^{\lambda x} + \beta - 1)^{\alpha+1}} = \alpha^n \lambda^n \beta^{n\alpha} e^{\lambda \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \left(e^{\lambda x} + \beta - 1\right)^{-(\alpha+1)} \tag{3.12}$$

Then, the log-likelihood function of L respectively is:

$$\ell = ln(L) = nln\alpha + nln\lambda + n\alpha ln\beta + \lambda \sum_{i=1}^{n} x_i - (\alpha+1) \sum_{i=1}^{n} ln(e^{\lambda x_i} + \beta - 1) \tag{3.13}$$

parameter α, β, λ and setting the result to zero, we get the following equation:

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + nln\beta - \sum_{i=1}^{n} ln\left(e^{\lambda x_i} + \beta - 1\right), \tag{3.14}$$

$$\frac{\partial \ell}{\partial \beta} = \frac{n\alpha}{\beta} - (\alpha+1) \sum_{i=1}^{n} \frac{1}{(e^{\lambda x_i} + \beta - 1)}, \tag{3.15}$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} + \sum_{i=1}^{n} x_i - (\alpha+1) \sum_{i=1}^{n} \frac{\lambda e^{\lambda x_i}}{(e^{\lambda x_i} + \beta - 1)} \tag{3.16}$$

3.3. Application and Discussion

This section discusses the parameter estimation using the data set, as well as the analysis of lifespan data using the Pareto type-II distribution with estimated value of α, β, the generated PEOD in order to compare and explain the obtained conclusions of the investigated data. Using the goodness of fit metrics like AIC, BIC, and others, the best eyesight may be obtained by comparing the Pareto type-II distribution, and the PEOD. The real data taken from Colorado Climate Center, Colorado state University (http://ulysses.atmos.colostate.edu). These data consist of 100 annual maximum precipitation (inch) for one rain gauge in Fort Collins, Colorado, from 1900 through 1999.

Table 3.1: the data from Aarset

239	232	434	85	302	174	170	121	193	71	168	148	116	132	132	144	183	223	96	110
298	97	116	146	84	230	138	170	117	285	115	132	125	156	124	189	193	71	176	215
105	93	354	60	151	160	219	142	117	103	87	223	215	108	354	213	306	169	184	443
71	98	96	218	176	121	161	321	102	185	269	98	271	95	212	151	136	240	162	199
162	112	348	95	249	103	181	152	135	463	115	134	297	187	203	146	94	129	183	241

The calculated parameters model of the Pareto type-II distribution, and PEOD are compared to the conventional mixed distribution known. Matlab was used to complete this estimate, and Table 3.2 provides the estimated parameter values, which are shown as follows:

Table 3.2: MLEs of parameters, Log-likelihood

Model	MLE'S of parameters
Pareto	$\hat{\alpha} = 5.084,\ \hat{\beta} = 6.793$
PEOD	$\hat{\alpha} = 0.005,\ \hat{\beta} = 0.859,\ \hat{\lambda} = 0.792$

The predicted values of the parameter α, β of the Pareto type-II distribution and the PEOD are vary and almost not identical. The influence of exceeding the number of parameters in the PEOD increase the efficiency of the analysis of data set. This can be shown by using the required data set with the maximum log-likelihood functions (ℓ), and then compute the AIC, BIC, CACI, and HQIC values. This allows for the necessary comparison of the two distributions. In fact, Table 3.3 displays the values of each quality, that are

Table 3.3: Log-likelihood, AIC, AICC, BIC and HQIC values of models fitted

Model	$\hat{\ell}$	AIC	BIC	CAIC	HQIC
Pareto	-1979.3	3962.6	3967.8	3962.8	3964.7
PEOD	-623.1884	1252.4	1260.2	1252.6	1255.5

It is clear that the values of goodness-of-fit measures (AIC, BIC, CAIC, HQIC) within the PEOD are quietly better than their values within Pareto type-II distribution.

3 Conclusion

Modern properties such as a new pdf, cdf, moments, and others have been produced by combining the two selected distributions (Pareto type-II distribution and Exponential distribution), which is defined as PEOD. The PEOD graphs show how the pdf graphs are skewed to the right with varying kurtosis coefficients. The goodness-of-fit to the Pareto type-II distribution and the PEOD for the given data set shows that the PEOD is somewhat better than its standard distribution, since its maximum likelihood value is greater than the maximum likelihood of the Pareto type-II distribution, and the values of the goodness-of-fit measures are better too.

References

[1] Joseph J., Joseph KK 2021. Gumbel - Pareto distribution and it's applications in modeling COVID data ،Biometrics & Biostatistics International Journal ،v 10(3), 125-128

[2] Akinsete, A., Famoye, F. and Lee, C., 2008. The beta-Pareto distribution. *Statistics*, *42*(6*), pp.547- 563.*

[3] Tahir, M.H., Cordeiro, G.M., Alzaatreh, A., Mansoor, M. and Zubair, M., 2016. A new Weibull–Pareto distribution: properties and applications. *Communications in Statistics-Simulation and Computation*, *45*(10), pp.3548-3567

[4] Zoografts, K. and Balakrishnan, N., 2009. On families of beta-and generalized gamma-generated distributions and associated inference. *Statistical methodology*, *6*(4), pp.344-362.

[5] S. Muhammad and K. Abed AL-Kadin, Odd Generalized Exponential Weibull Exponential Distribution (Journal of Engineering and Applied sciences, 2019), Vol. 16(11), pp. 10360-10386. .

[6]Asgharzadeh, A., Bakouch, H. S., & Esmaeili, L. (2013). Pareto Poisson–Lindley distribution with applications. *Journal of Applied Statistics*, *40*(8), 1717-1734.

[7] Mustafa, A., El-Desouky, B.S. and AL-Girahs, S., 2016. Weibull generalized exponential distribution. *axis preprint arXiv:1606.07378*.

[8] H. To rabi, & N. M. Heeds, The gamma-uniform distribution and its applications (Cybernetic, 2012), 48(1), 16-30, 2012.

[9]Tuij, B. M., & Al-Adilee, A. (2021). Rayleigh-uniform odd distribution and estimation of its parameters. *International Journal of Nonlinear Analysis and Applications*, *12*(Special Issue), 2223-2230.

[10]Baharith, L., & Alzahrani, H. (2019). New bivariate Pareto type II models. Entropy, 21(5), 473..

[11]Marshall, A. W., & Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, *62*(317), 30-44.

[12]H. Bozdogan, Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions, Psychometrika, 52 (1987), 345–370