FLEXIBILITY OF GENERALIZED DISTRIBUTIONS

Omekam, Ifeyinwa Vivian and Adejumo, Adebowale Olushola

Department of Statistics, Faculty of Physical Sciences, University of Ilorin, Ilorin, Nigeria

*** Corresponding author, Email address: omekam.iv@unilorin.edu.ng**

Abstract

Some existing distributions are limited in shapes of Probability Density Function (PDF) and Hazard Function (HF) which constrains their use in analysis of certain types of data. Generalizing these distributions often deal with this constraints on usage by introducing flexibility. Generalized distributions were derived using the Generalized Pareto Distribution (GPD) as base distribution. Exponentiated GPDs called Lehmann Type II GPD (LIIGPD) and Lehmann Type I GPD (LIGPD) having an additional parameter each were obtained by applying Lehmann Alternative 1 (LA1) and Lehmann Alternative 2 (LA2) parameter induction methods respectively. Flexibility of generalized distributions was established by comparing the shapes of probability density and hazard functions of LIIGPD and LIGPD with those of the GPD. No new probability density or hazard shape was introduced by LIIGPD but the new shape introduced by LIGPD demonstrated flexibility of generalized distributions. Generalized distributions do not always introduce new density and hazard shapes but often improve flexibility of distributions.

Keywords: Generalized distributions; Flexibility, Generalized Pareto Distribution; Probability Density Function; Hazard Function; Lehmann Alternative 1; Lehmann Alternative 2.

1. Introduction

Limitations in characteristics of some existing probability distributions motivate generalization of distributions to improve flexibility and goodness of fit. The process of parameter induction into an existing probability distribution is one technique for generating generalized distributions. GPD was introduced by Pickands (1975). The GPD is limited in the shapes of its probability density and hazard functions and is a skewed distribution parameterized with a shape, scale, and in some forms, location parameter, therefore, a suitable base distribution for generalization.

Tahir and Nadarajah (2015) discussed some existing proven families of generalized distributions comprising of Marshall-Olkin extended family, exponentiated family, beta-generated family, Kumaraswamy-generalized family, and McDonald-generalized family. Amongst the generalized families of distributions studied by authors is the exponentiated family having an additional parameter obtainable from applying Lehmann Alternatives (Lehmann, 1953).

Let $f(x)$, $F(x)$, $s(x) = \bar{F}(x)$, $h(x)$, and $r(x)$ be respective denotations of the following functions of a base continuous random variable, X; PDF, Survival Function (SF), HF, and the Reversed Hazard Function (RHF). The corresponding functions of the exponentiated family of distributions for a new continuous random variable Y, having an additional parameter (c > 0) obtained by applying LA1 are as follows;

$$f_Y(x) = cf(x)F(x)^{c-1} \qquad (1.1)$$

$$F_Y(x) = (F(x))^c \tag{1.1a}$$

$$s(x) = \bar{F}_Y(x) = 1 - F(x)^c \tag{1.1b}$$

$$h_Y(x) = cf(x)F(x)^{c-1}(1 - F(x)^c)^{-1} \tag{1.1c}$$

The corresponding functions of the exponentiated family of distributions for a new continuous random variable Y, having an additional parameter (c > 0) obtained from the application of LA2 are as follows;

$$f_Y(x) = cf(x)\big(1 - F(x)\big)^{c-1} = cf(x)(\bar{F}(x))^{c-1} \tag{1.2}$$

$$F_Y(x) = 1 - (1 - F(x))^c = 1 - (\bar{F}(x))^c \tag{1.2a}$$

$$\bar{F}_Y(x) = [1 - F(x)]^c = [\bar{F}(x)]^c \tag{1.2b}$$

$$h_Y(x) = cf(x)\big([1 - F(x)\big)^{-1} = cf(x)[\bar{F}(x)]^{-1} \tag{1.2c}$$

Gupta and Kundu (2009) presented alternative interpretations to these two exponentiated families obtainable from applying LA1 and LA2, observing them to be Proportional Reversed Hazard Model (PRHM) and Proportional Hazard Model (PHM) respectively.

Adeyemi and Adebanji (2004) introduced a three-parameter Exponentiated Generalized Pareto (EGP) distribution as a generalization of the two-parameter base generalized Pareto distribution

(GPD). Matheson et al. (2017) described the flexibility of the Generalized Gamma distribution in terms of the relationship among its three quartiles. Another way of characterizing flexibility is a comparison of probability density and hazard shapes of distributions of interest. Alzaatreh et al. (2013) introduced Weibull-Pareto distribution, a special case of the Weibull-X family by extending the base Pareto distribution. The distribution improved flexibility by introducing unimodal and left skewed shapes. Ghitany et al. (2018) proposed Generalized Truncated Log-Gamma distribution (GTLG) which generalized the log-gamma as well as the Pareto distribution. The distribution was derived from a monotonic transformation of the classical Gamma distribution. GTLG added flexibility as was observed in the decreasing and unimodal probability density and hazard shapes.

The Marshal-Olkin Alpha Power Pareto distribution (MOAPP) introduced by Almetwally and Haj Ahmad (2020) using the Marshall Olkin and Alpha Power transformation methods is another extension of Pareto distribution. The PDF shape of MOAPP is either decreasing or upside down bathtub curve and the hazard shape is either decreasing or upside down curve where the curve is right skewed. MOAPP also improved flexibility. T- Pareto(Y) families (Hamed et al., 2018) belonging to families of generalized Pareto distributions is yet another flexible distribution obtained from the T-R(Y) framework. The PDFs of these generalized families can have unimodal shapes and can also be skewed to both left and right with heavy tails. Lee and Kim (2019) proposed Exponentiated GPD (exGPD) created via log-transform of the GPD variable, which has less sample variability. This Study targets describing flexibility of generalized distributions by assessing the probability density and hazard shapes of generalized distributions derived from parameter induction into the probability distribution of the GPD.

2. Materials and Methods

Generalized distributions introducing an additional shape parameter each was first derived by applying LA1 and LA2 parameter induction methods to obtain two exponentiated GPDs; LIGPD and LIIGPD respectively. These generalized distributions generalized the GPD used as base distribution. Subsequently, to establish flexibility of generalized distributions, visual presentations of the shapes of probability density and hazard functions of LIGPD and LIIGPD were compared with those of the parent distribution (GPD). Introduction of new probability density and hazard shapes established flexibility of distributions.

3. Results and Discussions

3.1 Exponentiated Generalized Pareto Distributions

The base distribution (GPD) has the following functions;

$$f(x) = \begin{cases} \sigma^{-1}(1+\frac{kx}{\sigma})^{-1-1/k}; & k \neq 0 \\ \sigma^{-1}\exp\left(-\frac{x}{\sigma}\right); & k = 0 \end{cases} \quad (3a)$$

$$F(x) = \begin{cases} 1-\left(1+\frac{kx}{\sigma}\right)^{-\frac{1}{k}}; & k \neq 0 \\ 1-\exp\left(-\frac{x}{\sigma}\right); & k = 0 \end{cases} \quad (3b)$$

$$\bar{F}(x) = \begin{cases} (1+kx/\sigma)^{-1/k}; & k \neq 0 \\ \exp\left(-\frac{x}{\sigma}\right); & k = 0 \end{cases} \quad (3c)$$

$x < \infty$ when $k \geq 0$, and $0 < x < -\sigma/k$ when $k < 0$

Given a generalized pareto distributed base random variable X with previously stated denotations of functions in section 1, let $f(y), F(y),\ s(y) = \bar{F}(y),\ h(y)$, and $r(y)$ be respective corresponding denotations of the PDF, CDF, SF, HF and RHF of another continuous random variable Y, having an additional parameter (c>0) derived from the applications of either LA2 or LA1 parameter induction methods to the functions of X.

3.1.1 Lehmann Type II GPD (LIIGPD)

Functions of the generalized distribution (LIIGPD) are obtained by applying LA2 parameter induction method to the functions of the base distribution (GPD) in section 3.1 as follows;

Substituting $f(x)$ in (3a) for $f(x)$ in (1.2) and $\bar{F}(x)$ in (3c) for $\bar{F}(x)$ in (1.2)

$$\mathrm{f(y)} = \begin{cases} c\sigma^{-1}(1+\mathrm{ky}/\sigma)^{-1-c/k}, & k \neq 0 \\ c\sigma^{-1}\left(exp(-y/\sigma)\right)^{c}, & k = 0 \end{cases}$$

(3.1) Substituting $\bar{F}(x)$ in (3c) for $\bar{F}(x)$ in (1.2a)

$$F(y) = \begin{cases} 1-\left(1+\frac{ky}{\sigma}\right)^{-\frac{c}{k}}, & k \neq 0 \\ 1-\left(exp\left(-\frac{y}{\sigma}\right)\right)^{c}, & k = 0 \end{cases}$$

(3.1a) Substituting $\bar{F}(x)$ in (3c) for $\bar{F}(x)$ in (1.2b)

$$\bar{F}(y) = \begin{cases} (1+ky/\sigma)^{-c/k}, & k \neq 0 \\ \left(exp(-y/\sigma)\right)^{c}, & k = 0 \end{cases}$$

(3.1b)

Substituting $f(x)$ in (3a) for $f(x)$ in (1.2c) and $\bar{F}(x)$ in (3c) for $\bar{F}(x)$ in (1.2c)

$$h(y) = \begin{cases} c\sigma^{-1}(1+ky/\sigma)^{-1}, & k \neq 0 \\ c\sigma^{-1}, & k = 0 \end{cases}$$

(3.1c)

3.1.2 Lehmann Type I GPD (LIGPD)

Functions of the generalized distribution (LIGPD) are obtained by applying LA1 parameter induction method to the functions of the base distribution (GPD) in section 3.1 as follows;

Substituting $f(x)$ in (3a) for $f(x)$ in (1.1) and $F(x)$ in (3b) for $F(x)$ in (1.1)

$$f(y) = \begin{cases} c\sigma^{-1}(1+ky/\sigma)^{-1-1/k}\left(1-(1+ky/\sigma)^{-1/k}\right)^{c-1}, & k \neq 0 \\ c\sigma^{-1}exp(-y/\sigma)\left(1-exp(-y/\sigma)\right)^{c-1}, & k = 0 \end{cases}$$

(3.2)

Substituting $F(x)$ in (3b) for $F(x)$ in (1.1a)

$$\mathrm{F(y)} = \begin{cases} \left(1-(1+ky/\sigma)^{-1/k}\right)^{c}, k \neq 0 \\ \left(1-exp(-y/\sigma)\right)^{c}, k = 0 \end{cases}$$

(3.2a)

Substituting $F(x)$ in (3b) for $F(x)$ in (1.1b)

$$\bar{F}(y) = \begin{cases} 1-\left(1-\left(1+\frac{ky}{\sigma}\right)^{-\frac{1}{k}}\right)^{c}, k \neq 0 \\ 1-\left(1-exp\left(-\frac{y}{\sigma}\right)\right)^{c}, k = 0 \end{cases}$$

(3.2b)

Substituting $f(x)$ in (3a) for $f(x)$ in (1.1c) and $F(x)$ in (3b) for $F(x)$ in (1.1c)

$$h(y) = \begin{cases} \frac{c\sigma^{-1}(1+ky/\sigma)^{-1-1/k}\left(1-(1+k\ \ /\sigma)^{-1/k}\right)^{c-1}}{1-\left(1-(1+ky/\sigma)^{-1/k}\right)^{c}}, k \neq 0 \\ \frac{c\sigma^{-1}exp(-y/\sigma)\left(1-exp(-y/\sigma)\right)^{c-1}}{1-\left(1-exp(-y/\sigma)\right)^{c}}, \quad k = 0 \end{cases}$$

(3.2c)

3.2 Flexibility of Exponentiated GPDS

To establish flexibility, PDF and HF shapes of exponentiated GPDs are compared with those of GPD.

3.2.1 SHAPE OF THE PDF OF LIIGPD

Figures 1 and 2 provide visuals of LIIGPD PDF shapes at $\sigma = 1$ and varying values of c and k.

3.2.2 SHAPE OF THE HF OF LIIGPD

Hazard shapes of LIIGPD are represented in figures 3 and 4 below;

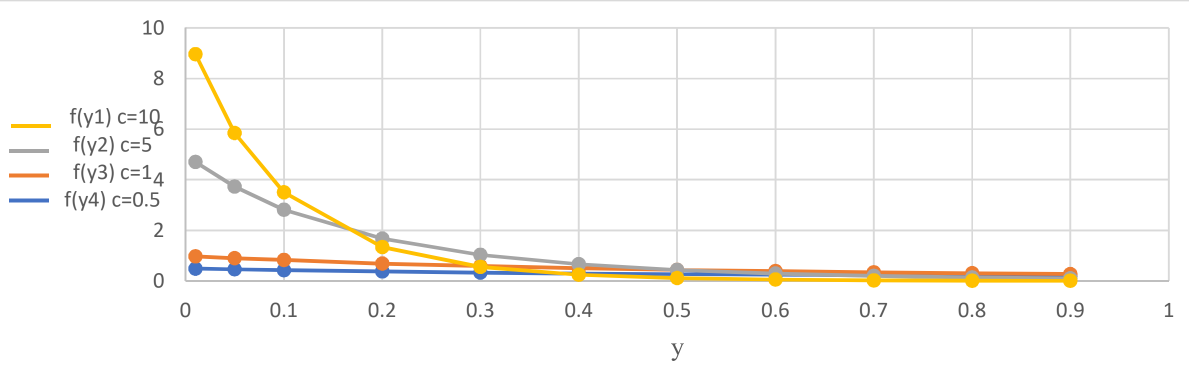

Figure 1: PDF Shapes of LIIGPD (k = 1,σ = 1, varying c)

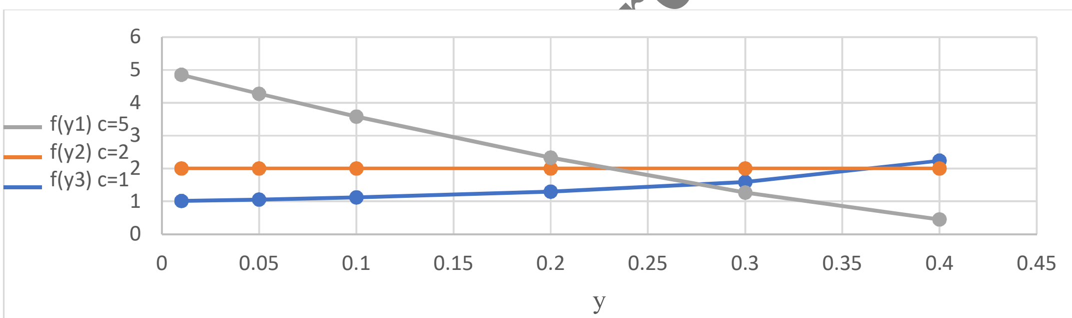

Figure 2: PDF Shapes of LIIGPD (k =-2,σ = 1, varying c)

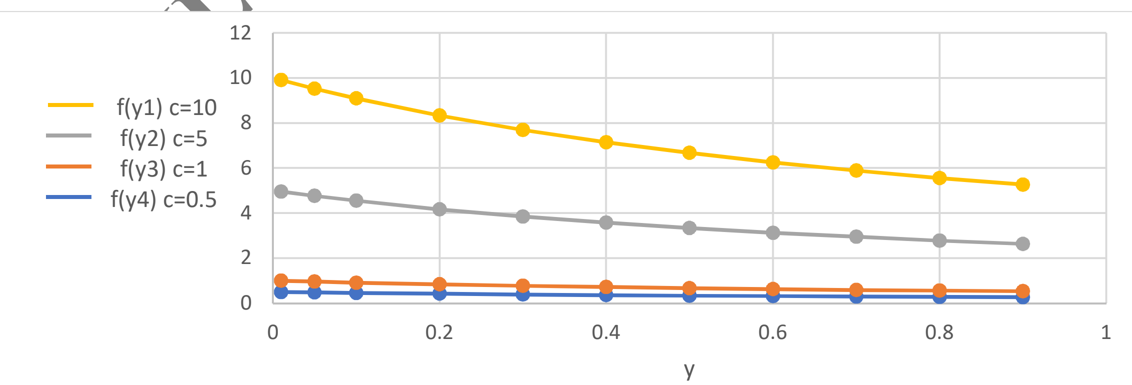

Figure 3: HF Shapes of LIIGPD (k = 1, σ = 1, varying c)

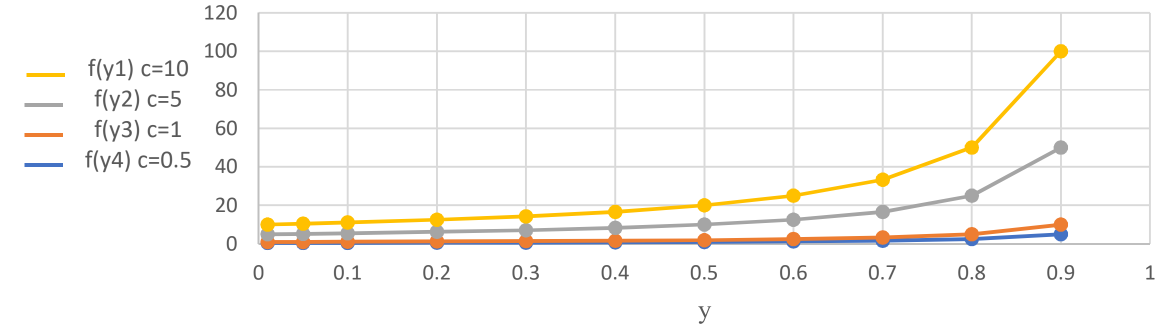

Figure 4: HF Shapes of LIIGPD (k = -1, σ = 1, varying c)

3.2.3 SHAPE OF THE PDF OF LIGPD

Figures 5, 6, and 7 show some PDF shapes of LIGPD for some specified values of distribution parameters. The LIGPD parallels GPD when c = 1, hence, at other values of c, figures show effect of introducing c on PDF shapes.

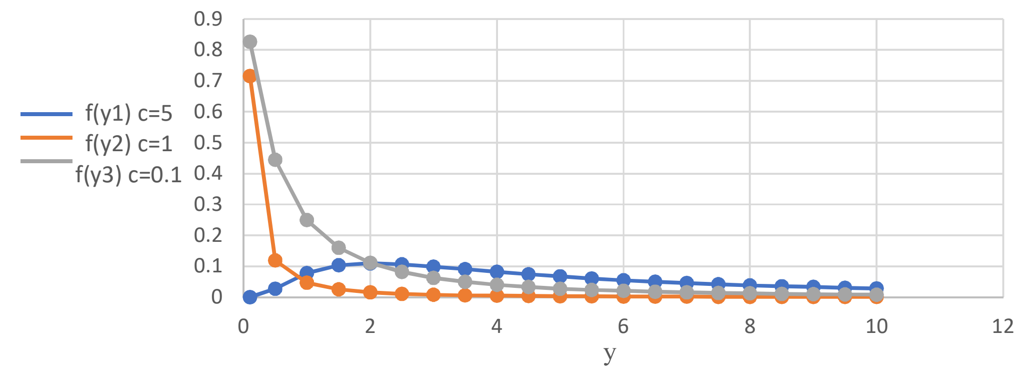

Figure 5: PDF shapes of LIGPD (K = 1,σ = 1, varying c)

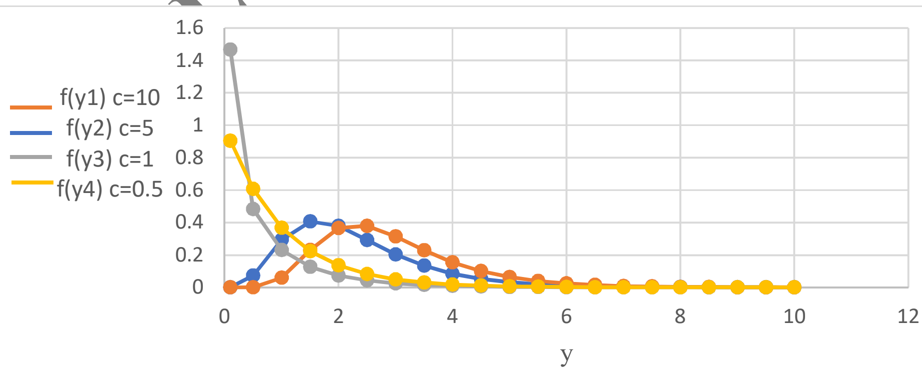

Figure 6: PDF Shapes of LIGPD (k = 0, σ = 1, varying c)

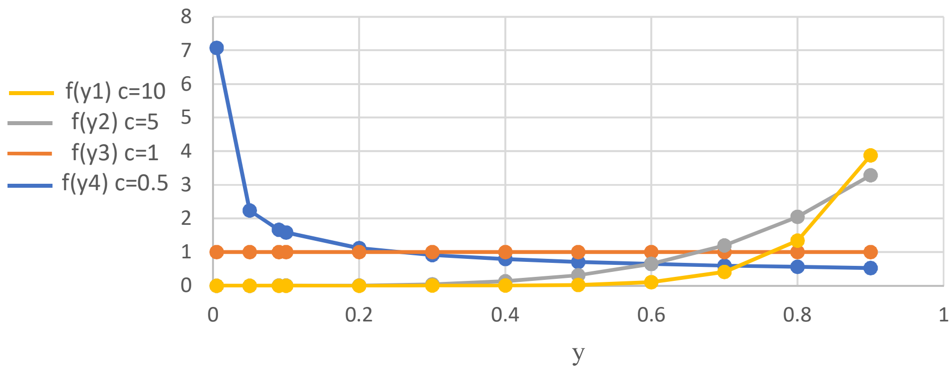

Figure 7: PDF Shapes of LIGPD (k = -1, σ = 1, varying c)

3.2.4 The shape of the HF of LIGPD

At some values of k, possible shapes of the hazard function are represented in figures 8, 9, and 10 below:

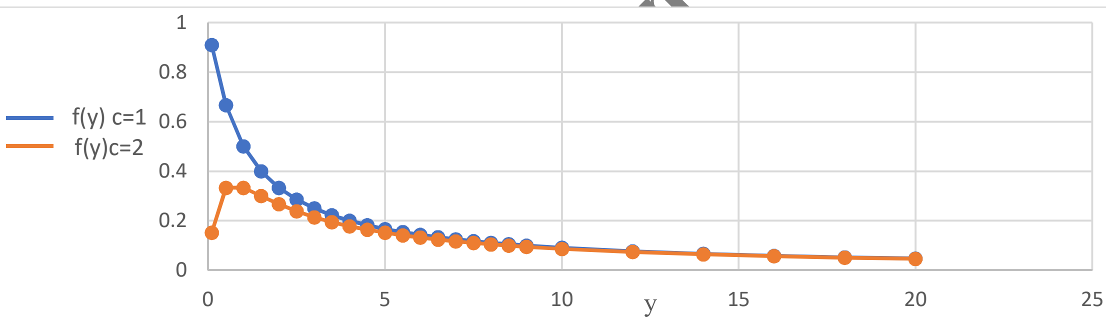

Figure 8: HF Shapes of LIGPD (k = 1, σ = 1, varying c)

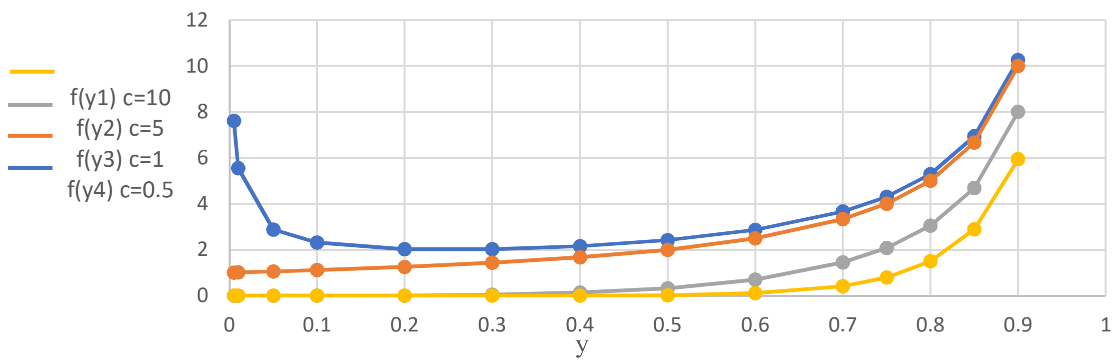

Figure 9: HF Shapes of LIGPD (k = -1, σ = 1, varying c)

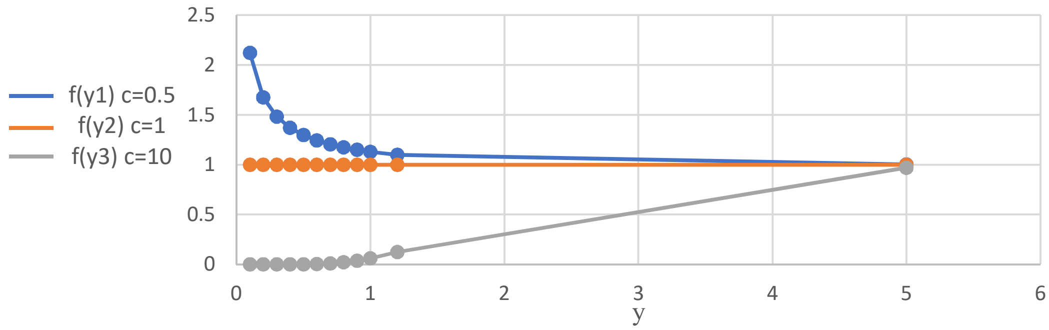

Figure 10: HF Shapes of LIGPD (k = 0, σ = 1, varying c)

3.3 Discussions

The LIIGPD parellels GPD when c=1. At other values of c and varying k values, LIIGPD with additional c did not introduce any new density or hazard shapes. Generalized distributions do not always introduce new shapes. Some generalizations of the Dirichlet distribution do not allow for a dependence and structure sufficiently richer than the Dirichlet (Ongaro & Migliorati, 2013). The hazard shapes of Generalized Gamma distribution (GG), an extension of the Gamma distribution (Stacy, 1962) include increasing, decreasing, bathtub, and unimodal shapes. Nevertheless, some generalizations of GG like Transmuted GG (Lucena et al., 2015), Kumaraswamy GG (Pascoa et al., 2011) and Modified GG (Mead et al., 2018) have similar hazard function with the base GG distributions. The GG is a flexible distribution for modeling many types of data, however, some of its extensions add very little to its capabilities and therefore GG is standard for parametric analysis of positive data considering the complexity of estimating these extended distribution (Matheson et al., 2017).

The two-parameter Weibull distribution, a generalization of the exponential distribution has increasing or decreasing hazard function. It is limited in its characteristics and is unable to show wide flexibility, However, density and hazard shapes of many other generalizations of the weibull distribution include the bathtub or inverted bathtub and unimodal shapes.

LIGPD reduces to GPD when c=1. At other values of c, probability and hazard shapes observed in figures 5-10 are the increasing, decreasing, constant and unimodal shapes. LIGPD introduced a new shape (unimodal) and improved flexibility by generalizing the GPD through LA1 parameter induction method. Ongaro and Migliorati (2013) proposed a new generalization of the Dirichlet distribution known as flexible Dirichlet exhibiting substantially greater flexibility in terms of shape of the density. Hazard rate of odd generalized exponential family introduced by Tahir et al. (2015) could have increasing, decreasing, J, reversed-J, bathtub and upside-down bathtub shapes. Most existing generalized distributions show wide flexibility by introduction of new shapes. Generalized distributions often improve flexibility of distributions.

4. CONCLUSION

Two modifications of Generalized Pareto Distribution belonging to the exponentiated class of distributions were derived for the purpose of establishing flexibility of generalized distributions. No new probability density or hazard shape was introduced by LIIGPD however, LIGPD derived from the application of LAI parameter induction method added flexibility by introducing new shape in the probability density and hazard functions. Generalized distributions do not always introduce new density and hazard shapes but often improves flexibility.

References

Adeyemi, S., & Adebanji, T. (2004). 'The Exponentiated Generalized Pareto Distribution". Ife Journal of Science, Vol. 6, pg. 127-133.

Almetwally, E.M. & Haj Ahmad, H.A. (2020). "A New Generalization of the Pareto Distribution and its Applications". Statistics in Transition New Series, Vol. 21, pg. 61-84.

Alzaatreh, A., Famoye, F., & Lee, C. (2013). "Weibull-Pareto Distribution and its Applications". Communications in Statistics-Theory and Methods, Vol. 42, pg. 1673-1691.

Ghitany, M.E., Gomez-Deniz, E., & Nadarajah S. (2018). "A New Generalization of the Pareto Distribution and its Application to Insurance data". Journal of Risk and Financial Management, Vol. 11, pg. 1-14.

Gupta, R.D., & Kundu, D. (2009). "Introduction of Shape/Skewness Parameter(s) in a Probability Distribution". Journal of Probability and Statistical Science, Vol. 7, pg. 153-171.

Hamed, D., Famoye, F., & Lee, C. (2018). "On Families of Generalized Pareto Distributions: Properties and Applications". Journal of Data Science, Vol. 16, pg. 377-396.

Lee, S. & Kim, J.H.T. (2018). "Exponentiated Generalized Pareto Distribution: Properties and Applications towards Extreme Value Theory". Communications in Statistics-Theory and Methods, Vol. 48, pg. 2014-2038.

Lehmann, E.L. (1953). "The Power of Rank Tests". The Annals of Mathematical Statistics, Vol. 24, pg. 23-43

Lucena, S.E.F., Silva, H.A., & Cordeiro, G.M. (2015). "The Transmuted Generalized Gamma Distribution: Properties and Application". Journal of Data Science, Vol. 13, pg. 187-206.

Matheson, M., Munoz A., & Cox C. (2017). "Describing the Flexibility of the Generalized Gamma and Related Distributions". Journal of Statistical Distributions and Applications, Vol. 4.

Mead, M., Nasssar, M.M., & Dey, S. (2018). "A generalization of Generalized Gamma Distributions". Pakistan Journal of Statistics and Operation Research, Vol.14, pg. 121-138.

Ongaro, A. & Migliorati, S. (2013). "A Generalization of the Dirichlet Distribution". Journal of Multivariate Analysis, Vol. 114, pg. 412-426.

Pascoa, M.R., Ortega, E., & Cordeiro, G.M. (2011). "The Kumaraswamy Generalized Gamma Distribution with Application in Survival Analysis". Statistical Methodology, Vol. 8, pg. 411-433.

Pickands, J. (1975). "Statistical Inference using Extreme Order Statistics". Annals of Statistics, Vol. 3, pg. 119-131.

Stacy, E.W. (1962). "A Generalization of the Gamma Distribution". Annals of Mathematical Statistics, Vol. 33, pg. 1187-1192.

Tahir, M.H., Cordeiro, G.M., Alizadeh, M., Mansoor, M., Zubair, M., & Hamedani, G.G. (2015). "The Odd Generalized Exponential Family of Distributions with Applications". Journal of Statistical Distributions and Applications, Vol. 2, pg. 1-28.

Tahir, M.H., & Nadarajah, S. (2015). "Parameter Induction in Continuous Univariate Distributions: Well-Established G Families". Anais da Academia Brasileire de Ciencias, Vol. 87, pg. 539-568.