Transactions of the Nigerian Association of Mathematical Physics
Volume 14, (January - March, 2021 Issue), **pp165 –172**

STATISTICAL DISTRIBUTION OF NOVEL CORONAVIRUS IN NIGERIA

Fantola J.O.[1], Adebiyi A.A.[2] and Ganiyu K.A.[1]

[1]Department of Mathematics and Statistics, The Polytechnic, Ibadan, Nigeria
[2]Department of Statistics, University of Ibadan, Ibadan, Nigeria

Abstract

The ongoing Corona Virus Disease (COVID – 19) pandemic, which was first detected in Wuhan; China in December 2019, is a disease from strain of Coronavirus SARS-CoV-2. It is spreading worldwide, and Nigeria like other African countries, has been affected. This virus has presented the greatest world health challenges and it has had adverse effects on the world economic. Currently, the most important priority of the public health practitioners is the prevention and control of the infection, hence the need to know the behavioural pattern and prediction of COVID-19, which is the sole aim of this research work, is important to them.
In this research, four families of the Generalized Gamma Distributions were used to model the reported daily cases of COVID-19 in Nigeria. The reported data, which was downloaded from https://ourworldindata.org/coronavirus spanned through the period of five months (24 March 2020 to 23 August 2020). The result shows that the Weibull distribution is the best fit for the data. The shape parameter is greater than one, meaning the hazard rate of this disease is increasing, the time plot of the reported cases support this view, but we can expect to pass the peak of the disease if proper measures is in place. According to the mean and median of the best fitted distribution, the daily cases is at increasing rate, but we can expect to pass the peak of the disease if proper preventive measures is taken by both the policy makers and the public at large.

***Keywords*:** Corona Virus, Distribution, Weibull, Epidemiology, Nigeria

BACKGROUND OF THE STUDY

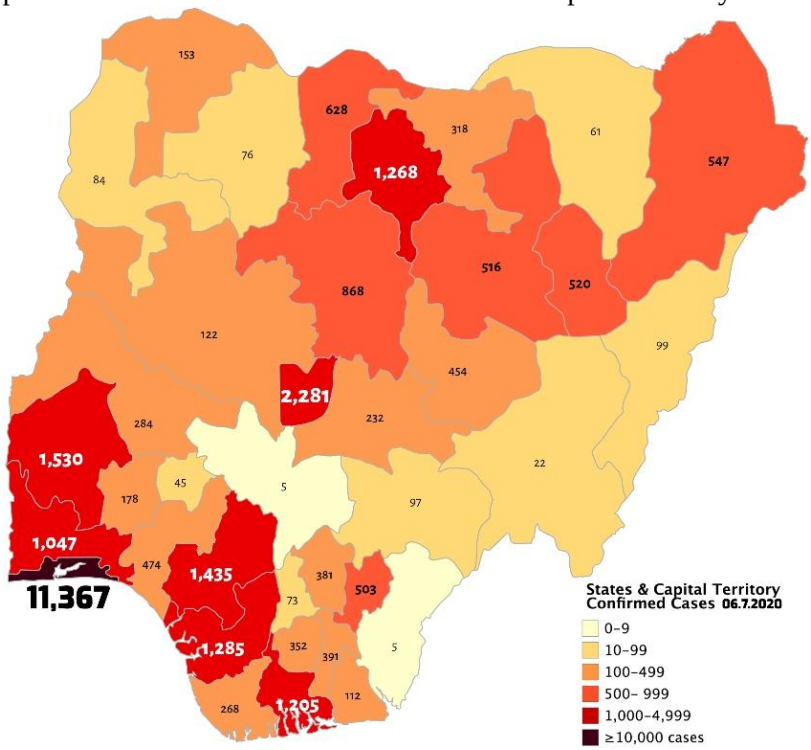
According to [1], COVID-19 is an infectious disease from a large family of coronaviruses which may cause illness in animals or humans. In humans, several coronaviruses are known to cause respiratory infections ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). On December 31st, 2019, the first case of COVID-19 was discovered by the Wuhan Municipal Health Commission Chinese in Wuhan City, Hubei province, China. The commission reported a group of pneumonia cases with unspecified etiology that had a history of exposure to Wuhan's Huanan Seafood Wholesale Market. On January 9th, 2020, The China Center for Disease and Control reported that a novel coronavirus (2019-nCoV) had been detected as the causative agent and the genome sequence was made known to the public [2]

The Corona Virus Disease (COVID-19) is spreading worldwide, and Nigeria like other African countries, has been affected. This virus has presented the greatest world health challenges and it has had adverse effects on the world economic. The early symptoms of COVID-19 involve fever, dry cough, fatigue, and myalgia. Other less common symptoms that may affects some patients include aches and pains, nasal congestion, headache,

Correspondence Author: Fantola J.O., Email: fantola.jubril@polyibadan.edu.ng, Tel: +2348056683809

conjunctivitis, sore throat, diarrhea, loss of taste or smell or a rash on skin or discoloration of fingers or toes. Most people (about 80%) recover from the disease without needing hospital treatment. Around 1 out of every 5 people who gets COVID-19 becomes seriously ill and develops difficulty breathing. Although, However, anyone can catch COVID-19 and become seriously ill, older people, and those with underlying medical problems like high blood pressure, heart and lung problems, diabetes, or cancer, are at higher risk of developing serious illness [1]. The world health technician has been working tirelessly to develop the vaccine that will permanently put an end to the pandemic, but till date, no successful vaccine or antiviral agents has been clinically approved. Therefore, the most important priority of the public health practitioners is the prevention and control of the infection [3].

The first case of Corona Virus (COVID-19) in Nigeria was dated back to 27 February 2020, when an Italian citizen in Lagos state tested positive for the virus [4]. As of 24 August 2020, according to Nigeria Health Authorities, there had been more than 50,000 confirmed cases of infection. The Nigeria map below showed the reported cases of the virus across the State and Capital Territory



https://upload.wikimedia.org/wikipedia/commons/3/3e/States_%26_Territory_by_confirmed_cases_of_COVID-19_in_Nigeria.jpg

Figure 1: Map showing confirmed cases of COVI9-19 in Nigeria

Reported cases of COVID-19 can be useful for epidemiology model fitting and testing once a certain number of cases have occurred. Extensive researches into the diseases are required to fully make clear its pathway and pathogenic mechanism; behavioural pattern, and the potential therapeutic target, which can be helpful to predict the common preventive and therapeutic measures, is an important problem. Therefore, in this study, the distribution of daily reported cases of COVID-19 in Nigeria were was modeled using Generalized Gamma Distribution which comprises Weibull, Gamma, Lognormal, and Exponential distribution.

METHODS

This study modeled the distribution of reported new cases of COVID-19 in Nigeria using a three-parameter Generalized Gamma Distribution (GDD) which comprises Weibull, Gamma, Lognormal and Exponential distribution. GDD is a continuous probability distribution which is sometimes used to determine which parametric model best described a given set of data [5].

Let X be a random variable representing the reported new cases of covid-19 in Nigeria, if X follows a GGD, then, the distribution of X is given by

$$f(x) = \begin{Bmatrix} \frac{\alpha^k}{\beta\Gamma(k)} \, exp^{\left(-\alpha x^{\frac{1}{\beta}}\right)} x^{\frac{k}{\beta}-1} & , \quad x > 0 \\ 0 & , \; elsewhere \end{Bmatrix} \tag{1}$$

Where $\beta>0$ is scale parameter, $\alpha>0$, and $k>0$ are shape parameters.

From equation (1), four families of generalized gamma distribution considered in this study can be gotten as follow:

When $\beta = k = 1$, we have the exponential distribution and the probability density function is

$$f(x) = \begin{Bmatrix} \alpha\, e^{-\alpha x}, & x > 0 \\ 0, & elsewhere \end{Bmatrix} \tag{2}$$

For $\beta = 1$, we have the Gamma distribution whose probability density function is

$$f(x) = \begin{Bmatrix} \frac{\alpha^k}{\Gamma(k)} \, exp^{(-\alpha x)} x^{k-1} & , \quad x > 0 \\ 0 & , \; elsewhere \end{Bmatrix} \tag{3}$$

For k = 1, we have the Weibull distribution with density function

$$f(x) = \begin{Bmatrix} \frac{\alpha}{\beta} \, exp^{\left(-\alpha x^{\frac{1}{\beta}}\right)} x^{\frac{1}{\beta}-1} & , \quad x > 0 \\ 0 & , \; elsewhere \end{Bmatrix} \tag{4}$$

Finally, with k approaching positive infinity ($k \to \infty$), we have Lognormal distribution in which the logarithm of a random variable X is normally distributed as

$$\ln(x) \sim N(\mu, \sigma^2) \tag{5}$$

The distribution was fitted using fitdistrplus package in R (latest version 4.0.1)

Furthermore, this study employed Kolmogorov-Smirnov (K-S), and Anderson-Darling (A-D) test the to compare the goodness of fit statistic [6]. These tests examined if particular set of data (continuous data) comes or follows a specified distribution. However, K-S is a non-parametric test the equality of only continuous distribution with estimation of location and scale parameters whereas A-D test assumes that there are no parameters to be estimated in the distribution being tested, in which case the test and it's set of critical values is distribution-free. The A-D test is an alternative to the chi-square and Kolmogorov-Smirnov goodness-of-fit tests.

Their null hypothesis is that particular data come from a specified distribution. The criteria to either reject or accept the null hypothesis depends on the p-value. If the p-value is less than significance level (0.05 or 0.01), the null hypothesis will be rejected meaning that the data do not follow such distribution. However, it will fail to reject if the p-value is greater than 0.05 indicating that the data follow the specified distribution. Hence, they are included to further determine the confirmed cases of covid which follows the GDD distribution.

The K-S and A-D test Statistic is given below:

Kolmogorov-Smirnov (K-S)

$$D = \max_{1 \le i \le N} \left(F(Y_i - \frac{i-1}{N}, \frac{i}{N} - F(Y_i) \right) \tag{6}$$

Anderson-Darling (A-D)

$$-N - \frac{1}{N}\sum_{i=1}^{N}(2i-1)\log(F_i(1 - F_{n+1-i})) \tag{7}$$

Where F is the cumulative distribution function of the specified distribution.

RESULTS OF ANALYSIS

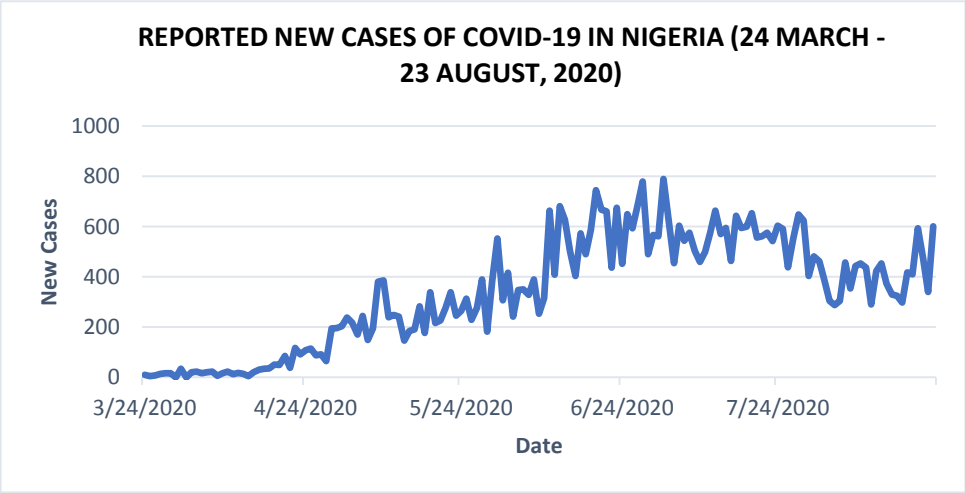
Exploratory Analysis

153 daily data of Corona Virus pandemic (from the period of 24 March 2020 to 23 August 2020) in Nigeria was used in the analysis. As of this period, the lowest number of cases is 0 while the highest number of cases is 790. The average, median, and the modal value of these news cases are respectively 339.1, 340.0, and 16. A time plot of reported number of COVID-19 new cases is shown in the Figure 2. The Figure indicates that there is both a seasonal variation and a fluctuated trend in the reported cases.

Figure 2: Time plot of daily new cases of COVID-19 in Nigeria

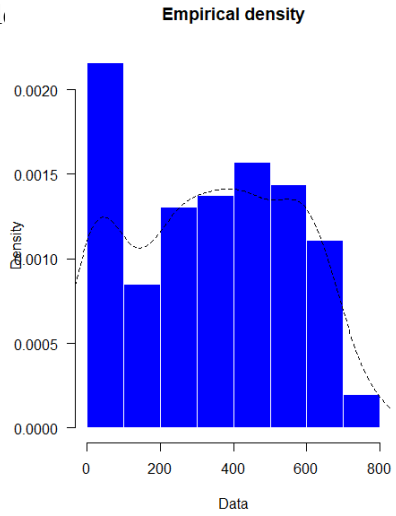
Fitting the distributions

Figure 3 shows two plots of the reported daily cases of COVID-19 in Nigeria; the first one (left-hand plot) is the empirical distribution function which was superimposed on the histogram, and the second one (right-hand plot) is the empirical cumulative distribution function (CDF)



Fitting the distribution

Figure 3 shows two plots of the reported cases of COVID-19 in Nigeria; the first one (3a, left-hand plot) is the empirical distribution function which was superimposed on the histogram, and the second one (3b, right-hand pl bution function (CD



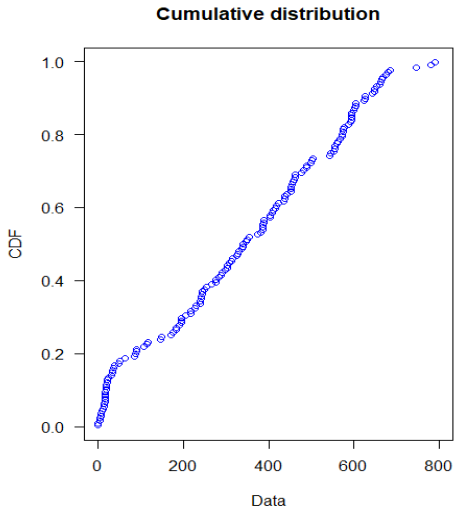


Figure 3a ***Figure 3b***

Histogram (3a) and CDF plots (3b) of an empirical distribution for COVID-19 daily reported cases.

In the call to fitdist function from the fitdistrplus R's package developed by Marie Laure Delignette-Muller and Christophe Dutang. The four families of a three-parameter Generalized Gamma Distribution were fitted using Maximum Likelihood Estimation (MLE) method. The numerical results returned by fitdist function, as posited by [7], are:

1. the parameter estimates,
2. the estimated standard error (by MLE),
3. the Akaike and Bayesian Information Criteria (AIC and BIC).
4. and the correlation between the parameters estimate of the fitted distribution

However, in this study, the AIC was used to determine the goodness of fit of the distributions. The distribution with the lowest AIC value will be the best fitted distribution.

Table 1 and Table 2 show the summary statistics and the goodness-of-fit of the four fitted distributions respectively. From Table 2, Kolmogorov-Smirnov statistic and Anderson-Darling statistic showed that the p-values of all the four distribution are not significant ($p>0.05$). This means that all the distributions (Weibull, Lognormal, Gamma, and Exponential) can be used to model COVID-19 daily cases in Nigeria. However, From the values of the Akaike Information Criteria (AIC), it can be conjectured that the best fitted distribution these cases follows the Weibull distribution which has the lowest value of AIC among others

Table 1: Summary statistics and AIC of COVID -19 daily cases in Nigeria

Distribution	**Parameters**	**Estimation**	**AIC**
Weibull	Shape	2.10	2080.74
	Scale	495.60	
Gamma	Shape	3.05	2093.19
	Rate	0.007	
Lognormal	Meanlog	1.00E+00	2112.76
	Sdlog	-5.45	
Exponential	Rate	0.002	2169.89

Source: The Authors' computation

Table 2: goodness-of-fits statistics of distribution COVID -19 daily cases in Nigeria

Source: The Authors' computation, *the distribution with the lowest AIC value

	Weibull	**gamma**	**lognormal**	**exponential**
Kolmogorov-Smirnov statistic	0.1043290	0.1086545	0.1295805	0.2124771
Cramer-von Mises statistic	0.3777274	0.5892118	0.9429584	2.6933610
Anderson-Darling statistic	3.1543055	4.1658274	6.1853357	15.0031660

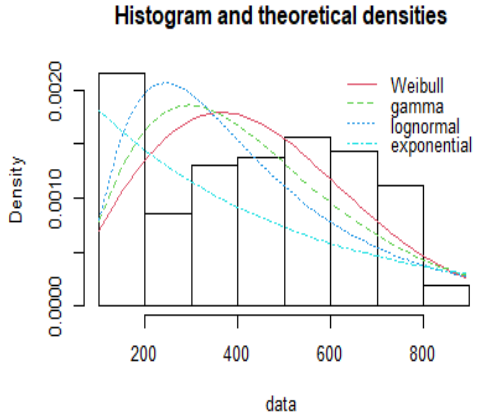
Goodness-of-fit criteria	**Weibull**	**gamma**	**lognormal**	**exponential**
Akaike's Information Criterion	2080.740*	2093.190	2112.763	2169.893
Bayesian Information Criterion	2086.801	2099.251	2118.823	2172.924

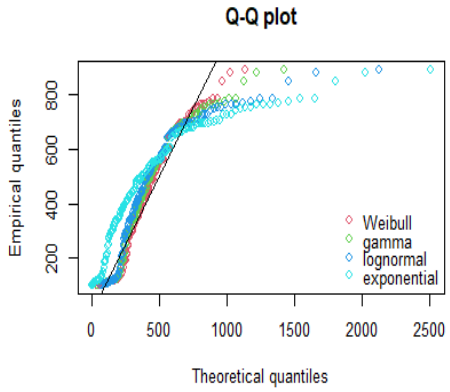
Distribution plots

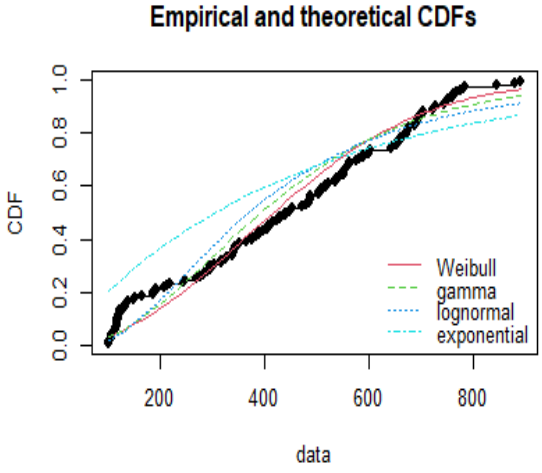
The plots of the four classical goodness-of-fit [8] are presented in Figure 3:

1. a density plot representing the density function of the fitted distribution along with the histogram of the empirical distribution of the daily reported COVID-19 cases (top-left)
2. a CDF plot of both the empirical distribution and the fitted distribution (top-right)
3. Q-Q plot representing the empirical quantiles (y-axis) against the theoretical quantiles (x-axis) (bottom-left), and
4. a P-P plot representing the empirical distribution function evaluated at each data point (y-axis) against the fitted distribution function (x-axis)

In the Figure 4, all the four distributions were plotted separately on each of the aforementioned four plots, this allows for easier comparison of the empirical distribution and multiple continuous distribution fitted on the same data set [8]. The density plot (Fig. 4a, top-right) shows that the Weibull distribution is the most fitted distribution that is close to being normally distributed followed by gamma distribution. Also, from the Q-Q plot (Fig. 4b, top-left), Empirical and Theoretical CDF plot (Fig. 4c, bottom-left), and P-P plot (Fig. 4d, bottom-right), it can be observed that Weibull distribution was tightly fitted to the empirical observation (the black solid line on the plot) of daily COVID-19 cases in Nigeria.







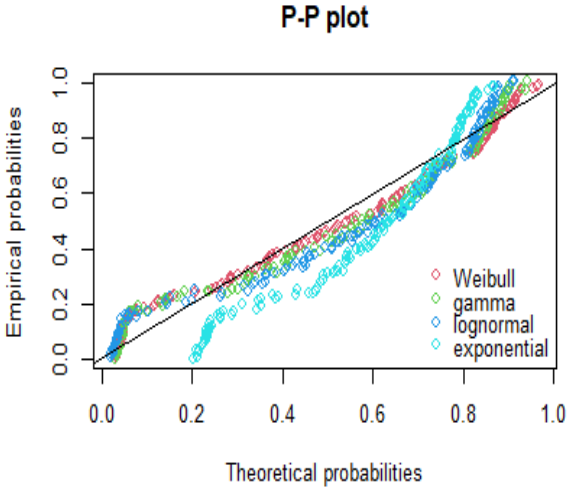


Figure 4c ***Figure 4d***

Four goodness-of-fit plots for various distributions fitted to daily COVID-19 cases in Nigeria

DISCUSSION

World Health Organization, WHO Director-General's opening remarks at the media briefing on COVID-19 on 11 March 2020, said "In the days and weeks ahead, we expect to see the number of cases, the number of deaths, and the number of affected countries climb even higher. We have therefore made the assessment that COVID-19 can be characterized as a pandemic". Hence, this present study gives an indication of the spread of this pandemic in Nigeria. We used four families of Generalized Gamma Distribution which are Weibull, Gamma, Lognormal, and Exponential distribution. We used real daily data reported by Nigeria Ministry of Health. Our result shows that the Weibull distribution best describes the pattern of the data, the shape parameters of the distribution is greater than one – indicating that case is still increasing. Also, with the assumption of this distribution, the mean and median of the data are 408.90, 386.49 which is high compared to the mean and median of the original data. It is therefore required that Policy makes and the public at large should ensure that proper preventive measures, otherwise the peak of the pandemic in the country will be very alarming. In conclusion, the Weibull distribution is the best fit of the COVID-19 daily cases in the country, according to this distribution, the cases are increasing, and we have yet to reach the peak of the disease. Proper preventive measures should well be accustomed to by the policy makers and the public at large.

REFERENCES

[1] World Health Organization, WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 (Accessed: 27 March 2020). Available at: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at10 the-media-briefing-on-covid-19---11-march-2020

[2] European Centre for Disease Prevention and Control, Risk assessment: Outbreak of acute respiratory syndrome associated with a novel coronavirus, Wuhan, China; first update, [March 27, 2020]. Available from: https://www.ecdc.europa.eu/en/publications-data/risk-assessment-outbreak-acuterespiratory-syndrome-associated-novel-coronavirus.

[3] Worldometer, COVID-19 Coronavirus Pandemic (Accessed: 27 March 2020). Available at: https://www.worldometers.info/coronavirus/

[4] FIRST CASE OF CORONA VIRUS DISEASE CONFIRMED IN NIGERIA". Nigeria Centre for Disease Control (NCDC). *28 February 2020*. Retrieved 10 March 2020.

[5] Box-Steffensmeier, Janet M.; Jones, Bradford S. (2004) Event History Modeling: A Guide for Social Scientists. Cambridge University Press. ISBN 0-521-54673-7 (pp. 41-43)

[6] Delignette-Muller ML, Pouillot R, Denis JB, Dutang C (2015). _tdistrplus: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data. R package version 1.0-4, URL http://CRAN.R-project.org/package=fitdistrplus.

[7] Delignette-Muller, M.L, Christophe D. (2015). fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software. Volume 64, Issue 4. http://www.jstatsoft.org/

[8] Cullen AC, Frey HC (1999). Probabilistic Techniques in Exposure Assessment. 1st edition. Plenum Publishing Co.