G-Transmuted Pareto Distribution for Modeling Failure Time Data

Abstract

In this paper, we investigate flexible and skewed families of transmuted distributions derived 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 provided, alongside an evaluation of its reliability characteristics. Using real-world data, we demonstrate the effectiveness of the proposed distribution in modeling and highlight its utility as a robust statistical tool for a real lifepractical application.

Keywords

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

1- Introduction

The Pareto distribution, introduced by Vilfredo Pareto, is a power-law probability distribution originally developed to model wealth allocation among individuals [1]. Since its inception, numerous extended forms of this distribution have been proposed, finding applications across various fields. Initially, the distribution was primarily utilized in studies of income distribution but later expanded to areas such as wealth allocation [2].

Subsequent research has explored various aspects of the Pareto distribution. Doostparast et al. examined Bayesian analysis methods for the two-parameter Pareto distribution, particularly in the context of record values [3]. The Pareto I model (Hill estimator) was found to be highly sensitive to threshold selection. Setting the threshold too low resulted in a significant underestimation of the tail index, which led to substantial overestimation of inequality [4]. Gyan Prakash focused on Bayesian prediction intervals for the generalized Pareto distribution [5], while Huda Alshanbari et al. introduced a novel mixture distribution derived from a combination of Fréchet-Weibull and Pareto distributions [6]. Due to its inherent property of heavy-tailed behavior, the Pareto distribution has frequently been applied to model the upper tail (or top income) of income distributions [7].

Recent advancements have further extended the Pareto distribution. A new generalization using the Marshall-Olkin generator and the method of alpha power transformation was introduced [8]. More recently, Sankara Narayanan et al. proposed a modified double sampling plan to optimize inspection processes with minimal sample sizes, ensuring product median life under a Weibull-Pareto framework while minimizing costs [9]. Yilmaz et al. proposed a new modification of transmutation based on the quadratic rank transmutation map (QRTM) [10], originally introduced by Shaw and Buckley [11]. This modification extended the range of the transmutation parameter from $[-1, 1]$ to $[-1, 2]$, enhancing the flexibility of the distribution family. Similar studies on the generalized transmuted G family by Nofal et al. [12] and the generalized transmuted Weibull distribution by Nofal and El Gebaly [13] further demonstrated this flexibility.

The G-Transmuted distribution's adaptability enables it to model complex datasets that may not align well with traditional distributions, providing researchers and practitioners with a robust tool for data analysis. These developments underscore the continued significance and versatility of the Pareto distribution and its extensions in modeling intricate real-world phenomena, emphasizing the need for further exploration across diverse scientific fields.

The structure of this paper is organized as follows: Section 2 discusses the main features of a three-parameter G-Transmuted Pareto distribution. Sections 3 and 4 present the statistical analysis of the G-Transmuted Pareto distribution (GTPD), including the survival and hazard rate functions. Section 5 provides expressions for the likelihood function method used to estimate the parameters of the distribution. In Section 6, we conduct a simulation study to compare the performance of the estimators using mean squared errors (MSEs) and biases through maximum likelihood estimation (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 extension is the G-Transmuted distribution, which introduces additional parameters to increase flexibility in modeling asymmetric data. The G-Transmuted Pareto distribution is particularly relevant in fields where data exhibit skewness and heavy tails. For example, it can be applied in financial modeling to assess risks associated with extreme market events, in insurance to model 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):

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

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 resulting in a steeper decline in the tail, 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 disproportionate 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:

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

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 $\alpha, \beta$and λ as follow

$$F(x, \alpha, \beta, \lambda) = \left(-1 + \left(\frac{\beta}{x}\right)^{\alpha}\right)^2 \left(1 + \left(\frac{\beta}{x}\right)^{\alpha} \lambda\right) \qquad , x > \beta, \alpha, \beta > 0, -1 < \lambda < 1 \qquad (4)$$

$$f(x, \alpha, \beta, \lambda) = \frac{\alpha \left(\frac{\beta}{x}\right)^{\alpha} \left(1 - \left(\frac{\beta}{x}\right)^{\alpha}\right) \left(2 + \left(-1 + 3\left(\frac{\beta}{x}\right)^{\alpha}\right) \lambda\right)}{x} \qquad , x > \beta, \alpha, \beta > 0, -1 < \lambda < 1 \qquad (5)$$

The PDF of the GTPDisunimodal as illustrated in Fig 1. and takes its maximum at

$$x_0 = 3^{\frac{1}{\alpha}}\beta\left(\frac{-1 + 2\lambda + \alpha(-2 + 4\lambda) + \sqrt{(1 + \lambda)^2 + 4\alpha(1 + \lambda)^2 + \alpha^2(4 + \lambda(2 + 7\lambda))}}{\lambda + 3\alpha\lambda}\right)^{-1/\alpha}$$

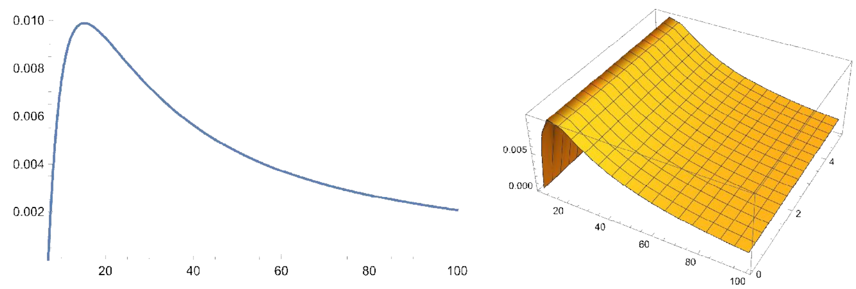

Fig1: The pdf for the GTPD.

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:

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

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

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

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

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

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:

$$\sigma^2 = \mu_2$$

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

Then the standard deviation will be

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

4- Reliability Analysis using G-Transmuted Pareto Distribution

The study of the characteristics of any probability distribution is fundamental to understanding the behavior of random variables. Among these characteristics are the survival function, the hazard rate function, the reversed hazard rate function, the cumulative hazard rate function, and the mean residual lifetime, all of 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 component within any system, with cumulative distribution function (CDF) F(x). The survival function, which represents the probability of not failing under specified conditions for a given period, is denoted by R(x) and is given by

$$R(x) = P(X > x) = 1 - F(x)$$
$$= -\frac{\left(\left(\frac{\beta}{x}\right)^{\alpha}\left(-1+\left(\frac{\beta}{x}\right)^{\alpha}\right)(x^{\alpha}\beta^{\alpha}(1-2\lambda)+x^{2\alpha}(-2+\lambda)+\beta^{2\alpha}\lambda)\left(2+\left(-1+3\left(\frac{\beta}{x}\right)^{\alpha}\right)\lambda\right)\right)}{\left((x^{\alpha}-\beta^{\alpha})(x^{\alpha}(-2+\lambda)-3\beta^{\alpha}\lambda)\right)} \quad (11)$$

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

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

The hazard rate function$h(x)$is unimodal as shown in Fig 2.

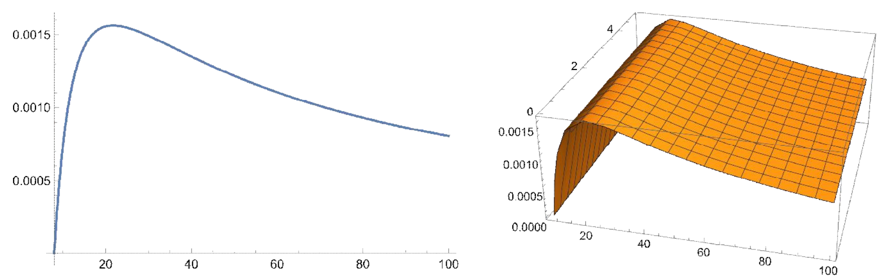

Fig2: The hazard rate function of the GTPD

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

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

5- Estimation of The Parameters

Parameter estimation is a fundamental aspect of statistical inference, allowing us to derive information about population parameters based on sample data. Among the various methods available, the method of maximum likelihood estimation (MLE) is particularly effective in providing point estimates and constructing confidence intervals for unknown parameters. Integer parameters often arise in discrete distributions and other applications. To address this, Dahiya [12] introduced an improved graphical method for estimating 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, offering valuable insights into 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 $X_1, X_2, \dots, X_n$ represent a sample of size n from the GTPD. Then, the likelihood function is given by

$$l(\alpha,\beta,\lambda) = \prod_{i=1}^{n} f(X_i\alpha,\beta,\lambda) = \prod_{i=1}^{n} \frac{-\alpha}{x_i} \prod_{i=1}^{n} (\frac{\beta}{x_i})^{\alpha}$$

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

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$\alpha$is to assume that it can take on a continuum of values and calculate derivatives with respect to the two parametersλ and α, then solve

$$U(\alpha,\beta,\lambda)_1 = \frac{\partial\mathcal{L}(\alpha,\beta,\lambda)}{\partial\alpha} = 0, U(\alpha,\beta,\lambda)_2 = \frac{\partial\mathcal{L}(\alpha,\beta,\lambda)}{\partial\beta} = 0 \quad \text{and } U(\alpha,\beta,\lambda)_3 = \frac{\partial\mathcal{L}(\alpha,\beta,\lambda)}{\partial\lambda} = 0$$

To solve theseequationswe 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 beingmultivariate normal. Thedigamma function$\varphi(k) = \frac{d}{dk}\ln\Gamma(k) = \frac{\Gamma'(k)}{\Gamma(k)}$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

$$U(\hat{\alpha}, \hat{\beta}, \hat{\lambda})_1 = 0 , U(\hat{\alpha}, \hat{\beta}, \hat{\lambda})_2 = 0 \text{and } U(\hat{\alpha}, \hat{\beta}, \hat{\lambda})_3.$$

6- SimulationStudy

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.

Table1. 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 using a real dataset and compared with several competing distributions, including the transmuted Lindley distribution, Monsef distribution, two-parameter Exponential distribution, and transmuted Monsef distribution. The estimated parameter values, log-likelihood statistic, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Kolmogorov–Smirnov (KS) statistic are presented in the following tables. The following data consist of the failure times for 23 communication transceivers from Modern Telecom Systems Information Technology Company (MTS-IT). The data are as follows:

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
Exponential	$\hat{\lambda}$ =1.013 $\hat{\gamma}$ = 1.21	56.15	116.316	118.587	116.916	116.887	116.916
Monsef	$\hat{\lambda}$ = 0.5	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	Cramer–von Mises	Waston	Waston2	Anderson
GTPD	0.09	0.02886	0.02878	5.2916	0.2059
Exponential	3.1109	97.909	25.577	105.57	---------
Monsef	0.73273	5.0187	1.29044	10.6738	42.6218
Trns Lindely	0.5066	2.0091	0.5748	7.5146	11.588
Trns Mons	0.5653	2.54114	0.70384	8.07896	16.2382

Conclusion

In conclusion, the G-Transmuted Pareto distribution represents a significant advancement in modeling asymmetric data. By incorporating additional parameters through the transmutation process, this distribution retains the foundational principles of the Pareto distribution while expanding its applicability. Additionally, real datasets and a simulation study are used to illustrate the advantages of the proposed distribution. It is evident that the GTPD distribution provides a better fit to the data compared to all other distributions, making it a competitive choice for the aforementioned dataset. Future research should focus on the empirical validation of the G-Transmuted distribution in various real-world datasets to assess its performance and robustness relative to other established distributions.

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. Arthur Charpentier, Emmanuel Flachaire, (2022). Pareto models for top incomes and wealth. Journal of Economic Inequality, 20 (1), pp. 1-25.

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, Article ID 9819200, 7 pages.

7. Muhanned Aslam, Nurulkamal, Mohd Azml. (2024). Examining Tail Index Estimators in New Pareto Distribution: Monte Carlo Simulations and Income Data Applications, Sains Malaysiana, 53(2), 461-476.

8. Ehab M. Almetwally, Hanan A. Haj Ahmad. (2020). A new generalization of the Pareto distribution and its applications, STATISTICS IN TRANSITION new series, Vol. 21, No. 5, pp. 61–84.

9. 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.

10. Mehmet Yilmaz, Sibel AcikKemaloglu. (2018). Two New Distribution Families Constructed by a Modified Rank. MATHEMATICAL SCIENCES AND

11. 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 KurtoticNormal Distributions.

12. 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.

13. 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

14. Dahiya, R. (1981). An improved method of estimating an integer-parameter by maximum likelihood. Am. Stat, 35(1):34–37.

15. Dahiya, R. (1986). Integer-parameter estimation in discrete distributions. Commun. Stat. - Theory Meth-ods, 15:709–725.

16. Olkin, I. ,Petkau, A. , Zidek, J. (1981). A comparison of n estimators for the binomial distribution. J. Am. Stat. Asso, 7, 685–689.

17. Miller, G. (1999). Maximum likelihood estimation for the Erlang integer parameter. Stat. Probab. Let, 43:335–341.