Transactions of the Nigerian Association of Mathematical Physics
Volume 11, (January - June, 2020 Issue), **pp15 – 30**

STOCHASTIC MODELLING OF COVID-19 CLOSED CASES IN NIGERIA

Ekum M.I., Akinmoladun O.M. and Ogunsanya A.S.

Department of Mathematics and Statistics, Lagos State Polytechnic, Ikorodu, Lagos, Nigeria

Abstract

COVID-19 infected individuals are increasing in an alarming rate in Nigeria. Just in 80 days, Nigeria has recorded over 5,000 cases. The trend shows that in the next 80 days, this figure might triple if nothing is done to flattened the curve. Developing a stochastic model is very important in studying the trend and pattern of the spread of this novel virus. In this research, a stochastic model called the Gamma-Power function distribution is developed using the T-Power{Y} framework. Many properties of the stochastic model were investigated. The maximum likelihood estimation method was used to estimate the parameters of the proposed distribution and numerical investigation of the parameters which are not in closed form were developed. Simulation study was carried out to test the consistency of the parameter estimates. The pro-posed distribution was applied to two COVID-19 continuous data collected from Nigeria Centre for Disease Control (NCDC) and the results were compared with existing models. The results showed that the proposed distribution performed favourably when compared with other models.

Keywords: COVID-19 pandemic, Closed cases, Exponential family; Gamma-power function distribution, Stochastic model

1. Introduction

The rise in the number of COVID-19 laboratory confirmed cases in Nigeria is alarming and in no time might hit 4-digit. COVID-19 is a disease caused by a novel coronavirus, which index case started from Wuhan City of Hubei Province in China in December 2019. The virus was first isolated on 7th January 2020 in China. The total cases reported by World Health Organisation (WHO) from 31st December 2019 to 20th January 2020 was 44 cases [1]. The first imported cases outside China was reported in Thailand on 13th January, 2020. WHO [1] reported that 282 confirmed cases of COVID-19 from four countries including 278 cases from China, 2 from Thailand, 1 from Japan and 1 from Republic of Korea; all from the same region on 20th January, 2020.

On the 23rd January, 2020, United States of America (USA) reported her first case, which are now leading in the number of laboratory confirmed cases worldwide. Vietnam recorded the first local transmission on the 24th January, 2020. The first death outside China was recorded in Philippines on 13th February, 2020. Egypt reported its first confirmed case of COVID-19 on the 15th February, 2020 to become the first country in Africa to have COVID-19. On 22nd February, 2020, France recorded the first death in Europe [2]. On 26th February, 2020, there were more new cases reported outside of China than in China for the first time, since the onset of COVID-19 disease.

On the 28th February, 2020, Nigeria reported her index case of COVID-19 in West African Sub-region from an Italian immigrant, and on the same day, WHO increased the assessment of the risk of spread and risk of impact of COVID-19 from high to very high at the global level and it was declared a pandemic [2]. The second case in Nigeria was confirmed after 10 days, on the 9th March, 2020. Strict measures were not put in place by Nigerian government to checkmate and screen immigrants at her borders to stop imported cases. It was too late before Nigerian government started putting strict measures to contain the virus. As at the 80th day of the virus in Nigeria, the total laboratory confirmed cases has hit 5,959 cases. Since 28th April 2020, new cases reported on daily basis are above 100 (an average of 231 cases) as reported by Nigeria Centre for Disease Control (NCDC). If this trend continues, it means the total reported cases will hit 4 digits in June 2020 [3].

COVID-19 closed cases are the sum of COVID-19 induced deaths and recovered cases. When closed cases are subtracted from infected individuals, what is left is the active cases. It is necessary to have continuous COVID-19 data. Thus, COVID-19 induced deaths per 100 closed cases and time between two COVID-19 induced deaths are two continuous data of interest. So, developing a probability

Correspondence Author: Ekum M.I., Email: matekum@yahoo.com, Tel: +2348032437234

distribution that fit the COVID-19 continuous data is very important because existing ones might not be a good fit. Generating continuous probability distributions and fitting them to emerging data is trending in the field of probability theory and statistics [4]. The *T*-R{Y} framework developed by [5] and later redefined by [6] is a framework for generating convoluted distributions. It was developed from Beta-X by [7] and extended to T-X by [8].

Different authors in recent years have used the *T*-R{Y} framework to generate new distributions. Some of the most recent works already published in literature include Normal-exponential{logistic} by [9], Weibull-normal{log-logistic} by [4], Reduced Beta Skewed Laplace by [10], Odd Lomax-Exponential{log-logistic} by [11], exponentiated-exponential-dagum{lomax} by [12]. None of these authors have explored the possibility of using the power function distribution as a baseline distribution in generating convoluted distribution using the *T*-R{Y} framework, despite its simple functional form and flexibility.

Epidemiological Summary History of COVID-19 Cases in Nigeria

Date	States Infected				Cum. State Affected
28-Feb-2020	Lagos				1
09-Mar-2020	Ogun				2
18-Mar-2020	Ekiti				3
21-Mar-2020	FCT				4
22-Mar-2020	Oyo				5
23-Mar-2020	Edo				6
24-Mar-2020	Bauchi				7
25-Mar-2020	Osun	Rivers			9
28-Mar-2020	Kaduna	Benue	Enugu		12
01-Apr-2020	Akwa Ibom				13
03-Apr-2020	Ondo				14
06-Apr-2020	Kwara				15
07-Apr-2020	Delta	Katsina			17
10-Apr-2020	Anambra	Niger			19
11-Apr-2020	Kano				20
19-Apr-2020	Jigawa	Borno			22
20-Apr-2020	Abia	Gombe	Sokoto		25
22-Apr-2020	Adamawa				26
23-Apr-2020	Plateau				27
24-Apr-2020	Zamfara				28
25-Apr-2020	Imo				29
26-Apr-2020	Bayelsa	Ebonyi	Kebbi	Taraba	33
28-Apr-2020	Nasarawa				34
29-Apr-2020	Yobe				35

Figure 1: Spread of COVID-19 Nationwide

The diagram in Figure 1 shows how COVID-19 spread from the first imported case in Lagos by an Italian immigrant to his contact in Ogun State, which was the second case and thereafter spread to other 33 states including Federal Capital Territory (FCT) Abuja. On 27th February, a 44-year old Italian citizen was diagnosed of COVID-19 in Lagos State. The case was the first to be reported in Nigeria since the first confirmed case in China. The case came through Murtala Muhammed International Airport, Lagos at 10pm on 24th February 2020 aboard Turkish airline from Milan, Italy. He moved from Lagos to his company site in Ogun state on 25th February. On 26th February, he presented at the staff clinic in Ogun and there was high index of suspicion by the managing physician. He was referred to IDH Lagos and COVID-19 was confirmed on 27th February [3].

The second case was reported on 9th March in Ogun State, a contact from the index case. Initially it was Lagos and Ogun States that were susceptible but by 18th March 2020, susceptible states increased to 3, with Ekiti State included. The virus continues to spread both from imported cases and local transmission. As at 29th April, 2020, 35 states of the federation have reported cases of COVID-19, making the 35 states to be susceptible. Only Cross River and Kogi States do not have any COVID-19 case as at 17th May 2020.

The map of Nigeria showing the total coverage of COVID-19 as at 17th March 2020 is depicted in Figure 2.

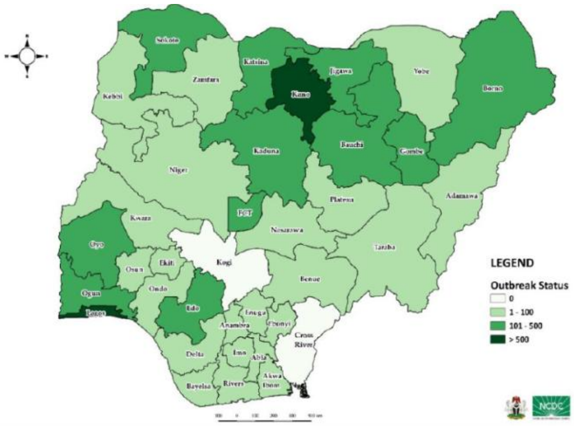

Figure 2: Map showing spread of COVID-19 Nationwide

Recently, the number of COVID-19 induced deaths have increased significantly from 14 deaths reported on the 23rd January, 2020 to 264,109 deaths reported on 6th May, 2020. Initially, as at 23rd January, the time between two successive COVID-19 induced deaths was 180 minutes (3 hours), but as at 6th May, the time between two successive COVID-19 induced deaths was 0.2367 minutes (14.2 seconds). This implies that every 14 seconds, someone dies of COVID-19. So, developing a probability distribution that fit the COVID-19 induced death rate is very important. Also, the number of time between two successive COVID-19 induced deaths is a continuous data that needs a continuous probability distribution to fit it. This distribution is distinct and has its peculiar features (see [13]). The new continuous distribution would be tested using two COVID-19 datasets, namely, data on COVID-19 induced death rate and time between two failures (COVID-19 induced death).

The remaining parts of the paper are unfolded as follows, in section 2, we present the materials and methods in which we derived the Gamma-Power function{log-logistic} distribution and its functions. Section 3 includes the results and discussion, while in Section 4, we made concluding remarks.

2 Materials and Methods

2.1 T-power{Y} Family of distribution

The T-R{Y} framework by [6] provides a convolution method for generalizing the R distribution using the T distribution with the Y quantile function. Combining these three distributions will produce a convoluted distribution with greater flexibility.

Let R be a random variable that follows a power function distribution defined by [14] with cdf, $F_R(x) = \frac{x^k}{\lambda^k}$ and pdf, $f_R(x) = \frac{kx^{k-1}}{\lambda^k}$. The cdf of T-Power function{Y} or simply T-P{Y} family of distribution is given by

$$F_X(x) = \int_a^{Q_Y\left(\frac{x^k}{\lambda^k}\right)} f_T(t)\,dt = F_T\left[Q_Y\left(\frac{x^k}{\lambda^k}\right)\right]$$

and the corresponding pdf is given by

$$f_X(x) = \frac{k}{\lambda^k} x^{k-1} \frac{f_T\left[Q_Y\left(\frac{x^k}{\lambda^k}\right)\right]}{f_Y\left[Q_Y\left(\frac{x^k}{\lambda^k}\right)\right]}$$

where T and Y are two random variables having the same support, except if Y is uniformly distributed. $F_T(t)$ is the cdf of T and $Q_Y(y)$ is the quantile function of Y. Also, $f_T(t)$ and $f_Y(y)$ are the pdfs of T and Y respectively. See [13] for the proof.

Remark 1. *If X is T -Power function{Y} distributed, then it follows that*
(i) $X = \lambda[F_Y(T)]^{1/k}$, *in distribution*
(ii) $Q_X(p) = \lambda\{F_Y[Q_T(p)]\}^{1/k}$,
(iii) If $T = Y$ *in distribution, then* $X = Power\ function(k)$ *in distribution and*
(iv) If $Y = Power\ function(k)$ *in distribution, then* $X = T$ *in distribution.*

The power function distribution has various forms. A simpler form is a one parameter power function distribution defined by [15] given by $f(x) = kx^{k-1}$. See [16] for other functional forms of the power function distribution.

2.2 *T*-power{log-logistic}

Lemma 1. *Let T and R be two random variables with cdfs $F_T(t)$ and $F_R(r)$ respectively. The cdf of a convoluted distribution derived from T and R is given by*

$$F_X(x) = F_T\left[\frac{F_R(x)}{S_R(x)}\right]$$

where $S_R(r)$ is the survival function of random variable R. The corresponding pdf is given by

$$f_X(x) = \frac{f_R(x)}{[S_R(x)]^2} f_T\left[\frac{F_R(x)}{S_R(x)}\right]$$

Proof.

Let T, R and Y be three random variables with cdfs $F_T(t)$, $F_R(r)$ and $F_Y(y)$ respectively. By definition, the cdf of a convoluted random variable X defined by [6] is given by

$$\mathrm{F_X}(x) = \int_{a}^{\mathrm{Q_Y}[\mathrm{F_R}(x)]} f_T(t)dt = P\{T \le \mathrm{Q_Y}[\mathrm{F_R}(x)]\} = F_T\{\mathrm{Q_Y}[\mathrm{F_R}(x)]\} \quad (1)$$

Let Y follows the log-logistic distribution with the standard quantile function $Q_Y(p)$ given by

$$Q_Y(p) = \frac{p}{1-p} \quad (2)$$

where $p = F_R(x)$. Let R be a random variable that follows the power function distribution with shape parameter k and scale parameter λ. The cdf of R is given by

$$F_R(x) = \left(\frac{x}{\lambda}\right)^k \quad (3)$$

Substitute equation (2) and (3) into (1) to have

$$F_X(x) = F_T\left[\frac{F_R(x)}{1-F_R(x)}\right] \quad (4)$$

But $S_R(x) = 1 - F_R(x)$. So that

$$F_X(x) = F_T\left[\frac{F_R(x)}{S_R(x)}\right] \quad (5)$$

By differentiating equation (5) with respect to x will give the pdf given by

$$f_X(x) = \frac{f_R(x)}{[S_R(x)]^2} f_T\left[\frac{F_R(x)}{S_R(x)}\right] \quad (6)$$

Thus, equations (5) and (6) complete the proof.

It should be noted that equation (6) can also be written as

$$f_X(x) = \frac{h_R(x)}{S_R(x)} f_T\left[\frac{F_R(x)}{S_R(x)}\right]$$

The ratio $\left[\frac{F_R(x)}{S_R(x)}\right]$ is called the odd-ratio of R random variable and $h_R(x)$ is its hazard function, which confirms that $f_X(x)$ is a weighted hazard function of the baseline distribution R. This is one of the major features of the T-R{Y} family of distributions defined by [13].

Lemma 2. *Let T $[0, \infty)$ be any random variable with cdf $F_T(t)$ and pdf $f_T(t)$, then the cdf and pdf of T - Power{log-logistic} family are respectively given by*

$$F_X(x) = F_T\left(\frac{x^k}{\lambda^k - x^k}\right)$$

and

$$f_X(x) = \frac{k\lambda^k x^{k-1}}{(\lambda^k - x^k)^2} f_T\left(\frac{x^k}{\lambda^k - x^k}\right), k, \lambda > 0; 0 \le x \le \lambda$$

Proof

Let R be a random variable that follows the Power function distribution with parameters k and λ, the cdf of R is given by

$$F_R(x) = \left(\frac{x}{\lambda}\right)^k \quad (7)$$

and the survival function is given by

$$S_R(x) = 1 - \left(\frac{x}{\lambda}\right)^k \quad (8)$$

Substitute equations (7) and (8) into (5) to have

$$F_X(x) = F_T\left(\frac{x^k}{\lambda^k - x^k}\right) \quad (9)$$

and differentiating equation (9) with respect to x will give the pdf of the T -Power{log-logistic} family as

$$f_X(x) = \frac{k\lambda^k x^{k-1}}{(\lambda^k - x^k)^2} f_T\left(\frac{x^k}{\lambda^k - x^k}\right), k, \lambda > 0; 0 \le x \le \lambda \quad (10)$$

Thus, equations (9) and (10) complete the proof

2.3 Gamma-power{log-logistic} distribution (GPLD)

The Gamma-power{log-logistic} distribution (GPLD), which is a convolution of the gamma, power function and log-logistics distributions using a *T*-R{Y} was first developed by [13]. Here, the power function is the baseline distribution. Many other distributions can be derived by replacing gamma distribution with any other distribution and/or replacing log-logistic with other distribution, making sure that *T* and *Y* have the same support. For other *T* and *Y* already used, see [13]. The Weibull-Power function distribution by [16] is a special case of this family. It can be called Weibull-power{Cauchy} distribution. In this section, we will show how the *T* -Power{log-logistic} family was derived and in particular, the Gamma-Power{log-logistic} distribution. All of its important functions are also derived in this section.

2.4 Derivation of GPLD

One may be interested in modelling fluctuations in variables such as income of employees in a firm, expenditure of households in a community, strength of material produced by a machine, quantity of liquid in a reservoir, quantity of oil spilled in bbl, etc. Gamma distribution plays an important role in reliability theory and life testing for these mentioned variables. These variables are not just non-negative but also have known upper bounds. The power function distribution provides such a great flexibility with a real number λ as the upper bound and zero as the lower bound.

2.5 Cumulative distribution function of Gamma-Power{log-logistic}

One of the most important ways of characterizing a probability distribution is through its cumulative distribution function (cdf).

Theorem 1. *Let X be a random variable that follows the Gamma-Power{log-logistic} distribution with parameters α, β, k and λ, then the cdf of X defined on a closed interval* [0, λ] *is given by.*

$$F_X(x) = \frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]; \alpha, \beta, k, \lambda > 0, 0 \le x \le \lambda$$

where $\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]$ *is an incomplete gamma function*

Proof

Let *T* follows the gamma distribution with shape parameter α and scale parameter β, that is

$$f_T(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1}\exp(-\beta t), \alpha, \beta > 0, t \ge 0$$

The cdf of *T* is given by

$$F_T(x) = \frac{1}{\Gamma(\alpha)}\gamma(\beta t) \quad (11)$$

Substitute equation (11) into (9) to have the cdf of GPLD

$$F_X(x) = \frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]; \alpha, \beta, k, \lambda > 0, 0 \le x \le \lambda \quad (12)$$

Equation (12) completes the proof. Equation (12) is the cdf of the proposed gamma-power function log-logistic distribution (GPLD).

2.6 Probability Density Function of GPLD

The probability density function (pdf) of GPLD is derived by differentiating the cdf in equation (12) with respect to *x*. The pdf of the proposed Gamma-Power{log-logistic} distribution is therefore given by

$$f_X(x) = \frac{k\lambda^k \beta^\alpha x^{\alpha k-1}}{\Gamma(\alpha)(\lambda^k - x^k)^{\alpha+1}} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]; \alpha, \beta, k, \lambda > 0, 0 \le x \le \lambda \quad (13)$$

From now hence forth the newly proposed Gamma-Power{log-logistic} distribution will be regarded as GPLD.

Figure 3 illustrates some possible shapes of the density function of the GPLD, for selected parameter values. The density function can take various forms depending on the parameter values. It is obvious that the GPLD has higher flexibility than the gamma and the power function distributions, because of the additional parameters, which allow for a high degree of flexibility of the GPLD. It shows that for different parameter values *α, β, k* and for a constant λ, GPLD can be positively or negatively skewed and can be leptokurtic or platykurtic. The pdf has various shapes as displayed in Figure 3. So, the new distribution would be very useful in many practical situations for modelling positive real data sets.

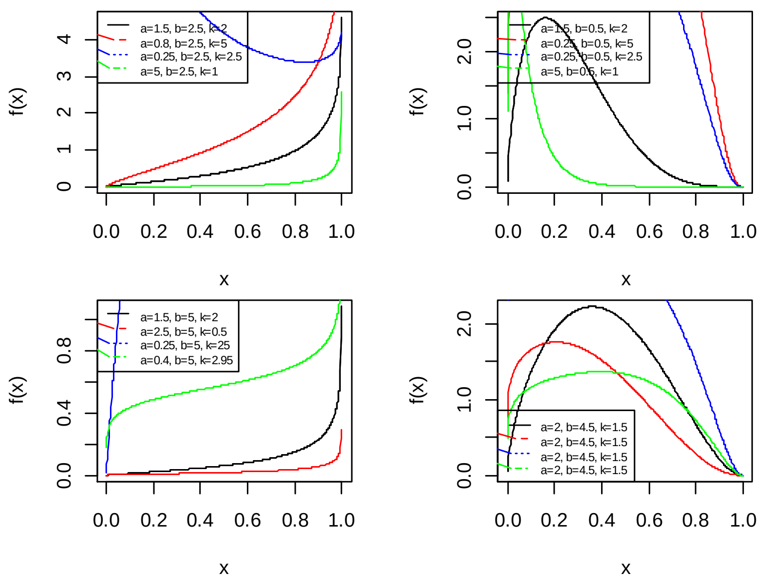

Figure 3: The pdf of GPLD Distribution for $\lambda = 1$. Source: [13]

2.7 Survival Function of GPLD

Given the cdf in equation (12), the survival function of GPLD is given by

$$S_X(x) = 1 - \frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right] \tag{14}$$

2.8 Hazard Function of GPLD

Theorem 2. *The hazard function of a random variable X that follows a GPLD with parameters α, β, k, λ exist and it is given by*

$$h_X(x) = \frac{k\lambda^k\beta^\alpha x^{\alpha k-1} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]}{\left(\lambda^k - x^k\right)^{\alpha+1}\left\{\Gamma(\alpha) - \gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right\}}.$$

Proof

By definition, the hazard function of a random variable X is given by

$$h_X(x) = \frac{f_X(x)}{S_X(x)}. \tag{15}$$

Substitute the pdf in equation (13) and the survival function in equation (14) into equation (15), we derive the $h_X(x)$ of GPLD as

$$h_X(x) = \frac{\frac{k\lambda^k\beta^\alpha x^{\alpha k-1}}{\Gamma(\alpha)\left(\lambda^k - x^k\right)^{\alpha+1}} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]}{1 - \frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]}.$$

Solve further to arrive at

$$h_X(x) = \frac{k\lambda^k\beta^\alpha x^{\alpha k-1} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]}{\left(\lambda^k - x^k\right)^{\alpha+1}\left\{\Gamma(\alpha) - \gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right\}}. \tag{16}$$

Equation (16) completes the proof.

2.9 Cumulative Hazard Function of GPLD

By definition, the cumulative hazard function of a random variable X is given by

$$H_X(x) = -log[S_X(x)] \tag{17}$$

Let X be random variable that follows a GPLD with survival function given in equation (14). The cumulative hazard function, $H_X(x)$ of GPLD is derived by substituting equation (14) into (17) to have

$$H_X(x) = -ln\left\{1 - \frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right\} \tag{18}$$

Equation (18) is the cumulative hazard function of GPLD.

2.10 Reverse Hazard Function of GPLD

Theorem 3. *The revered hazard function of a random variable X that follows a GPLD with parameters α, β, k, λ exist and it is given by*

$$\tau_X(x) = \frac{k\lambda^k\beta^\alpha x^{\alpha k-1} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]}{\left(\lambda^k - x^k\right)^{\alpha+1}\left\{\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right),\alpha\right]\right\}}$$

Proof.

By definition, the reversed hazard function of a random variable *X* is given by

$$\tau_X(x) = \frac{f_X(x)}{S_X(x)}. \qquad (19)$$

Substitute the pdf in equation (13) and the cdf in equation (12) into equation (19), we derive the $\tau_X(x)$ of GPLD as

$$\tau_X(x) = \frac{k\lambda^k\beta^\alpha x^{\alpha k-1} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]}{\left(\lambda^k - x^k\right)^{\alpha+1}\left\{\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right),\alpha\right]\right\}}. \qquad (20)$$

2.11 Quantile Function of GPLD

Definition 1. *The quantile function of a random variable X is the value at which the probability of the random variable is less than or equal to the given probability. It is the inverse function of the cdf and it is defined as*

$$Q_X(x) = F^{-1}(x). \qquad (21)$$

Recall the pdf of *T*-Power{log-logistic} given in equation (10).

Lemma 3. *Let T be a random variable with pdf* $f_T(x)$, *then random variate,* $X = \lambda\left(\frac{T}{1+T}\right)^{1/k}$ *follows T-Power{log-logistic} family of distribution in equation (10), provided T is supported on the interval 0 to ∞, i.e., T ∈ [0, ∞). The log-logistic parameters, scale = shape = 1, where k and λ are the parameters from the power function distribution.*

Proof.

It is easy to see the result from Remark 1i.

Lemma 4. *It follows from Lemma 1 that the quantile functions of T-Power function{log-logistic} distribution is given by*

$$Q_X(p) = \lambda\left[\frac{Q_T(p)}{1+Q_T(p)}\right]^{1/k} \qquad (22)$$

Proof.

It is easy to see the result from Remark 1ii.

Theorem 4. *If T (α, β) follows a gamma distribution with parameters α and β, then the quantile of GPLD with parameters α, β, k, λ is given by*
where QT(_,_) is the quantile function of gamma distribution with parameters _, _; and k and _ are the parameters from the power function distribution.

$$Q_X(p) = \lambda\left(\frac{Q_{T(\alpha,\beta)}(p)}{1+Q_{T(\alpha,\beta)}(p)}\right)^{1/k}$$

where $Q_{T\ (\alpha,\beta)}$ *is the quantile function of gamma distribution with parameters α, β; and k and λ are the parameters from the power function distribution.*

Proof.

Following Remarks (1i) and (1ii), and Lemma (3) and (4). Just substitute the quantile function of gamma distribution into Lemma (4) to have

$$Q_X(p) = \lambda\left(\frac{Q_{T(\alpha,\beta)}(p)}{1+Q_{T(\alpha,\beta)}(p)}\right)^{1/k} \qquad (23)$$

where k is a shape parameter and λ is a scale parameter from the power function distribution. It is easy to generate *T* using R codes. The rgamma generates random values of gamma distribution, *T*. Then, use the transformation in Theorem (4) with known α and β to generate random variates that follow GPLD.

The quantile function returns the value *x* such that

$F(x) = P(X \leq x) = p$

The quantile function of a particular distribution is used in Monte Carlo method to simulate random variates that follows such distribution. The quantile function can be used to partition a distribution into different non-overlapping continuous sections. We can determine the quartiles, octiles, deciles and percentiles using the quantile function.

2.12 Asymptotes of GPLD

2.12.1 Vertical Asymptotes of GPLD

This is the value x will approach for $f_X(x)$ to approach 1.

$$\lim_{x\to a} f_X(x) = \infty \quad (24)$$

For GPLD, the functions $f_X(x)$ and $h_X(x)$ will be undefined if

$$\lambda^k - x^k = 0 \quad (25)$$

and $x = \lambda$. Thus, the vertical asymptote of $f_X(x)$ is given by

$$\lim_{x\to\lambda} f_X(x) = \infty. \quad (26)$$

and the vertical asymptote of $hX(x)$ is given by

$$\lim_{x\to\lambda} h_X(x) = \infty. \quad (27)$$

So, equations (26) and (27) are the vertical asymptotes of the pdf and hazard function of GPLD respectively.

2.12.2 Horizontal Asymptotes of GPLD

If $f_X(x)$ and $h_X(x)$ are the pdf and hazard functions of GPLD distribution. Then the horizontal asymptotes are horizontal lines that the functions approach as $x \to \infty$.

$$\lim_{x\to\infty} f_X(x) = a \quad (28)$$

The horizontal asymptote of $f_X(x)$ is given by

$$\lim_{x\to\infty} f_X(x) = 0 \quad (29)$$

and the horizontal asymptote of $h_X(x)$ is given by

$$\lim_{x\to\infty} h_X(x) = 0 \quad (30)$$

So, equations (29) and (30) are the horizontal asymptotes of the pdf and hazard function of GPLD respectively. Note that x is bounded above and so this may not exist in the real sense of GPLD.

2.13 Moment of GPLD

The moment of a distribution is a very important function for deriving the mean of the distribution. The series expansion of the pdf of GPLD is given by

$$f_X(x) = \frac{k\beta^\alpha}{\Gamma(\alpha)} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^i (\alpha+i+j)! x^{k(\alpha+i+j)-1}}{i! j! \lambda^{k(\alpha+i+j)}} \quad (31)$$

If $i = j = 0$, the series expansion of the pdf of GPLD given in equation (31) will reduce to

$$f_X(x) = \frac{\alpha! k\beta^\alpha}{\Gamma(\alpha)\lambda^{\alpha k}} x^{\alpha k-1} \quad (32)$$

See the complete proof in [13].

The rth moment of GPLD using the linear expansion pdf in equation (31) is given by

$$E(X^r) = \frac{k\lambda^r \beta^\alpha}{\Gamma(\alpha)} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^i (\alpha+i+j)!}{i! j! [k(\alpha+i+j)+r]} \quad (33)$$

If $r = 1$, we have the mean of GPLD given by

$$E(X) = \frac{k\lambda\beta^\alpha}{\Gamma(\alpha)} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^i (\alpha+i+j)!}{i! j! [k(\alpha+i+j)+1]} \quad (34)$$

If $i = j = 0$, the mean of GPLD becomes

$$E(X) = \frac{(\alpha-1)! \alpha k \lambda \beta^\alpha}{\Gamma(\alpha)(\alpha k+1)} \quad (35)$$

and the variance is given by

$$Var(X) = \frac{\alpha! k\lambda^2 \beta^\alpha}{\Gamma(\alpha)(\alpha k+2)} \left| 1 - \frac{\alpha! k\beta^\alpha (\alpha k+2)}{(\alpha k+1)} \right| \quad (36)$$

2.14 Gini Index of GPLD

The Gini index or coefficient measures the inequality among values of a distribution. A Gini index of zero expresses perfect equality, where all values are the same (for example, where everyone has the same income). A Gini index of one (or 100%) expresses maximal inequality among values. For instance, if a large number of people, where only one person has all the income or consumption, and all others have none, then the Gini index will be very nearly one [17].

Definition 2. *Let X be a continuous random variable with survival function $S_X(x)$, and has a finite mean μ, the Gini index, G, which is zero for all negative values of X is given by*

$$G = 1 - \frac{1}{\mu} \int_0^{\infty} [S_X(x)]^2 \, dx \quad (37)$$

Theorem 5. *Let X be a random variable that follows a GPLD with survival function $S_X(x)$ and with μ, the Gini index, G of X is only dependent on parameter λ and it is given by*

$$G = \left|1 - \left(\lambda - \lambda^2 + \frac{\lambda^3}{3}\right)\right|$$

Proof. For GPLD, the Gini index is derived by

$$G = 1 - \frac{1}{\mu}\int_0^\lambda [S_X(x)]^2\, dx \qquad (38)$$

where $S_X(x) = 1 - F_X(x)$ given in equation (14) and μ is the mean of GPLD. So that equation (38) can become
Put equations (40) and (41) into (39) to have

$$G = 1 - \frac{1}{\mu}\int_0^\lambda [1 - F_X(x)]^2\, dx. \qquad (39)$$

Let

$$w = F_X(x) \qquad (40)$$

So,

$$\frac{dw}{dx} = f_X(x)$$

$$dx = \frac{dw}{f_X(x)} \qquad (41)$$

Put (40) and (41) into (39) to have

$$\mu(1 - G) = \int_0^\lambda (1 - w)^2 \frac{dw}{f_X(x)}$$

$$\mu(1 - G) f_X(x) = \int_0^\lambda (1 - w)^2\, dw = \int_0^\lambda (1 - 2w + w^2)^2\, dw$$

$$\mu(1 - G) f_X(x) = \lambda - \lambda^2 + \frac{\lambda^3}{3} \qquad (42)$$

Continuous sum of both sides of equation (42) on the bounded interval [0, λ], that is, integrate for side with respect to *x*.

$$\mu(1 - G)\int_0^\lambda f_X(x)\, dx = \int_0^\lambda \left(\lambda - \lambda^2 + \frac{\lambda^3}{3}\right) dx \qquad (43)$$

But we know that

$$\int_0^\lambda f_X(x)\, dx = 1$$

So, that equation (43) becomes

$$\mu(1 - G) = \left(\lambda - \lambda^2 + \frac{\lambda^3}{3}\right) x \qquad (44)$$

Sum through equation (44) to have

$$n\mu(1 - G) = \left(\lambda - \lambda^2 + \frac{\lambda^3}{3}\right)\sum_{i=1}^{n} x_i$$

$$\mu(1 - G) = \left(\lambda - \lambda^2 + \frac{\lambda^3}{3}\right)\frac{1}{n}\sum_{i=1}^{n} x_i = \left(\lambda - \lambda^2 + \frac{\lambda^3}{3}\right)\bar{x}$$

where *n* is the number of observation, that is x_i, $i = 1, 2, ..., n$.

Taking expectation of both sides gives

$$\mu(1 - G) = \left(\lambda - \lambda^2 + \frac{\lambda^3}{3}\right) E(\bar{x}) = \left(\lambda - \lambda^2 + \frac{\lambda^3}{3}\right)\mu \qquad (45)$$

Solving for *G* from equation (45), we have the Gini index of GPLD as

$$G = \left|1 - \left(\lambda - \lambda^2 + \frac{\lambda^3}{3}\right)\right| = \left|1 - \lambda\left(1 - \lambda + \frac{\lambda^2}{3}\right)\right| \qquad (46)$$

Equation (46) completes the proof and *G* depends only on parameter λ. When λ = 1, the Gini index for GPLD is a constant value and it is equal to 2/3 or 66.7%. The world Gini index for 2008 was 70% and that of 2013 was 65%. So, the Gini index of GPLD for λ = 1 can be used as an approximation to world Gini index.

2.15 Order Statistics of GPLD

2.15.1 1st Order Statistics of GPLD

Lemma 5. *Let $X_1, X_2, ..., X_n$ be a random sample from the GPLD distribution and $X_{(1)}, X_{(2)}, . . . , X_{(n)}$, such that, $X_{(1)} \leq X_{(2)} \leq .... . . \leq X_{(n)}$, are order statistics obtained from the sample. Then the pdf $f_{X1}(x)$ of the 1st order statistics, $X_{(1)}$ is given by*

$$f_{X_1}(x) = \frac{nk\lambda^k\beta^\alpha x^{\alpha k-1}}{\Gamma(\alpha)\left(\lambda^k - x^k\right)^{\alpha+1}}\left\{1 - \frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right\}^{n-1} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]$$

Proof.

By definition, 1*st* order statistic of a random variable *X* is given by

$$f_{X_1}(x) = -\frac{d}{dx}\prod_{i=1}^{n}[1 - F_X(x)]^n = n[1 - F_X(x)]^{n-1}f_X(x) = n[S_X(x)]^n h_X(x) \quad (47)$$

Substitute the cdf and pdf of GPLD in equations (12) and (13) into (47) to have the 1[st] order statistic of GPLD derived as

$$f_{X_1}(x) = \frac{nk\lambda^k\beta^\alpha x^{\alpha k-1}}{\Gamma(\alpha)\left(\lambda^k - x^k\right)^{\alpha+1}}\left\{1 - \frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right\}^{n-1} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right] \quad (48)$$

Equation (48) completes the proof.

2.15.2 n^{th} Order Statistics of GPLD

Lemma 6. *Let $X_1, X_2, ..., X_n$ be a random sample from the GPLD distribution and $X_{(1)}, X_{(2)}, . . . , X_{(n)}$, such that, $X_{(1)} \le X_{(2)} \le .... . . \le X_{(n)}$, are order statistics obtained from the sample. Then the pdf $f_{X_n}(x)$ of the n^{th} order statistics, $X_{(n)}$ is given by*

$$f_{X_n}(x) = \frac{nk\lambda^k\beta^\alpha x^{\alpha k-1}}{\Gamma(\alpha)\left(\lambda^k - x^k\right)^{\alpha+1}}\left\{\frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right\}^{n-1} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]$$

Proof.

By definition, n*th* order statistic of a random variable *X* is given by

$$f_{X_n}(x) = -\frac{d}{dx}\prod_{i=1}^{n}[1 - F_X(x)]^n = n[F_X(x)]^{n-1}f_X(x) = n[F_X(x)]^n \tau_X(x) \quad (49)$$

Substitute the cdf and pdf of GPLD in equations (12) and (13) into (49) to have the n^{th} order statistic of GPLD derived as

$$f_{X_n}(x) = n\left[\frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right]^{n-1}\frac{k\lambda^k\beta^\alpha x^{\alpha k-1}}{\Gamma(\alpha)\left(\lambda^k - x^k\right)^{\alpha+1}} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]. \quad (50)$$

Solve equation (50) further to have

$$f_{X_n}(x) = \frac{nk\lambda^k\beta^\alpha x^{\alpha k-1}}{\Gamma(\alpha)\left(\lambda^k - x^k\right)^{\alpha+1}}\left\{\frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right\}^{n-1} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right] \quad (51)$$

Equation (51) completes the proof.

2.15.3 General Order Statistics of GPLD

Lemma 7. *Let $X_{(1)}, X_{(2)}, ..., X_{(n)}$ denote the order statistics of a random sample that follows GPLD distribution, $X_1, X_2, ..., X_n$, from a continuous population with cdf, $F_X(x)$ and pdf $f_X(x)$. Then the pdf $f_{X(j)}(x)$ of GPLD is given by*

$$f_{X_{(j)}}(x) = \frac{n!\,k\lambda^k\beta^\alpha x^{\alpha k-1}}{(j-1)!\,(n-j)!\,\Gamma(\alpha)\left(\lambda^k - x^k\right)^{\alpha+1}} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]\Lambda$$

where

$$\Lambda = \left\{\frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right\}^{j-1}\left\{1 - \frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right\}^{n-j}$$

Proof.

By definition, j*th* order statistic of a random variable *X* is given by

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!}f_X(x)[F_X(x)]^{j-1}[F_X(x)]^{n-j} \quad (52)$$

Substitute the cdf and pdf of GPLD in equations (12) and (13) into (52) to have the j^{th} order statistic of GPLD derived as

$$f_{X_{(j)}}(x) = \frac{n!k\lambda^k\beta^\alpha x^{\alpha k-1}}{(j-1)!(n-j)!\Gamma(\alpha)\left(\lambda^k - x^k\right)^{\alpha+1}} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]\Lambda \quad (53)$$

where

$$\Lambda = \left\{\frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right\}^{j-1}\left\{1 - \frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]\right\}^{n-j} \quad (54)$$

Equation (53) completes the proof.

2.16 Odd Ratio of GPLD

Definition 3. *Let X be a random with cdf $F_X(x)$. The odd ratio is given by*

$$OR = \frac{F_X(x)}{1 - F_X(x)} \quad (55)$$

The odd-ratio is a non-decreasing function, which made it possible for the proofing of Lemma (1). It is a very useful function in survival analysis. Its first derivative is the inverse of its survival function.

Lemma 8. *Let X be a random variable that follows the GPLD with cdf $F_X(x)$, its odd-ratio exists and it is given by*

$$OR = \frac{\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]}{\Gamma(\alpha) - \gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]}$$

Proof.

Substitute the cdf of GPLD into equation (55) of Definition (5) to have.

$$OR = \frac{\frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]}{1 - \frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]} \quad (56)$$

Solve equation (56) further to have the odd-ratio of GPLD given by

$$OR = \frac{\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]}{\Gamma(\alpha) - \gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]} \quad (57)$$

Equation (57) completes the proof.

2.17 Likelihood Ratio of GPLD

The likelihood ratio *lr* is the ratio of two probabilities.

Lemma 9. *If $f_X(x)$ be the pdf of GPLD with parameters $(\alpha, \beta, k_1, \lambda)$* and *$g_X(x)$ be the pdf of GPLD with parameters $(\alpha, \beta, k_2, \lambda)$, such that $k_1 > k_2$. Then*

$$\lim_{x \to 0} l(x) = \log\left(\frac{k_1 \lambda^{\alpha k_2}}{k_2 \lambda^{\alpha k_2}}\right) \geq 0.$$

Proof.

The likelihood ratio of $f_X(x)$ and $g_X(x)$ is defined as

$$lr = \frac{f_X(x)}{g_X(x)} \quad (58)$$

Substitute the pdf of GPLD into equation (58) to have

$$lr = \frac{k_1}{k_2}\left(\frac{\lambda^{k_2} - x^{k_2}}{\lambda^{k_1} - x^{k_1}}\right)^{\alpha+1} \lambda^{k_1 - k_2} x^{\alpha(k_1 - k_2)}\, exp\left\{-\beta\left[\left(\frac{x^{k_1}}{\lambda^{k_1} - x^{k_1}}\right) - \left(\frac{x^{k_2}}{\lambda^{k_2} - x^{k_2}}\right)\right]\right\} \quad (59)$$

Take the log of equation (59) to have

$$loglr = \log k_1 - logk_2 + (\alpha+1)log\left(\frac{\lambda^{k_2} - x^{k_2}}{\lambda^{k_1} - x^{k_1}}\right) + (k_1 - k_2)log\lambda + [\alpha(k_1 - k_2)]logx - \beta\left[\left(\frac{x^{k_1}}{\lambda^{k_1} - x^{k_1}}\right) - \left(\frac{x^{k_2}}{\lambda^{k_2} - x^{k_2}}\right)\right] \quad (60)$$

Take the limit of *loglr* as *x* tends to zero, where $l(x) = loglr$

$$\lim_{x \to 0} l(x) = \log k_1 - logk_2 + (\alpha+1)(k_2 - k_1)log\lambda + (k_1 - k_2)log\lambda \quad (61)$$

Solve equation (61) further to arrive at

$$\lim_{x \to 0} l(x) = \log\left(\frac{k_1 \lambda^{\alpha k_2}}{k_2 \lambda^{\alpha k_2}}\right) \geq 0. \quad (62)$$

Equation (62) completes the proof.

2.18 Monotone Likelihood Ratio Property of GPLD

In statistics, the monotone likelihood ratio property (MLRP) is a property of the ratio of two probability density functions (pdfs).

Definition 4. *Let $f_X(x)$ and $g_X(x)$ be the pdfs of two distributions with respect to x with this property, if for every $x_2 > x_1$*

$$\frac{f_X(x_2)}{g_X(x_2)} = \frac{f_X(x_1)}{g_X(x_1)} \quad (63)$$

that is, if the ratio is non-decreasing in the argument x.

If the functions are first-differentiable, the property is sometimes stated as

$$\frac{\partial}{\partial x}\left[\frac{f_X(x)}{g_X(x)}\right] \geq 0 \quad (64)$$

Then $f_X(x)$ and $g_X(x)$ have the MLRP in x.

For two distributions that satisfy the definition with respect to some argument *x*, we say they have the MLRP in *x*. For a family of distributions that all satisfy the definition with respect to some statistic *T* (*X*), we say they have the MLR in *T* (*X*), where *T* (*X*) is a sufficient statistic. The MLRP is used to represent a data-generating process that enjoys a straightforward relationship between the magnitude of some observed variable and the distribution it draws from. If $f_X(x)$ satisfies the MLRP with respect to $g_X(x)$, the higher the observed value *x*, the more likely it was drawn from distribution *f* rather than *g*. As usual for monotonic relationships, the likelihood ratio's monotonicity comes in handy in statistics, particularly when using maximum-likelihood estimation. Also, distribution families with MLR have a number of well-behaved stochastic properties, such as first-order stochastic dominance and increasing hazard ratios. Monotone likelihoods are used in several areas of statistical theory, including point estimation and hypothesis testing, as well as in probability models.

Theorem 6. *If $x_1, x_2, ..., x_n$ is a random sample from gamma distribution with parameters α, β having pdf $g_X(x)$ and $x_{(1)}, x_{(2)}, . . . , x_{(n)}$, such that, $x_{(1)} \leq x_{(2)} \leq .... . . \leq x_{(n)}$, are order statistics obtained from the*

sample, then $f_X(x)$ and $g_X(x)$ have the monotone likelihood ratio property in x, where $f_X(x)$ is the pdf of GPLD, such that
$\frac{\partial}{\partial x}\left[\frac{f_X(x)}{g_X(x)}\right] \geq 0$

Proof.
Recall the pdf of gamma distribution given as

$$g_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)\,;\ \alpha, \beta > 0, x \geq 0 \tag{65}$$

and the pdf, $f_X(x)$ of GPLD given in equation (13). The likelihood ratio of $f_X(x)$ and $g_X(x)$ is given as

$$lr = \frac{f_X(x)}{g_X(x)} = \frac{\frac{k\lambda^k \beta^\alpha x^{\alpha k-1}}{\Gamma(\alpha)\left(\lambda^k - x^k\right)^{\alpha+1}} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right]}{\frac{1}{\Gamma(\alpha)}\gamma\left[\beta\left(\frac{x^k}{\lambda^k - x^k}\right), \alpha\right]} \tag{66}$$

If we solve equation (66) further, we have

$$lr = \frac{k\lambda^k \beta^\alpha x^{\alpha(k-1)}}{\left(\lambda^k - x^k\right)^{\alpha+1}} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right] \tag{67}$$

To complete the proof, from Definition (4), we have to show that *lr* is a non-decreasing function of *x* for all $x \geq 0$ for the parameter space ($\alpha, \beta, k, \lambda$). To do this, we have to take the first derivative of *lr* with respect to x.

$$\frac{\partial lr}{\partial x} = \frac{\partial}{\partial x} \frac{k\lambda^k \beta^\alpha x^{\alpha(k-1)}}{\left(\lambda^k - x^k\right)^{\alpha+1}} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right] \tag{68}$$

$$\frac{\partial lr}{\partial x} = \frac{k\lambda^k \beta^\alpha x^{\alpha(k-1)}}{\left(\lambda^k - x^k\right)^{\alpha+1}} exp\left[-\beta\left(\frac{x^k}{\lambda^k - x^k}\right)\right](A - B) \tag{69}$$

where

$$A = \frac{k(\alpha+1)x^k - \alpha(k-1)\left(\lambda^k - x^k\right)}{x\left(\lambda^k - x^k\right)} \tag{70}$$

and

$$B = \beta\left[\frac{kx^k}{x\left(\lambda^k - x^k\right)} + \frac{k\left(x^k\right)^2}{x\left(\lambda^k - x^k\right)^2} - 1\right] \tag{71}$$

Note that $x \geq 0$ and $max(x) < \lambda$, so the term $(\lambda^k - x^k) > 0$. Also, $A, B > 0$ and $A > B$. Thus, the derivative in (69) is greater than zero.

2.19 Maximum Likelihood Estimation (MLE)

Recall the pdf in (13), the MLE is derived thus.
The likelihood of (13) gives

$$L(\alpha, \beta, k, \lambda) = \frac{k^n \lambda^{kn} \beta^{\alpha n}}{[\Gamma(\alpha)]^n} \prod_{i=1}^n \frac{x_i^{\alpha k-1}}{\left(\lambda^k - x_i^k\right)^{\alpha+1}} exp\left[-\beta \sum_{i=1}^n \left(\frac{x_i^k}{\lambda^k - x_i^k}\right)\right] \tag{72}$$

Take the log to have

$$l = logL(\alpha, \beta, k, \lambda) = nlogk + nklog\lambda + anlog\beta - nlog\Gamma(\alpha) + (\alpha k - 1)\sum_{i=1}^n logx_i - (\alpha+1)\sum_{i=1}^n log\left(\lambda^k - x_i^k\right) - \beta \sum_{i=1}^n \left(\frac{x_i^k}{\lambda^k - x_i^k}\right) \tag{73}$$

The maximum likelihood estimation parameters of the GPLD are given by differentiating ` partially with respect to each of the parameters and equating the results to zero and solve for each parameter.

$$\frac{\partial l}{\partial \alpha} = -nln\beta + k\sum_{i=1}^n lnx_i - \sum_{i=1}^n \ln\left(\lambda^k - x_i^k\right) \tag{74}$$

$$\frac{\partial l}{\partial \beta} = \frac{\alpha n}{\beta} - \sum_{i=1}^n \left(\frac{x_i^k}{\lambda^k - x_i^k}\right) \tag{75}$$

$$\frac{\partial l}{\partial k} = \frac{n}{k} + nln\lambda + \alpha\sum_{i=1}^n lnx_i - k(\alpha+1)\sum_{x=0}^n \frac{x_i^{k-1}}{\left(\lambda^k - x_i^k\right)} - \beta\lambda^k k \sum_{i=1}^n \frac{x_i^{k-1}}{\left(\lambda^k - x_i^k\right)^2} \tag{76}$$

$$\frac{\partial l}{\partial \lambda} = \frac{\partial}{\partial \lambda} nlogk + nklog\lambda + anlog\beta - nlog\Gamma(\alpha) + (\alpha k - 1)\sum_{i=1}^n logx_i - (\alpha+1)\sum_{i=1}^n log\left(\lambda^k - x_i^k\right) - \beta\sum_{i=1}^n \left(\frac{x_i^k}{\lambda^k - x_i^k}\right) \tag{77}$$

Equate (75) to zero and solve for β to obtain

$$\hat{\beta} = \frac{\hat{\alpha} n}{\sum_{i=1}^{n}\left(\frac{x_i^k}{\lambda^k - x_i^k}\right)} \tag{78}$$

Equate (74) to zero gives

$$\hat{\beta} = \exp\left[\frac{1}{n}\sum_{i=1}^{n}\log\left(\lambda^k - x_i^k\right) - \frac{k}{n}\sum_{i=1}^{n} log x_i\right] \tag{79}$$

Insert (78) into (79) to have

$$\frac{\hat{\alpha} n}{\sum_{i=1}^{n}\left(\frac{x_i^k}{\lambda^k - x_i^k}\right)} = \exp\left[\frac{1}{n}\sum_{i=1}^{n}\log\left(\lambda^k - x_i^k\right) - \frac{k}{n}\sum_{i=1}^{n} log x_i\right] \tag{80}$$

Solve for $\hat{\alpha}$ in (80) gives

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i^k}{\lambda^k - x_i^k}\right)\exp\left[\frac{1}{n}\sum_{i=1}^{n}\log\left(\lambda^k - x_i^k\right) - \frac{k}{n}\sum_{i=1}^{n} log x_i\right] \tag{81}$$

The equations obtained by setting the partial derivatives ` with respect to k to zero is not in closed form and the values of the parameter k must be found by using numerical methods. The estimate of k, denoted by $\hat{\hat{k}}$ is estimated using Newton-Raphson numerical method. The R package (maxLik or optim) can also be used to estimate this parameter [18].

The parameter λ is estimated by

$$\hat{\lambda} = \max(x_i) + \sigma_{\bar{x}} \tag{82}$$

or λ is the least upper bound

$$\hat{\lambda} = \text{Sup}(x_i) \tag{83}$$

where $\sigma_{\bar{x}}$ is the standard error of X and x is the value of a random variable X, that is $x \in X$.

3 Results and Discussion

3.1 Simulation Study

The maximum likelihood method for estimating the performance of GPLD is evaluated using Monte Carlo simulation for a total of eighteen parameter combinations with 1000 replications. Three different sample sizes n = 20, 200 and 1000 were considered, for small, medium and large samples respectively. The actual values, maximum likelihood estimates, absolute bias and standard errors of the parameter estimates were presented in Table 1. From Table 1, it is noted that the maximum likelihood parameter estimates performed well for estimating the distribution parameters. As the sample size increases, the absolute bias and standard error decrease.

Table 1. Actual values, Average Estimates and Standard errors for various parameter values

	Actual values				Estimates				Absolute Bias				Std. Error			
n	α	β	k	λ	$\hat{\alpha}$	$\hat{\beta}$	$\hat{k}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{k}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{k}$	$\hat{\lambda}$
20	0.5	0.5	1.0	5.0	0.39	0.42	1.04	3.70	0.11	0.08	0.04	1.30	0.1	0.157	0.534	0.065
	0.5	1.0	2.0	10.0	1.01	3.15	2.30	9.22	0.51	2.15	0.30	0.78	0.282	1.124	0.974	0.89
	1.0	0.5	1.0	5.0	1.24	0.65	1.03	5.47	0.24	0.15	0.03	0.47	0.353	0.227	0.541	0.0113
	1.0	1.0	2.0	10.0	1.24	1.01	2.39	10.94	0.24	0.01	0.39	0.94	0.354	0.353	1.39	0.0439
	1.5	0.5	1.0	5.0	1.41	0.65	0.90	5.31	0.09	0.15	0.10	0.31	0.405	0.223	0.371	0.0049
	1.5	1.0	2.0	10.0	1.41	1.30	2.43	10.37	0.09	0.30	0.43	0.37	0.405	0.445	0.94	0.0068
200	0.5	0.5	1.0	5.0	0.46	0.52	0.92	4.54	0.04	0.02	0.08	0.46	0.038	0.069	0.41	0.0023
	0.5	1.0	2.0	10.0	0.46	0.95	0.92	11.51	0.04	0.05	1.08	1.51	0.038	0.126	0.874	0.2123
	1.0	0.5	1.0	5.0	1.10	0.51	1.28	5.79	0.10	0.01	0.28	0.79	0.082	0.056	0.341	0.0031
	1.0	1.0	2.0	10.0	1.06	1.01	2.04	11.33	0.06	0.01	0.04	1.33	0.094	0.114	0.628	0.0089
	1.5	0.5	1.0	5.0	1.39	0.50	1.24	5.69	0.11	0.00	0.24	0.69	0.126	0.054	0.663	0.0024
	1.5	1.0	2.0	10.0	1.39	0.99	2.28	11.02	0.11	0.01	0.28	1.02	0.126	0.109	0.94	0.0052
1000	0.5	0.5	1.0	5.0	0.48	0.51	0.98	5.85	0.02	0.01	0.02	0.85	0.018	0.03	0.086	0.0007
	0.5	1.0	2.0	10.0	0.51	1.04	1.86	11.60	0.01	0.04	0.14	1.60	0.019	0.06	0.475	0.0026
	1.0	0.5	1.0	5.0	1.03	0.53	1.26	5.79	0.03	0.03	0.26	0.79	0.041	0.027	0.197	0.0006
	1.0	1.0	2.0	10.0	1.00	1.04	2.02	11.53	0.00	0.04	0.02	1.53	0.04	0.053	0.447	0.0023
	1.5	0.5	1.0	5.0	1.58	0.56	0.98	5.67	0.08	0.06	0.02	0.67	0.065	0.027	0.208	0.0004
	1.5	1.0	2.0	10.0	1.58	1.11	2.20	11.00	0.08	0.11	0.20	1.00	0.065	0.054	0.94	0.001

3.2 Consistency of the Parameter Estimates

Table 1 shows that the estimates of parameters are consistent as shown by the values of absolute biases and standard errors. The absolute biases and standard errors converge to zero as the sample size, n increases from 20 to 200 to 1000.

3.3 Application

In this section, two applications to real data sets were provided to illustrate the uses and importance of the proposed Gamma-Power function distribution (GPLD), especially in medicine and survival analysis. The distribution parameters were estimated by the method of maximum likelihood and three goodness-of-fit criteria and three goodness-of-fit statistics were evaluated to compare the flexibility of the GPLD distribution with other known existing distributions: Gamma, Weibull and Normal distributions. The goodness-of-fit criteria used are log-likelihood, Akaike information criterion (AIC), Bayesian information criterion, while the goodness-of-fit statistics include Kolmogorov-Smirnov statistic (K-S), Cramer-von Mises statistic (W) and Anderson-Darling statistic (A). These criteria and statistics were computed to compare the fitted distributions to the datasets. The required computations were carried out in the R-language [18].

3.3.1 Application 1: Deaths per 100 COVID-19 concluded cases

The first real data set represents Deaths per 100 COVID-19 concluded cases. COVID-19 concluded cases are those individuals that were once infected but have either died or recovered. If these concluded cases are subtracted from the infected cases, we have the active cases. The data used here was derived by

$$Y_t = \frac{D_t}{D_t + R_t} x 100 \quad 84$$

where D_t is the deaths at time t, R_t is the deaths at time t. The denominator $D_t + R_t$ is the number of concluded cases. Y_t represents the number of COVID-19 deaths out of every 100 concluded cases per day.

Table 2. COVID-19 induced deaths per 100 closed cases in Nigeria

100.0	16.7	28.6	11.1	10.0
12.5	20.0	11.1	3.3	4.0
36.4	22.2	33.3	5.3	25.0
100.0	100.0	8.3	17.6	22.7
50.0	11.9	36.8	23.8	33.3
20.0	20.0	7.2	8.6	5.6
11.4	14.3	31.3	6.8	9.5
5.1	2.7	2.8	3.2	4.7

Table 2 shows the dataset, also depicted in Figure 4 and shows that there is a gap in the histogram with positive skewness (2.2221) and kurtosis (7.3552). Table 3 displays the maximum likelihood estimates of the parameters with their corresponding standard errors in brackets. Table 3 shows all the parameters of the GPLD distribution and other distributions.

Table 3. MLE of Parameters and Standard Errors for Death per 100 COVID-19 concluded cases

Distribution	Parameter Estimates (Standard errors)			
GPLD	$\hat{\alpha}$	$\hat{\beta}$	$\hat{k}$	$\hat{\lambda}$
	1.512	1.628	0.494	108
	(0.190)	(0.242)	(0.045)	(1.547)
Gamma	$\hat{\alpha}$	$\hat{\beta}$		
	1.728	0.088		
	(0.219)	(0.013)		
Weibull	$\hat{k}$	$\hat{\lambda}$		
	1.340	21.445		
	(0.097)	(1.650)		
Normal	$\hat{\mu}$	$\hat{\sigma}$		
	19.614	15.780		
	(1.540)	(1.089)		

Table 4. Goodness-of-fit Criteria and Statistics for Death per 100 COVID-19 concluded cases

	Criteria			Statistics		
Distribution	**-logL**	**AIC**	**BIC**	**K-S**	**W**	**A**
GPLD	92.4112	188.8223	194.1303	0.1788	0.7077	4.3374
Gamma	409.4749	822.9499	828.2578	0.0722	0.0978	0.7123
Weibull	410.6393	825.2785	830.5864	0.0840	0.1106	0.8286
Normal	438.6585	881.3169	886.6248	0.1474	0.5766	4.0785

Table 4 clearly shows that the GPLD distribution provides the best fit to the first data. Figures 5 also support the results in favour of the GPLD model.

3.3.2 Application 2: Time (in minutes) between successive COVID-19 induced deaths

The COVID-19 data are reported on daily basis. The data given is the number of COVID-19 induced deaths per 24 hours. So, to get the time (in hours) between successive COVID-19 induced deaths, we have $\left(\frac{24}{D_t}\right)$.

Table 5. Time (in hours) between two successive COVID-19 induced deaths in Nigeria

24.00	24.00	12.00	24.00	24.00
24.00	8.00	24.00	24.00	24.00
6.00	12.00	12.00	24.00	8.00
8.00	8.00	24.00	8.00	4.80
6.00	3.43	3.43	2.40	1.41
12.00	4.00	4.80	4.80	6.00
2.40	2.18	1.60	2.67	4.00
4.00	8.00	6.00	4.80	4.00

The skewness and kurtosis of the data are 0.8293 and 2.0329 respectively. The data is positively skewed and peaked as depicted in Figure 5.

Table 6. Maximum likelihood estimates of parameters and standard errors for minutes between COVID-19 induced deaths

Distribution	Parameter Estimates (Standard errors)			
GPLD	$\hat{\alpha}$	$\hat{\beta}$	$\hat{k}$	$\hat{\lambda}$
	5.038	3.541	0.226	203
	(0.676)	(0.498)	(0.042)	(2.322)
Gamma	$\hat{\alpha}$	$\hat{\beta}$		
	0.398	0.032		
	(0.044)	(0.006)		
Weibull	$\hat{k}$	$\hat{\lambda}$		
	0.527	6.895		
	(0.040)	(1.351)		
Normal	$\hat{\mu}$	$\hat{\sigma}$		
	12.627	23.679		
	(2.311)	(1.634)		

The maximum likelihood estimates of the parameters of the fitted distributions with their corresponding standard errors in brackets are given in Table 4. All the parameters of the GPLD are significant at the 5% significance level. The GPLD provides a better fit to the yarn data than the WPC, PC, Gamma and Power function distributions as shown in Table 5.

Table 7. Goodness-of-fit Statistics and Criteria for minutes between COVID-19 induced deaths

	Criteria			Statistics		
Distribution	-logL	AIC	BIC	K-S	W	A
GPLD	93.8537	191.7074	197.0153	0.1014	0.1463	1.1643
Gamma	326.2997	656.5995	661.9074	0.1875	0.7863	4.7969
Weibull	322.9465	649.8930	655.2009	0.1815	0.8012	4.8109
Normal	481.2709	966.5418	971.8497	0.2994	2.4548	13.2806

Table 5 shows that GPLD AIC, A and W approach zero faster than that of others, and has the smallest K-S statistic value compared to the other models.

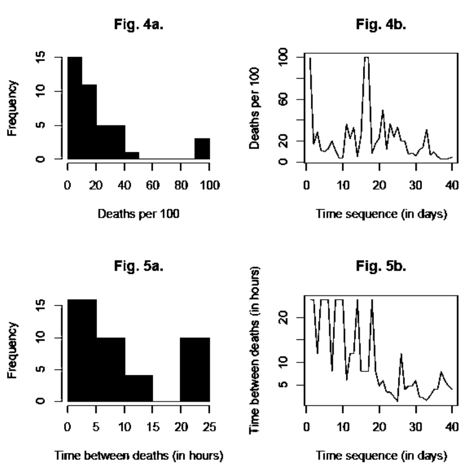

Figure 4 and 5. Histogram and Times Plots of deaths per 100 and time between deaths

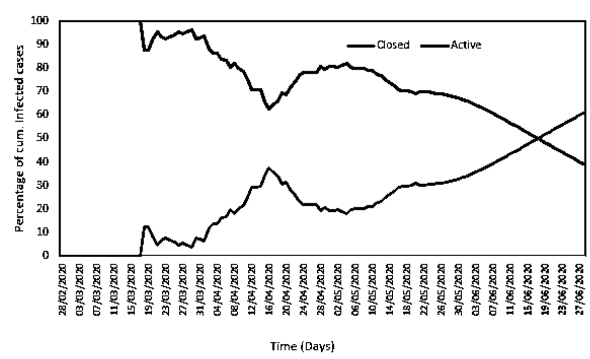

Figure 6. Expected Closed and Active Cases

Figure 6 shows that in the coming days, the closed cases will increase while the active cases will reduce as time (in days) increases. At a point, there will be equilibrium point, where closed cases equate active cases. As time increase, closed cases approach 100%, while active cases approach 0%. This shows that there is hope for better days ahead if susceptible individuals are compliant to NCDC recommendations of physical distancing, wearing of face mask, washing of hands with alcohol base sanitisers, not touching of surfaces and so on.

4 Conclusion

This research developed a new univariate continuous probability distribution called Gamma-Power function distribution with log-logistic quantile function (GPLD) using the T-R{Y} framework. The GPLD is a member of the T-Power{Y} family and results on its statistical properties are presented, such as the cumulative distribution function, density function, the quantile function, survival function, hazard function, cumulative hazard function, moments, and Shannon entropy. The maximum likelihood estimation of the parameters of the model were derived. GPLD distribution was applied to two real datasets on COVID-19 concluded cases and the results of its performance were compared favourably with Gamma, Weibull and Normal distributions. This is a clear indication that a convoluted distribution is a better model compared to known individual distributions. The Weibull distribution is also shown to be a member of this T -Power function family of distributions. The GPLD can be used to fit generalized regression model on COVID-19 induced death rate and time between successive COVID-19 deaths. The GPLD would be used on data, where gamma distribution does not provide a good fit.

References

[1] WHO (2020). There is a current outbreak of coronavirus (COVID-19) disease. https://www.who.int/health-topics/coronavirus.

[2] WHO-China Joint Mission on Coronavirus Disease 2019 (2020). Report of the WHO China Joint Mission on Coronavirus Disease 2019 (COVID-19), 16-24 February 2020. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19final-report.pdf.

[3] NCDC (2020). Nigeria Centre for Disease Control. COVID-19 Situation Report.www.covid19.ncdc. gov.ng

[4] Famoye, F., Akarawak, E., and Ekum, M.I. (2018). Weibull-Normal Distribution and its Applications. *Journal of Statistical Theory and Applications* 17(4), 719–727. DOI: 10.2991/jsta.2018.17.4.12.

[5] Aljarrah, M.A., Lee, C., Famoye, F. (2014). On generating T-X family of distributions using quantile functions, *Journal of Statistical Distributions and Applications* 1(2), doi:10.1890/13-1452.1.

[6] Alzaatreh, A., Lee, C. and Famoye, F. (2014). T-normal family of distributions: a new approach to generalize the normal distribution. *Journal of Statistical Distributions and Applications* 1(16), 1-18.

[7] Eugene, N., Lee, C. and Famoye, F. (2002). Beta-Normal distribution and its applications, *Communications in Statistics - Theory and Methods* 31(4), 497-512.

[8] Alzaatreh, A., Lee, C., Famoye, F. (2013). A new method for generating families of continuous distributions, *METRON* 71,63–79. DOI 10.1007/s40300-013-0007-y.

[9] Zubair, M., Alzaatreh, A., Cordeiro, G.M., Tahir, M.H. and Mansoor, M. (2018). On Generalized Classes of Exponential Distribution using T-X Family Framework *Filomat* 32(4), 1259-1272.

[10] Arowolo1 O. T., Nurudeen T. S., Akinyemi J. A., Ogunsanya A. S. and Ekum M. I. (2019). Reduced Beta Skewed Laplace Distribution with Application to Failure-Time of Electrical Component Data. *Annals of Statistical Theory and Applications (ASTA)* 1, 31-41.

[11] Ogunsanya A. S., Sanni O.O. and Yahya W. B. (2019). Exploring Some Properties of Odd Lomax-Exponential Distribution. *Annals of Statistical Theory and Applications (ASTA)* 1, 21-30.

[12] Ekum, M. I., Adamu, M. O. and Akarawak, E. E. (2020). *T*-Dagum: A Way of Generalizing Dagum Distribution Using Lomax Quantile Function. *Journal of Probability and Statistics*, 1-17. ID 1641207, https://doi.org/10.1155/2020/1641207.

[13] Ekum, M. I., Adamu, M. O., Adeleke, I. A., Akarawak, E. E. and Arowolo, O. T. (2020). Class of Generalized Power Function Distributions: Properties and Applications. preprint, Dept. of Mathematics, University of Lagos. matekum@yahoo.com.

[14] Ahsanullah, M., Shakil, M. and Golam Kibria, B.M.G. (2013). A characterization of the power function distribution based on lower records, *ProbStat Forum* 6, 68–72.

[15] Azedine, G. (2013). Characterization of the power function distribution based on lower records, *Applied Math-ematical Sciences* 7, 5259–5267.

[16] Tahir, M., Alizadehz M., Mansoor M. Cordeiro G. M. Zubair M. (2014). The Weibull-power function dis-tribution with applications. *Hacettepe university bulletin of natural sciences and engineering series b: mathematics and statistics*, DOI: 10.15672/HJMS.2014428212.

[17] Blomquist, N. (1981). A comparison of distributions of annual and lifetime income: Sweden around 1970. *Review of Income and Wealth*. 27(3): 243–264.

[18] R Development Core Team R- A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Austria, Vienna, 2009).