International Journal of Mathematics Trends and Technology
ISSN: 2231-5373/ https://doi.org/10.14445/22315373/IJMTT-V68I12P5111
Volume 68 Issue 12, 100-105, December 2022

Original Article

Applied Some Probability Density Function for Frequency Analysis of New Cases Covid-19 in Indonesia

Vinny Anugrah[1], Rado Yendra[2], Muhammad Marizal[3], Rahmadeni[4]

[1,2,3,4] *Department of Mathematics, Faculty of Science and Technology, Universitas Islan Negeri Sultan Syarif Kasim Riau, Indonesia*

Received: 29 October 2022 Revised: 04 December 2022 Accepted: 16 December 2022 Published: 31 December 2022

Abstract - *Changes in the number of positive COVID-19 patients in Indonesia greatly affect the number of beds in hospitals. Indonesia as a country that has a large population requires an accurate frequency analysis to make policies in dealing with changes in the number of patients. This study focuses on analyzing the frequency of COVID-19 patients by using probability modeling. Probability modeling will be carried out using four probability density functions (pdf) namely Gamma, Amarendra, Rani, and Sujatha 2 Parameters will be used in this study. The estimated parameter from the pdf used in this study will be obtained using the maximum likelihood technique. The distribution will be chosen using a few Good of Fit Test techniques, including graphical (pdf plot and CDF plot) and numerical (Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) methods. The majority of the time, graphical approaches get identical results, however their AIC and BIC results differ. The distribution with the lowest AIC and BIC values is chosen as the best appropriate outcome. The Sujatha two parameter distribution has generally been deemed to be the best model.*

Keywords – *Covid 19, Amarendra distribution, Gamma distribution, Rani distribution, Sujatha distribution.*

1. Introduction

Covid 19 is an infectious disease that spreads quickly and became very viral after it was first discovered in December 2019 in Wuhan City of China [1]. The first case outside China was found in Thailand on 13 January 2020 [2], Quick spread in both local and international transmissions [3] including Indonesia. The COVID-19 pandemic in Indonesia is part of the ongoing worldwide pandemic of coronavirus disease 2019 (COVID-19). It was confirmed to have spread to Indonesia on March 2, 2020, after a dance instructor and her mother tested positive for the virus. Both were infected by a Japanese national.[4]. Probability distributions have been used by academics from a range of disciplines to model and analyze the frequency of daily COVID-19 data cases. Short summaries of the relevant literature are provided below. Soltani-Kermanshahi et al. [5] investigated the probability modeling of fresh coronavirus data in Iran. Based on daily reported data, this study examines the normal, log-normal, and weibull distributions of COVID-19 cases, three different forms of parametric distributions. Yusuf et al. [6] Forecasting monthly data on patients with COVID-19 disease that occurred in Pakistan. Zafar, J. et al [7] Compared several daily data models of COVID-19 patients in Pakistan. Weibull distribution (WD), Power function distribution (PFD), Log-Logistics Distribution (LLD), Log-Normal Distribution (LND), Inverse Weibull Distribution (IWD), Gumbel Distribution (GuD), Burr III Distribution (BIIID), Burr XII Distribution (BXIID), and Birnbaum Saunders Distribution make up the 2-parameter probability distribution model that was used (BSD). The Susceptible-Infectious-Recovered model and the log-linear regression model, two straightforward epidemiological mathematical models, were employed to examine the daily and cumulative incidence of COVID-19 in Spain and Italy. [8]. Forecasting the COVID-19 pandemic that occurred in Turkey was modeled using the Bayesian negative multilevel binomial model with mixed effects [9]. The choosing of the optimal season for the eradication of this illness is more vulnerable because of the unique nature of the virus. To better manage societal, financial, social, and general medical issues, transitory deciding is crucial even for the tiniest insight into the upcoming month [10]. The behavior of pandemies, crop harvesting, commercial data mining, e-commerce fraud, as well as other practical challenges, has been described using data science methodologies. [11–22]. Scientists have just recently developed or applied proven scientific and quantifiable methods to forecast the number of COVID-19 cases and associated outcomes. The summed-up strategic model demonstrates that the spread of the plague in China was exponential [23]. According to the prognosis, the situation will worsen across all of Europe, and the USA will become the focus of new cases around the middle of April 2020. [24]. In this study, several probability models will be used to analyze the frequency of positive COVID-19 sufferers in Indonesia. Indonesia is a country with a very large population that is greatly affected by the outbreak of the covid 19 virus. Not many studies have been based on probability modeling on the data of COVID-19 cases, therefore in this study, several probability models will be used to analyze the frequency of positive COVID-19 cases. 19 in Indonesia. However, research has found that uses a mathematical model using data on new cases of covid 19 [25]. Simulation of the number of COVID-19 sufferers in Indonesia is also carried out through a mathematical model [26]. This study aims to select the best model for analyzing the frequency of cases of covid 19 in Indonesia, the model used is the distribution of Gamma, Rani, Amarendra, and Sujatha 2 parameters. The maximum likelihood method will be used in estimating Parameters. The two best methods in determining the best model, such as the graphical method and the numerical method, will be carried out in this study. The graphical method will be carried out by comparing the pdf and CDF of each model while the numerical values of AIC and BIC will complete this research as a representation of the numerical method. Model selection using the graphical method often gives the same results for each model but using the numerical method will result in the selection of the best model. easy based on the lowest value generated by AIC and BIC

2. Data and Study Area

Data on daily new cases of covid 19 obtained for the period March 12 2020 to May 11, 2020, from the public report of the National Institute of Health, Indonesia. Summary of data presented in descriptive statistics as in table 1. In the table it can be concluded that the skewness value is very small and the kurtosis value does not exceed 2, this indirectly indicates that some distributions with 2 parameters such as log-normal, Weibull, and Normal can be used in this modeling

Table 1. Desciptive Statistics of New Cases Positif Covid 19 in Indonesia

Mean	Standrard deviasion	Minimum	Maximum	Skewness	Kurtosis
222.7	145.72	2.0	533.0	8.94×10^{-19}	1.80

3. Methods

3.1. Probability Density Function (PDF) and Cumulative Distribution Function (CDF)

The new cases covid 19 modeling require analysis of the number of daily patients with positive covid 19 data over a number of years. Probability density functions are the main techniques for describing the novel situations with 19 properties. The two-parameter gamma distribution with shape parameter and scaling parameter β is one of four probability density functions, the Rani distribution with parameter θ, the Amarendra distribution with parameter **Ω** and Sujatha two parameters distribution with scale parameters λ and shape parameters γ associated with modeling daily new cases covid 19, are thought about in this paper. Table 2 shows the pdf and CDF for each distribution we take into consideration, where y stands for the observed values of the random variable corresponding to the event of interest. Several researchers have used the probability density function in this study for various purposes [27–33].

Table 2. Probability Density Function (pdf) and distribution function (cdf) four distributions

	Distribution	pdf (*f(y)*) dan cdf (*F(y)*)
1	Gamma (y;α,β)	$f(y)=\frac{1}{\beta^{\alpha}\Gamma(\alpha)}y^{\alpha-1}e^{-\frac{y}{\beta}},\quad y>0,\quad \alpha,\beta>0$ $F(y)=\frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_{o}^{y}x^{\alpha-1}e^{-x/\beta}dx$
2	Rani (y;θ)	$f(y)=\frac{\theta^5}{\theta^5+24}(\theta+y^4)e^{-y\theta}, y>0,\theta>0$ $F(y)=1-\left[1+\frac{\theta x(\theta^3y^3+4\theta^2y^2+12\theta y+24)}{\theta^5+24}\right]e^{-\theta y}$
3	Amarendra(y;**Ω**)	$f(y)=\frac{\Omega^4}{\Omega^3+\Omega^2+2\Omega+6}(1+y+y^2+y^3)e^{-y\beta}\ , y>0,\Omega>0$ $F(y)=1-\left[1+\frac{\Omega^3y^3+\Omega^2(\Omega+3)y^2+\Omega(\Omega^2+2\Omega+6)y}{\Omega^3+\Omega^2+2\Omega+6}\right]e^{-y\Omega}$
4	Sujatha (y;λ,γ)	$f(y)=\frac{\lambda^3}{\gamma\lambda^2+\lambda+2}(\gamma+y+y^2)e^{-y\lambda}, y>0,\lambda,\gamma>0$ $F(y)=1-\left[1+\frac{\lambda y(\lambda y+\lambda+2)}{\gamma\lambda^2+\lambda+2}\right]e^{-y\lambda}$

The choice of the model design and parameter estimation techniques is crucial for choosing the best fit model. The maximum likelihood approach is used to calculate the parameter estimation of the distribution function. We won't go into detail in this work on the implicit and complex maximum likelihood function for this model. The Newton Rapshon numerical approach must be used to solve the non-linear equation produced by the maximum log-likelihood function (ln L). However, the iteration system has used this approach to obtain the answer. For this technique, several beginning values have been examined. It is thought to be the selected estimated parameter if the starting value falls within the range of the same value. Both numerically and graphically, the process of goodness of fit testing for model selection is presented

3.2. Maximum Likelihood Estimate (MLE) and Goodness of Fit Tests (GOF)

Let $(y_1, y_2, \cdots, y_n)$ be a random sample from four PDFs, The natural log likelihood (ln L) is presented in Table 3. The MLE $\hat{\theta}\ of\ \theta$ is the solution of the equation $\frac{d \ln L}{d\theta} = 0$ and thus it is the solution of the following nonlinear equation. The best distribution is determined utilizing the findings from various goodness-of-fit tests. The GOF tests taken into consideration are based on numerical criteria and the graphical inspection probability density function (PDF). To ascertain the distributions' goodness-of-fit standards, Akaike's information criterion (AIC) and Bayesian information criterion (BIC) were used. The majority of the time, graphical inspection produced the same outcome, however the AIC and BIC findings varied. The distribution with the lowest AIC and BIC values was determined to be the best fit outcome. The following is the formula for calculating AIC and BIC:

$$AIC = -2\ln L + 2k,$$

$$BIC = -2\ln L + k\ln n,$$

where k = the number of parameters, n = the sample size

Table 3. Log Likelihood function (*ln L*)

Distribution	*ln L*
Gamma	$n\alpha log(\beta) + (\alpha - 1)\sum log(x) - \beta\sum x - nlog(\Gamma(\alpha)$
Rani	$nlog\left(\frac{\theta^5}{\theta^5 + 24}\right) + \sum log(\theta + y^4) - n\theta\bar{y}$
Amarendra	$n\left(log\left(\frac{\Omega^4}{\Omega^3 + \Omega^2 + 2\Omega + 6}\right)\right) + \sum log(1 + y + y^2 + y^3) - n\Omega\bar{y}$
Sujatha	$n\big(3\ log(\lambda) - log(\gamma\lambda^2 + \lambda + 2)\big) + \sum log(\gamma + y + y^2) - n\lambda\bar{y}$

4. Results and Discussion

In this study, the daily data histogram of new Covid 19 cases in Indonesia will be approximated by the four distributions or probability density functions used in this study. This of course can be obtained based on the estimated parameter values as presented in table 2. The maximum likelihood estimation approach is used to estimate the distributional parameters..

Table 4. Computed parameters of different distribution

Distribution	Parameters					
	α	**β**	**θ**	**Ω**	**λ**	**γ**
Gamma	1.20	0.0054				
Rani			0.022			
Aramendra				0.018		
Sujatha 2 Paremeters					7432.36	0.011

Figure 1 shows that the Gamma distribution is not good in approaching the frequency of daily Covid-19 case data in Indonesia, while the Rani and Amarendra distributions are also not very good at capturing the frequency of these data. Different results are shown by the 2-parameter Sujatha Distribution, this distribution is able to approach the histogram or the frequency of daily Covid-19 case data that occurs in Indonesia.

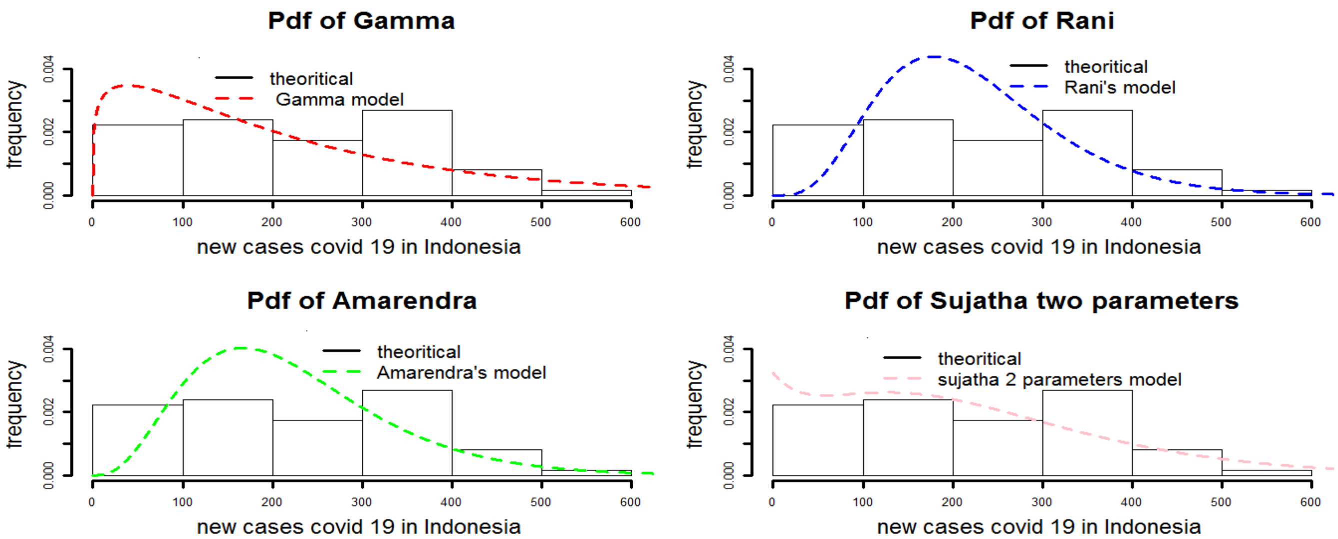

Fig. 1 Pdf plot for comparisons Predicted and observed daily new cases covid 19 in Indonesia

Furthermore, the plot of the cumulative distribution function will be used to make more convincing conclusions in choosing the best distribution in this study. Figure 2 is thus provided for this purpose as well. From the figure 2, it is very clear that the Gamma, Rani, and Amarendra distributions are not good in approaching the observed distribution function, while the opposite result is shown by a 2-parameter Sujatha distribution. The resulting distribution function is very good at capturing the observation distribution function.

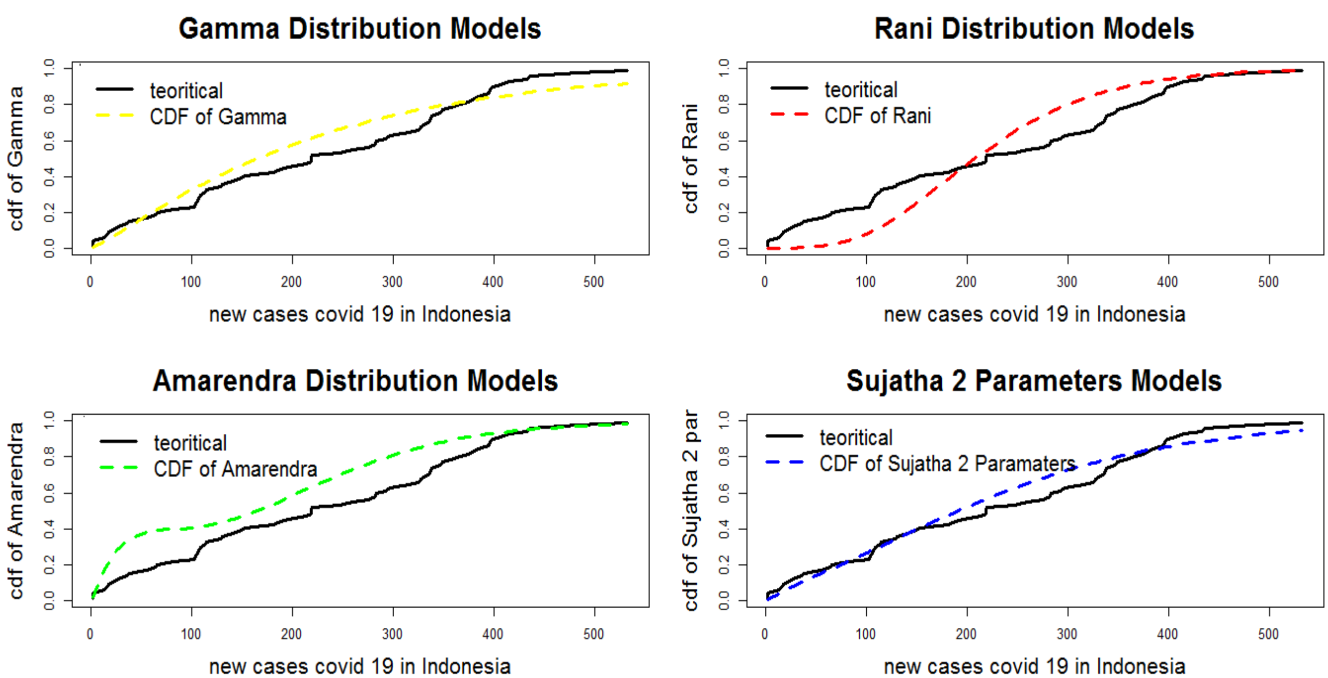

Fig. 2 Cdf plot for comparisons Predicted and observed daily new cases covid 19 in Indonesia

From the goodness of fit test of the model using this graphical method, it is very clear that the pdf and CDF plots give very clear results that the 2-parameter Sujatha distribution is the best model in analyzing the frequency of covid 19 cases in Indonesia. Numerical methods such as AIC and BIC values for the goodness-of-fit test were also used in this study. These two values for each distribution used will be presented in table 3. Based on the values in the table, it can be concluded that the Sujatha 2 parameter distribution is the best model because it has the smallest AIC and BIC values.

Table 5. The Goodness and fit test result of the daily new cases covid 19 in Indonesia

	Gamma	Rani	Amarendra	Sujatha 2 Parameters
AIC	809.8525	936.1560	887.3286	798.6539
BIC	810.0525	936.2216	936.2216	798.8539

5. Conclusion

In order to choose the 4 most suitable models or distributions that may be used to represent the distribution of the daily number of new COVID-19 cases, this study will investigate the daily case data for the disease in Indonesia. Conclusion: When compared to other well-known distributions, the Sujatha 2 parameters distribution produced better results. This judgment is supported by popular goodness-of-fit test models like AIC and BIC.. The graphical technique namely pdf and CDF plots were also observed comparing the empirical distributions with the adjusted by the Sujatha 2 parameters distribution. In addition, through the best model in this study, we can use the distribution of the quantile function to simulate the number of daily new cases of covid 19 in the future. These results are interesting for estimating the number of hospital beds during resource allocation planning or social isolation policies.

References

[1] Catharine I. Paules, Hilary D. Marston, and Anthony S. Fauci, "Coronavirus Infections—More than Just the Common Cold," *JAMA*, vol. 323, no. 8, pp. 707–708, 2020. *Crossref*, http://doi.org/10.1001/jama.2020.0757

[2] Nungruthai Suntronwong et al., "Impact of COVID-19 Public Health Interventions on Influenza Incidence in Thailand," *Pathogens and Global Health*, vol. 114, no. 5, pp. 225-227, 2020. *Crossref*, http://doi.org/10.1080/20477724.2020.1777803

[3] Jimmy Whitworth, "COVID-19: A Fast-Evolving Pandemic," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 114, no. 4, pp. 241-248, 2020. *Crossref*, http://doi.org/10.1093/trstmh/traa025

[4] I. R. Rachmawati, and Kumiko Shishido, "Travelers' Motivation to Travel Abroad During Covid-19 Outbreak," *International Journal of Applied Sciences in Tourism and Events*, vol. 4, no. 1, pp. 1-11, 2020. *Crossref*, http://dx.doi.org/10.31940/ijaste.v4i1.1772

[5] Elham Gholami, Kamyar Mansori, and Mojtaba Soltani-Kermanshahi, "Statistical Distribution of Novel Coronavirus in Iran," *International Journal of One Health*, vol. 6, no. 2, pp. 143-146, 2020. *Crossref*, http://doi.org/10.14202/IJOH.2020.143-146

[6] Muhammad Yousaf et al., "Statistical Analysis of Forecasting COVID-19 for Upcoming Month in Pakistan," *Chaos Solitons Fractals*, vol. 138, p. 109926, 2020. *Crossref*, https://doi.org/10.1016/j.chaos.2020.109926

[7] Muhammad Ahsan-ul-Haq et al., "Modeling of COVID-19 Cases in Pakistan Using Lifetime Probability Distributions," *Annals of Data Science*, vol. 9, pp. 141–152, 2022. *Crossref*, https://doi.org/10.1007/s40745-021-00338-9

[8] Jeffrey Chu, "A Statistical Analysis of the Novel Coronavirus (COVID-19) in Italy and Spain," *PLoS ONE*, vol. 16, no. 3, p. e0249037, 2021. *Crossref*, https://doi.org/10.1371/journal.pone.0249037

[9] Aybar Can Acar et al., "Projecting the Course of Covid-19 in Turkey: A Probabilistic Modeling Approach," *Turkish Journal of Medical Sciences*, vol. 51, no. 1, pp. 16-27, 2021. *Crossref*, https://doi.org/10.3906/sag-2005-378

[10] Fotios Petropoulos, and Spyros Makridakis, "Forecasting the Novel Coronavirus COVID-19," *PLoS ONE*, vol. 15, no. 3, pp. e0231236, 2020. *Crossref*, https://doi.org/10.1371/journal.pone.0231236

[11] Sanjay Kumar, "Monitoring Novel Corona Virus (COVID-19) Infections in India by Cluster Analysis," *Annals of Data Science*, vol. 7, pp. 417–425, 2020. *Crossref*, https://doi.org/10.1007/s40745-020-00289-7

[12] Aman Khakharia et al., "Outbreak Prediction of COVID-19 for Dense and Populated Countries Using Machine Learning," *Annals of Data Science*, vol. 8, pp. 1-19, 2021. *Crossref*, https://doi.org/10.1007/s40745-020-00314-9

[13] Jianping Li et al., "Culture vs Policy: More Global Collaboration to Effectively Combat COVID-19," *The Innovation*, vol. 1, no. 2, p.100023, 2020. *Crossref*, https://doi.org/10.1016/j.xinn.2020.100023

[14] Yang Liu et al., "What are the Underlying Transmission Patterns of COVID-19 Outbreak? An Age-Specific Social Contact Characterization," *eClincial Medicine*, vol. 22, p. 100354. 2020. *Crossref*, https://doi.org/10.1016/j.eclinm.2020.100354

[15] Paula Ianishi et al., "Probability on Graphical Structure: A Knowledge-Based Agricultural Case," *Annals of Data Science*, vol. 9, pp. 327-345, 2022. *Crossref*, https://doi.org/10.1007/s40745-020-00311-y

[16] David L. Olson, and Shi Yong, *Introduction to Business Data Mining*, New York: McGraw-Hill/Irwin, 2007.

[17] Shi Yong et al., *Optimization Based Data Mining: Theory and Applications*, Berlin: Springer, 2011.

[18] James M. Tien, "Internet of Things, Real-Time Decision Making, and Artificial Intelligence," *Annals of Data Science*, vol. 4, no. 2, pp. 149–217, 2017. *Crossref*, https://doi.org/10.1007/s40745-017-0112-5

[19] Diego C. Nascimento et al., "Risk Management in E-Commerce: A Fraud Study Case Using Acoustic Analysis through Its Complexity," *Entropy*, vol. 21, no. 11, pp. 1087, 2019. *Crossref*, https://doi.org/10.3390/e21111087

[20] Pedro L. Ramos et al., "Modeling Traumatic Brain Injury Lifetime Data: Improved Estimators for the Generalized Gamma Distribution under Small Samples," *PLoS ONE*, vol. 14, no. 8, p. e0221332, 2019. *Crossref*, https://doi.org/10.1371/journal.pone.0221332

[21] Xuetao Cao, "COVID-19: Immunopathology and its Implications for Therapy," *Nature Reviews Immunology*, vol. 20, no. 5, pp. 269–270, 2020. *Crossref*, https://doi.org/10.1038/s41577-020-0308-3

[22] Ricardo J. Pais, and Nuno Taveira, "Predicting the Evolution and Control of COVID-19 Pandemic in Portugal," *medRxiv*, 2020. *Crossref*, https://doi.org/10.1101/2020.03.28.20046250

[23] K.Roosa et al., "Real-Time Forecasts of the COVID-19 Epidemic in China from February 5th to February 24th, 2020," *Infectious Disease Modelling*, vol. 5, pp. 256–263, 2020. *Crossref*, https://doi.org/10.1016/j.idm.2020.02.002

[24] Pavan Kumar et al., “Forecasting the Dynamics of COVID-19 Pandemic in Top 15 Countries in April 2020: ARIMA Model with Machine Learning Approach,” *MedRxiv*, 2020. *Crossref,* https://doi.org/10.1101/2020.03.30.20046227

[25] Dipo Aldila et al., “A Mathematical Study on the Spread of COVID-19 Considering Social Distancing and Rapid Assessment: The Case of Jakarta, Indonesia,” *Chaos, Solitons and Fractals*, vol. 139, p. 110042, 2020. *Crossref,* https://doi.org/10.1016/j.chaos.2020.110042

[26] Nuning Nuraini, Kamal Khairudin, and Mochamad Apri, “Modeling Simulation of COVID-19 in Indonesia Based on Early Endemic Data,” *Communication in Biomathematical Sciences*, vol. 3, no. 1, pp. 1-8, 2020. *Crossref,* https://doi.org/10.5614/cbms.2020.3.1.1

[27] Rama Shanker, “Rani Distribution and its Application,” *Biometrics & Biostatistics International Journal,* vol. 6, no. 1, pp. 256-265, 2016. *Crossref,* https://doi.org/10.15406/bbij.2017.06.00155

[28] Rama Shanker, “Akash Distribution and Its Applications,” *International Journal of Probability and Statistics*, vol. 4, no. 3, pp. 65-75, 2015. *Crossref,* https://doi.org/10.5923/j.ijps.20150403.01

[29] Rama Shanker, “Shanker Distribution and its Applications,” *International Journal of Statistics and Applications*, vol. 5, no. 6, pp. 338-348, 2015. *Crossref,* https://doi.org/10.5923/j.statistics.20150506.08

[30] Rama Shanker, “Aradhana Distribution and its Applications,” *International Journal of Statistics and Applications,* vol. 6, no. 1, pp. 23-34, 2016. *Crossref,* https://doi.org/10.5923/j.statistics.20160601.04

[31] Rama Shanker, “Sujatha Distribution and its Applications,” *Statistics in Transition-New Series*, vol. 17, no. 3, pp. 391-410, 2016. *Crossref,* https://doi.org/10.21307/stattrans-2016-02

[32] Rama Shanker, “Amarendra Distribution and its Applications,” *American Journal of Mathematics and Statistics*, vol. 6, no. 1, pp. 44-56, 2016. *Crossref,* https://doi.org/10.5923/j.ajms.20160601.05

[33] R. Shanker. “Devya Distribution and its Applications,” *International Journal of Statistics and Applications,* vol. 6, no. 4, pp. 189-202, 2016. *Crossref,* https://doi.org/10.5923/j.statistics.20160604.01