G-Transmuted Pareto Distribution for Modeling Aircraft Windshields

Abstract

In this paper, we investigated flexible and skewed families of transmuted distributions obtained by introducing one or more additional parameters to a baseline distribution. Specifically, we propose a novel three-parameter G-Transmuted Pareto Distribution. A comprehensive analysis of its mathematical properties is presented, along with an evaluation of its reliability characteristics. Using real-world data, we illustrate the distribution's effectiveness in modeling and demonstrate its utility as a robust statistical tool for various practical applications.

Keyword

Pareto Distribution, G-Transmuted, Hazard rate function, Maximum likelihood, Reliability Analysis.

1. INTRODUCTION

The Pareto distribution, introduced by Vilfredo Pareto in 1897[1], is a power-law probability distribution originally developed to model wealth allocation among individuals. Since its inception, numerous extended forms of this distribution have been proposed and have found applications across different fields. Initially, the distribution was primarily utilized in studies of income distribution, later expanding to areas such as wealth allocation (Wold & Whittle, 1957)[2].

Subsequent research has explored various aspects of the Pareto distribution. Doostparast et al. (2011) examined Bayesian analysis methods for the two-parameter Pareto distribution, particularly in the context of record values [3]. Fader et al. (2011) developed the beta-geometric/NBD model for applications in various purchasing environments [4]. Gyan Prakash (2012) focused on Bayesian prediction intervals for the generalized Pareto distribution [5]. Huda Alshanbari et al. (2021) introduced a novel mixture distribution derived from a combination of Fréchet-Weibull and Pareto distributions [6]. More recently, Sankara Narayanan et al. (2024) proposed a modified double sampling plan to optimize inspection processes that require minimal sample sizes, ensuring product median life under a Weibull-Pareto framework while minimizing costs [7].

Yilmaz et al. (2018) proposed a new modified transmutation based on the quadratic rank transmutation map (QRTM) [8] originally introduced by Shaw and Buckley (2007) [9], extending the range of the transmutation parameter from $[-1,1]$ to $[-1,2]$, thus enhancing the flexibility of the distribution family. This increased flexibility is further demonstrated in similar studies on the generalized transmuted G family by Nofal et al. (2017) [10] and the generalized transmuted Weibull distribution by Nofal and El Gebaly (2017). The flexibility of the G-Transmuted distribution facilitates the modeling of complex datasets that may not fit traditional distributions, thus providing researchers and practitioners with a powerful tool for data analysis [11].

These developments highlight the continued significance and versatility of the Pareto distribution and its extensions in modeling intricate real-world phenomena, emphasizing the need for further investigation across diverse scientific fields.

The structure of this paper is organized as follows: in Section 2 is devoted to the main features of a three-component G-Transmuted Pareto distribution. In Section 3 and 4, we present GTPD with its statistical analysis, including the survival and hazard rate functions. Section 5 provides expressions for the likelihood function method to estimate the parameters of the distribution. In Section 6, we conduct a simulation study to compare the performance of the estimators based on mean squared errors (MSEs) and biases using MLE. Finally, in Section 7, we apply the G-Transmuted Pareto Distribution model to real-life datasets to demonstrate its practical utility.

2- A three-component G-Transmuted Pareto Distribution

The Pareto distribution is widely recognized for modeling phenomena characterized by the unequal distribution of resources, such as wealth and income. Its foundational role in both theoretical and applied statistics has led researchers to extend its applicability through various modifications and transformations. One significant modification is the G-Transmuted distribution, which introduces additional parameters for greater flexibility in modeling asymmetric data. The G-Transmuted Pareto distribution is particularly relevant in fields where data exhibit skewness and heavy tails. For instance, it can be applied in financial modeling to assess risks associated with extreme market events, in insurance for modeling claim sizes, and in environmental studies to describe the distribution of resource consumption.

Let X be a random variable from a Pareto distribution with the following cumulative distribution function (cdf):

$$F(x) = 1 - \left(\frac{\beta}{x}\right)^{\alpha} \ ; \ x \geq \beta,, \alpha > 0, \beta > 0 \qquad (1)$$

The associated probability density function (PDF) is given by:

$$f(x) = \alpha f(x) = \frac{\alpha\,\beta^{\alpha}}{x^{\alpha+1}} \ , \qquad ; \alpha > 0, \beta > 0 \ for \ x > \beta \qquad (2)$$

The shape parameter α governs the distribution's behavior, with larger values leading to a steeper decline in the tail of the distribution, while the scale parameter β shifts the distribution along the x-axis. The heavy-tailed nature of the Pareto distribution makes it particularly suitable for modeling extremes, where a small number of observations have a significant impact.

The G-Transmuted distribution is a transformation of the original Pareto distribution that introduces additional flexibility. The CDF of a G-Transmuted random variable can be expressed as:

$$G(x) = (1+\lambda)F^{2}(x) - \lambda\, F^{3}(x); \ \lambda \in [-1, 2] \qquad (3)$$

This transformation allows for a wider array of shapes in the resultant distribution, making it particularly useful in contexts where the data exhibit characteristics not adequately captured by the standard Pareto model.

3- Statistical Properties of G-Transmuted Pareto Distribution

By using the transformation given by equation (3), we can get the cdf and the pdf of GTPD with two parameters θ, β and λ as follow

$$F(x,\alpha,\beta,\lambda) = \left(-1+\left(\frac{\beta}{x}\right)^{\alpha}\right)^{2}\left(1+\left(\frac{\beta}{x}\right)^{\alpha}\lambda\right) \tag{4}$$

$$f(x,\alpha,\beta,\lambda) = \left[-\frac{\alpha}{x}\left(\frac{\beta}{x}\right)^{\alpha}\left(-1+\left(\frac{\beta}{x}\right)^{\alpha}\right)\left(2+\left(-1+3\left(\frac{\beta}{x}\right)^{\alpha}\right)\lambda\right)\right] \tag{5}$$

Fig.1 illustrate some of the possible shapes of the PDF of the GTPD at different values of the parameters.

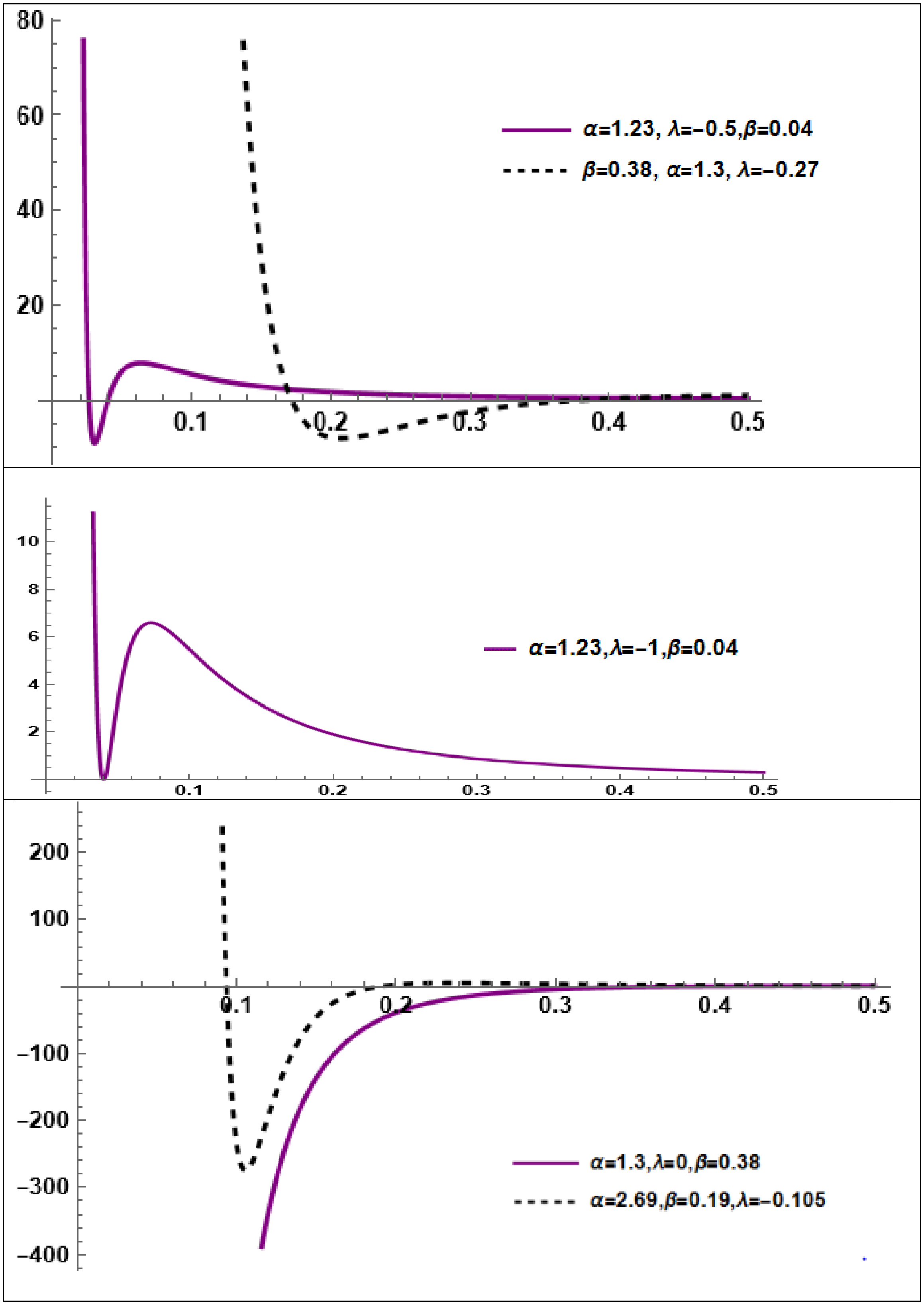

Fig1: Different shapes of the pdf for the GTPD at different values of α, β and λ.

In the following section, we will derive the noncentral and central moments of the GTPD and present the moment generating function. By analyzing these statistical properties, we aim to provide a comprehensive understanding of the GTPD and its implications in various applications.

So we can get the first non-central moment about the origin (mean μ) of TMD as follows:

$$\mu`_1 = \mu = \frac{2\alpha^2\beta(-1+3\alpha-\lambda)}{(-1+\alpha)(-1+2\alpha)(-1+3\alpha)} \tag{6}$$

Also, the second non-central moment $\mu`_2$ is given by

$$\mu`_2 = \frac{\alpha^2\beta^2(3\alpha-2(1+\lambda))}{(-2+\alpha)(-1+\alpha)(-2+3\alpha)} \tag{7}$$

Finally, the r^{th} non-central moment can be written as

$$\mu`_r = \frac{2\alpha^2\beta^r(r-3\alpha+r\lambda)}{(r-3\alpha)(r-2\alpha)(r-\alpha)} \tag{8}$$

We can also get the r^{th} moment about the mean (central moment) of GTPD by the following relation: Where the variance is the second central moment which can be obtained as follow:

$$\sigma^2 = \mu_2$$

$$\frac{\left(\alpha^2\beta^2\left((1-3\alpha)^2(-2+3\alpha)(-1+5\alpha)-2(-1+3\alpha)\left(1+\alpha\left(-8+\alpha(7+4\alpha)\right)\right)\lambda-4(-2+\alpha)\alpha^2(-2+3\alpha)\lambda^2\right)\right)}{(4-8\alpha+3\alpha^2)(-1+6\alpha-11\alpha^2+6\alpha^3)^2} \tag{9}$$

Then the standard deviation will be

$$\boldsymbol{\sigma} = \sqrt{\frac{\left(\alpha^2\beta^2\left((1-3\alpha)^2(-2+3\alpha)(-1+5\alpha)-2(-1+3\alpha)\left(1+\alpha\left(-8+\alpha(7+4\alpha)\right)\right)\lambda-4(-2+\alpha)\alpha^2(-2+3\alpha)\lambda^2\right)\right)}{(4-8\alpha+3\alpha^2)(-1+6\alpha-11\alpha^2+6\alpha^3)^2}} \tag{10}$$

4- Reliability Analysis using G-Transmuted Pareto Distribution

The study of characteristics of any probability distributions is fundamental in understanding the behavior of random variables. Among these characteristics, the survival function, the hazard rate function, the reversed hazard rate function, the cumulative hazard rate function, and the mean residual lifetime which are particularly important in the fields of reliability theory, survival analysis, and risk assessment. In this section, we analyze these functions for the GTPD.

Let **X** be a random variable, representing the lifetime of an item or a component within any system, has cdf $\boldsymbol{F(x)}$. The survival function in this case, which is the probability of not failing under specified conditions for a given period of time, can be defined by $\boldsymbol{R(x)}$ and is given by

$$R(x) = P(X > x) = 1 - F(x)$$

$$= -\frac{\left(\left(\frac{\beta}{x}\right)^{\alpha}\left(-1+\left(\frac{\beta}{x}\right)^{\alpha}\right)(x^{\alpha}\beta^{\alpha}(1-2\lambda)+x^{2\alpha}(-2+\lambda)+\beta^{2\alpha}\lambda)\left(2+\left(-1+3\left(\frac{\beta}{x}\right)^{\alpha}\right)\lambda\right)\right)}{\left((x^{\alpha}-\beta^{\alpha})(x^{\alpha}(-2+\lambda)-3\beta^{\alpha}\lambda)\right)} \tag{11}$$

Also, its hazard rate function $\boldsymbol{h(x)}$, which is defined as the instantaneous rate at which events occur given that no event has occurred up to time *t*, can be expressed as:

$$h(x) = \lim_{\Delta x \to 0} \frac{pr(x < X < x + \Delta x \;/\; X > x)}{\Delta x} = \frac{f(t)}{1 - F(x)}$$
$$= \frac{\alpha(x^{\alpha} - \beta^{\alpha})(x^{\alpha}(-2+\lambda) - 3\beta^{\alpha}\lambda)}{x(x^{\alpha}\beta^{\alpha}(1-2\lambda) + x^{2\alpha}(-2+\lambda) + \beta^{2\alpha}\lambda)} \tag{12}$$

The behavior of the hazard rate function $h(x)$ might be constant, increasing or decreasing depending on the values of the parameters involved as shown in Fig 2.

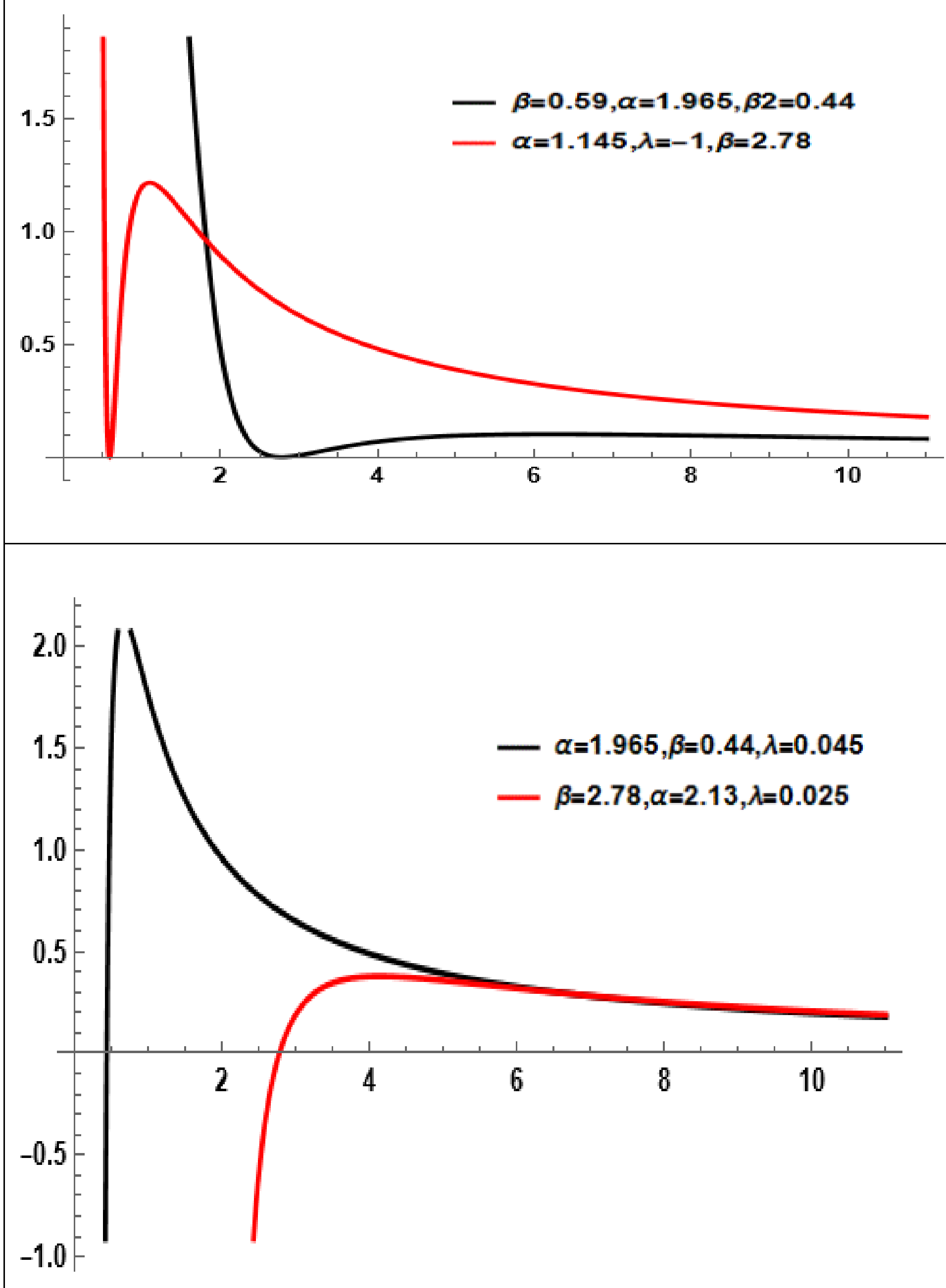

Fig2: Different shapes of the hazard rate function of the GTPD

The reversed hazard rate function $r(x)$, which is the probability of observing an outcome in a neighborhood of x, conditional on the outcome being no more than x, is given by

$$r(x) = \lim_{\Delta x \to 0} \frac{pr(< X < x + \Delta x | X \leq x)}{\Delta x} = \frac{f(x)}{F(x)} =$$
$$\frac{\alpha(\frac{\beta}{x})^{\alpha}(-1 + (\frac{\beta}{x})^{\alpha})(2 + (-1 + 3(\frac{\beta}{x})^{\alpha})\lambda)}{x + \frac{x(\frac{\beta}{x})^{\alpha}(-1 + (\frac{\beta}{x})^{\alpha})(x^{\alpha}\beta^{\alpha}(1 - 2\lambda) + x^{2\alpha}(-2 + \lambda) + \beta^{2\alpha}\lambda)(2 + (-1 + 3(\frac{\beta}{x})^{\alpha})\lambda)}{(x^{\alpha} - \beta^{\alpha})(x^{\alpha}(-2 + \lambda) - 3\beta^{\alpha}\lambda)}} \tag{13}$$

5- Estimation of The Parameters

Parameter estimation is a fundamental aspect of statistical inference, enabling us to get information about population parameters based on sample data. Among the different methods available, the method of maximum likelihood estimation (MLE) which is particularly effective in providing point estimates and constructing confidence intervals for unknown parameters. Integer parameters often arise in discrete distributions and other applications so Dahiya [12] introduced an improved graphical method to estimate integer-valued parameters using MLE, which was further refined in subsequent works (Dahiya [13], Olkin et al. [14]). Additionally, Miller [15] specifically addressed the MLE of integer-valued parameters within the Erlang distribution, contributing valuable insights to this area of study.

In this section, we will investigate different techniques for estimating parameters based on the behavior of the likelihood function. By analyzing a sample of size n from the Truncated Modified Discrete (GTPD) distribution, we aim to identify effective estimation strategies and develop robust confidence intervals for the unknown parameters. Let $X_{\cdot}, X_2, \ldots, X_n$ be a sample size n from GTPD, then the likelihood function (l) is given by

$$l(\alpha, \beta, \lambda) = \prod_{i=1}^{n} f(X_i \alpha, \beta, \lambda) = \prod_{i=1}^{n} \frac{-\alpha}{x_i} \prod_{i=1}^{n} (\frac{\beta}{x_i})^{\alpha}$$
$$\prod_{i=1}^{n} \left(-1 + (\frac{\beta}{x_i})^{\alpha}\right) \prod_{i=1}^{n} \left(2 + \lambda \left(-1 + 3(\frac{\beta}{x_i})^{\alpha}\right)\right) \tag{14}$$

Then, the log-likelihood function will be

$$\mathcal{L}(\alpha,\beta,\lambda) = lnl(\alpha,\beta,\lambda) = \ln(-n\alpha) - \sum_{i=1}^{n} lnx_i + \alpha ln\beta - \alpha \sum_{i=1}^{n} lnx_i$$

$$+\sum_{i=1}^{n}\left(-1+(\frac{\beta}{x_i})^{\alpha}\right) + \sum_{i=1}^{n}\left(2+\lambda\left(-1+3(\frac{\beta}{x_i})^{\alpha}\right)\right) \qquad (15)$$

A natural way for estimating the parameter θ is to assume that it can take on a continuum of values and calculate derivatives with respect to the two parameters λ and θ, then solve

$U(\alpha,\beta,\lambda)_1 = \frac{\partial\mathcal{L}(\alpha,\beta,\lambda)}{\partial\alpha} = 0$ (16) , $U(\alpha,\beta,\lambda)_2 = \frac{\partial\mathcal{L}(\alpha,\beta,\lambda)}{\partial\beta} = 0$ (17)

and $U(\alpha,\beta,\lambda)_3 = \frac{\partial\mathcal{L}(\alpha,\beta,\lambda)}{\partial\lambda} = 0$ (18)

To solve the equations (26) through (18), we can use nonlinear optimization algorithms such as quasi-Newton algorithm to numerically maximize the log-likelihood function. for computing standard errors and asymptotic confidence intervals, we rely on the l large sample approximation, where the maximum likelihood estimates (MLEs) can be approximated as being multivariate normal. The digamma function $\varphi(k) = \frac{d}{d\,k}\ln\Gamma(k) = \frac{\Gamma'(k)}{\Gamma(k)}$ plays a crucial role in these calculations. The MLEs $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\lambda}$ of are obtained by solving the non-linear equations $U(\hat{\alpha},\hat{\beta},\hat{\lambda})_1 = 0$, $U(\hat{\alpha},\hat{\beta},\hat{\lambda})_2 = 0$ and $U(\hat{\alpha},\hat{\beta},\hat{\lambda})_3$.

6- Simulation Study

In this section, various simulation schemes, as outlined in Table 1, were employed to examine the mean square error (MSE) and average bias (Bias) of the parameters of the G-Transmuted of Pareto Distribution. For this purpose,100 random samples of sizes 20, 40, 60, 80, and 100 were generated.

Table 1. MSE and Bias values for the parameter $\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\lambda}$ of GTPD

	$\boldsymbol{\alpha}$= 0.6		$\boldsymbol{\beta}$ =0.4		$\boldsymbol{\lambda}$= 0.3	
N	MSE	Bais	MSE	Bais	MSE	Bais
20	0.0123567	0.0240951	0.00570608	0.0260873	0.213325	0.0409148
40	0.0108294	0.00274165	0.00220754	0.0122004	0.288877	0.105248
60	0.00472744	0.00237286	0.00117708	0.00351447	0.225311	0.0270988
80	0.00570612	-0.0039380	0.000706654	0.00296982	0.179755	0.0401196
100	0.00574952	-0.0203739	0.000692137	0.00855779	0.252593	0.164376

7- Application

In this section, the flexibility of the new GTPD distribution is investigated based on a real data set as compared with some other competing distributions such as transmuted Lindely distribution, Monsef distribution, two parameter Exponential distribution and transmuted Monsef distribution. The estimated values of the parameters, log-likelihood statistic, Akaike information criterion (AIC), Bayesian information criterion (BIC), and the Kolmogorov–Smirnov (KS) statistics are presented in the following tables.

The following data consist the failure times of 23 communication transceiver and aircraft windshield. The data set is: 0.2887, 0.822, 0.3646, 0.3014, 49.059 , 0.522, 2.007, 0.145, 0.1509, 0.327, 3.931, 0.194,

0.3079,0.71024,0.2213, 0.6064, 0.5745, 0.74437,0.1477, 0.1193, 1.868,1.28023, 0.276587.

Table 2. Fitted estimates for different distributions for data

Distribution	MLEs	-Log	AIC	BIC	AICC	HQIC	CAIC
GTPD	$\hat{\alpha} = 0.8102$ $\hat{\beta} = 0.1057$ $\hat{\lambda} = 0.5396$	19.2152	44.4305	47.837	45.693	45.287	45.693
TwoParameter Exponential	$\hat{\lambda}$ =1.0136 $\hat{\gamma} = 1.21$	56.15	116.316	118.587	116.916	116.887	116.916
MonsefDis	$\hat{\lambda}$ = ٠.٥	79.3072	160.614	161.75	160.805	160.89	160.805
TrnsLindely	$\hat{\lambda} = 0.91906$, $\hat{\theta}$=0.5	51.268	106.538	108.809	107.138	107.19	107.139
TrnsMons	$\hat{\lambda} = 0.8914$, $\hat{\theta}$= 0.7466	63.911	131.82	134.091	132.421	132.392	132.42

Table 3. Goodness-of-fit tests for data

Distribution	Kolmogorov–Smirnov Statistic	Cramer–von Mises	Waston	Waston2	Anderson
GTPD	0.09	0.02886	0.02878	5.2916	0.2059
TwoParam. Exponential	3.1109	97.909	25.577	105.57	Indeterminate

MonsefDis	0.73273	5.0187	1.29044	10.6738	42.6218
TrnsLindely	0.5066	2.0091	0.5748	7.5146	11.588
TrnsMons	0.5653	2.54114	0.70384	8.07896	16.2382

8- Conclusion

In conclusion, the G-Transmuted Pareto distribution represents a significant advancement in the modeling of asymmetrical data. By incorporating additional parameters through the transmutation process, this distribution maintains the foundational principles of the Pareto while expanding its applicability. Future research should focus on empirical validation of the G-Transmuted distribution in various real-world datasets to assess its performance and robustness compared to other established distributions.

9- References

1. Pareto, V. 1897. Cours d'econmic politique. Rouge and Cie, Paris.
2. Wold, H. O. A. and Whittle, P. 1957. A model explaining the Pareto distribution of wealth. Econometrica 25 : 591-595.
3. Doostparast, M., Akbari, M. G. and Balakrishna, N. 2011. Bayesian analysis for the two-parameter Pareto distribution based on record values and times. Journal of Statistical Computation and Simulation 81(11):1393-1403.
4. Fader, P. S., Hardie, B. G. S. and Lee, K. L. 2011. Non perspectives on customer death using ageneralization of Pareto/NBD model. Marketing Research 30 : 866-886.
5. Gyan Prakash 2012. Central coverage Bayes prediction Intervals for the generalized Pareto distribution Statistics Research Letters 1(1) :1-5.
6. Huda M. Alshanbari , Abd Al-Aziz Hosni El-Bagoury , Ahmed M. Gemeay , E. H. Hafez , and Ahmed Sedky Eldeeb (2021), "A Flexible Extension of Pareto Distribution:Properties and Applications", Computational Intelligence and Neuroscience Volume 2021, Article ID 9819200, 7 pages.
7. Sankara Narayanan, M., Jeyadurga, P. and Balamurali, S. (2024), "Optimal and economic design of modified double sampling plan under new Weibull–Pareto distribution", International Journal of Quality & Reliability Management, Vol. 41 No. 5, pp. 1235-1252.
8. Mehmet Yilmaz, Sibel Acik Kemaloglu.2018. "Two New Distribution Families Constructed by a Modified Rank". MATHEMATICAL SCIENCES AND APPLICATIONS E-NOTES 6 (1) 106-112.
9. Shaw, W.T and Buckley, I.R.C.(2007), "The Alchemy of Probability Distributions: Beyond Gram-Charlier and Cornish Fisher Expansions, and Skew-Normal or Kurtotic-Normal Distributions".
10. Nofal, Z. M., Afify, A. Z., Yousof, H. M., and Cordeiro, G. M. , (2017), "The generalized transmuted-G family of distributions". Communications in Statistics-Theory and Methods, 46(8) 4119-4136.
11. Nofal, Z. M., and El Gebaly, Y. M., (2017), "The Generalized Transmuted Weibull Distribution for Lifetime Data". Pakistan Journal of Statistics and Operation Research, 13(2), 355-378

12. Dahiya, R. (1981). "An improved method of estimating an integer-parameter by maximum likelihood". Am. Stat. 35(1):34–37.
13. Dahiya, R. (1986)." Integer-parameter estimation in discrete distributions". Commun. Stat. - Theory Meth-ods. 15:709–725.
14. Olkin, I., Petkau, A., Zidek, J. (1981). "A comparison of n estimators for the binomial distribution". J. Am.Stat. Assoc. 70:685–689.
15. Miller, G. (1999). "Maximum likelihood estimation for the Erlang integer parameter". Stat. Probab. Lett. 43:335–341.
16. D. Kundu, M.Z. Raqab, "Estimation of $R = P(Y < X)$ for three-parameter Weibull distribution". Stat. Probab. Lett. 79(17) (2009), 1839-1846.