

Modeling of Corona Virus Infection Probability in Indonesia

Bustami*, Fachroh Fiddin

Department of Business Administration

State Polytechnic of Bengkalis

Bengkalis, Indonesia

*Bustami@Polbeng.ac.id, FachrohFiddin@Polbeng.ac.id

Abstract—Coronavirus (COVID-19) was reported at the end of December 2020 in Wuhan, China, and has become a global pandemic. Indonesia's first case of the Coronavirus infection was on March 2nd, 2020, and the number grows continuously. For the mentioned cases, the authors are interested in studying how to model coronavirus infection probability in Indonesia. The purpose of this study is to determine the model of infection probability from humans to humans in Indonesia associated with government intervention in handling the pandemic. The Modeling is done by collecting data and modeling it using the probability theory and simulate the data with the help of software to obtain a proper model. The result of this study shows that the best model found is Johnson SB distribution. This model hopefully can contribute to handling the pandemic such as helping the government to predict the possible additional number of people get infected daily by the virus and make anticipation steps to cure them. Besides, this study is also aimed to introduce another way to predict the number of people get infected by the COVID-19 that has been introduced by previous researches.

Keywords—*coronavirus, infection, probability*

I. INTRODUCTION

On December 29, 2019, 4 patients were reported by local hospitals in Hubei Province as “pneumonia of unknown etiology” in the national infectious disease surveillance system. These patients were all related to the Huanan (Southern China) Seafood Wholesale Market. Wuhan Health Commission announced an outbreak of pneumonia of unknown etiology in Wuhan on December 31, 2019. On January 8, a Novel Corona Virus was officially announced as the causative pathogen of the outbreak by the Chinese Center for Disease Control and Prevention. At present, Corona Virus Disease 2019 (COVID-19) cases have been reported in many countries of the world, causing a serious social burden and a public health emergency of international concern declared by the World Health Organization. Asymptomatic infected persons or asymptomatic periods before the onset of the cases are not easy to identify so it is easy to cause omission, but the epidemiological significance cannot be ignored. Now the transmission capacity of cases as a source of infection during the asymptomatic period (incubation period) is not well known [1].

The presence of the Corona Virus causes many damages such as many people have postponed their activities; industrial processes are interrupted; business trips are delayed; tourism is stopped; even Umrah is forced to cancel. In Indonesia, the economy has also weakened the stock price index; investors have withdrawn their money; the Rupiah weakened and also expected to have an impact on economic growth.

The coronavirus outbreak is severely disrupting the global economy. Almost all the nations are struggling to slow down the transmission of the disease by testing and treating patients, quarantining suspected persons through contact tracing, restricting large gatherings, maintaining complete or partial lockdown, etc. [2].

On the problem early mentioned, we are interested in doing research related to the coronavirus. The study is done based on the data released by the Indonesian government on its website (covid19.go.id) released daily starting from March to June 2020. The model can hopefully be used to estimate the number of people who can be infected by the virus. So it is hoped that the government benefitted from this study.

Meanwhile, various studies have been undertaken related to this research i.e.:

- The research conducted by Dejen Ketema Mamo entitled “Model the transmission dynamics of COVID-19 propagation with public health intervention” In this work, a researcher develops the SHEIQRD (Susceptible–Stay-at-home–Exposed–Infected–Quarantine–Recovery–Death) coronavirus pandemic, spread model. The disease-free and endemic equilibrium points are computed and analysed. The basic reproduction number R_0 is acquired, and its sensitivity analysis is conducted. COVID-19 pandemic spread dies out when $R_0 \leq 1$ and persists in the community whenever $R_0 > 1$ [3].
- The research conducted by Dongming et al. entitled “Epidemiological characteristics and transmission model of Corona Virus Disease 2019 in China”. They found that from the respect of transmission character, they could found an obvious geographic discrepancy

over time. And the disease spread to almost the whole of China just for 14 days. Several factors may attribute to this phenomenon, for instance, the movement of asymptomatic and potential infected cases from Wuhan, the density of population in these cities, the control measures in these cities, etc. Of which, the first one may be the most important. They also collected the personnel flow information from Wuhan to other cities according to the information of bus, train, and plane. The data confirmed their inference. In conclusion, they found the number of COVID-19 cases is still increasing greatly by the study deadline. Meanwhile, they found an obvious geographic discrepancy on the cases over time about the transmission character [4].

- The research conducted by Fábio A.M. Cássaro and Luiz F. Pires entitled “Can we predict the occurrence of COVID-19 cases? Considerations using a simple model of growth”. This study aimed to present a simple model to follow the evolution of the COVID-19 (CV-19) pandemic in different countries. The cumulative distribution function (CDF) and its first derivative were employed for this task. The simulations showed that it is almost impossible to predict based on the initial CV-19 cases (1st 2nd or 3rd weeks) how the pandemic will evolve. However, the results presented here revealed that this approach can be used as an alternative for the exponential growth model, traditionally employed as a prediction model, and serve as a valuable tool for investigating how protective measures are changing the evolution of the pandemic [5].
- The research conducted by B. Malavika et al. entitled “Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models”. It says The Logistic growth curve model predicts accurately the short-term scenario for India and high incidence states. The prediction through the SIR model may be used for planning and prepare the health systems. The study also suggests that there is no evidence to conclude that there is a positive impact of lockdown in terms of reduction in new cases [6].
- The research conducted by Kenji Mizumoto and Gerardo Chowell entitled “Transmission potential of the novel coronavirus (COVID-19) on board the Diamond Princess Cruises Ship, 2020”. It says an outbreak of COVID-19 developed aboard the Princess Cruises Ship during January-February 2020. Using mathematical modelling and time-series incidence data describing the trajectory of the outbreak among passengers and crew members, we characterize how the transmission potential varied throughout the outbreak. Our estimate of the mean reproduction number in the confined setting reached values as high as ~ 11 , which is higher than mean estimates reported from community-level transmission dynamics in China and Singapore (approximate range: 1.1-7). Our findings suggest that R_t decreased substantially compared to values during the early phase after the Japanese government implemented an enhanced quarantine control. Most recent estimates of R_t reached values largely below the epidemic threshold, indicating that a secondary outbreak of the novel coronavirus was unlikely to occur aboard the Diamond Princess Ship [7].
- The research conducted by Navid Feroze entitled “Forecasting the patterns of COVID-19 and Causal Impacts of Lockdown in Top Ten Affected Countries using Bayesian Structural Time Series Models”. They achieved better levels of accuracy as compared to ARIMA models. The forecasts for the next 30 days suggest that India, Brazil, USA, Russia, and the UK are expected to have a 101.42%, 85.85%, 46.73%, 32.50%, and 15.17% increase in the number of confirmed cases, respectively. On the other hand, there is a chance of 70.32%, 52.54%, 45.65%, 19.29%, and 18.23% growth in the death figures for India, Brazil, Russia, the USA, and the UK, respectively [8].
- The research conducted by Mohsen Maleki et al. entitled “Time series modelling to forecast the confirmed and recovered cases of COVID-19”. Autoregressive time series models based on two-piece scale mixture normal distributions, called TP-SMN-AR models, is a flexible family of models involving many classical symmetric/asymmetric and light/heavy tailed autoregressive models. In this paper, they use this family of models to analyse the real world time series data of confirmed and recovered COVID-19 cases [9].
- The research conducted by Ivorra et al. entitled “Mathematical modelling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China”. We show a good agreement between the reported data and the estimations given by our model. We also study the behaviour of the outputs returned by our model when considering incomplete reported data (by truncating them at some dates before and after the peak of daily reported cases). By comparing those results, we can estimate the error produced by the model when identifying the parameters at the early stages of the pandemic. Finally, taking into account the advantages of the novelties introduced by our model, we study different scenarios to show how different values of the percentage of detected cases would have changed the global magnitude of COVID-19 in China, which can be of interest for policy makers [10].
- The research conducted by Sheng Zhang et al entitled “Estimation of the reproductive number of Novel Coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess Cruise ship: A data-driven analysis”. The Maximum-Likelihood (ML) value of R_0 was 2.28 for the COVID-19 outbreak at an early stage on the ship. The median with 95% confidence interval (CI) of R_0 values was 2.28 (2.06-2.52)

estimated by the bootstrap resampling method. The probable number of new cases for the next ten days would gradually increase, and the estimated cumulative cases would reach 1514 (1384-1656) on the tenth day in the future. However, if R0 value was reduced by 25% and 50%, the estimated total number of cumulative cases would be reduced to 1081 (981-1177) and 758 (697-817), respectively [11].

- The research conducted by Piu Samui et al. entitled “A mathematical model for COVID-19 transmission dynamics with a case study of India”. Based on estimated data, our model predicts that about 60 days the peak will be higher for COVID-19 in India, and after that, the curve will plateau but the coronavirus diseases will persist for a long time [12].

II. THE OBJECTIVES OF RESEARCH

The purpose of this study is to determine the model of infection probability model from human to human in Indonesian regions based on the data released by the Indonesian government through the website mentioned.

III. RESEARCH METHODOLOGY

This research begins with the identification and formulation of the problem, the study of literature from several studies that have been done before, searching and collecting data, making transmission distribution models with the help of software, and then analyzing and finally drawing conclusions. Broadly speaking, the stages of research can be illustrated in figure 1.

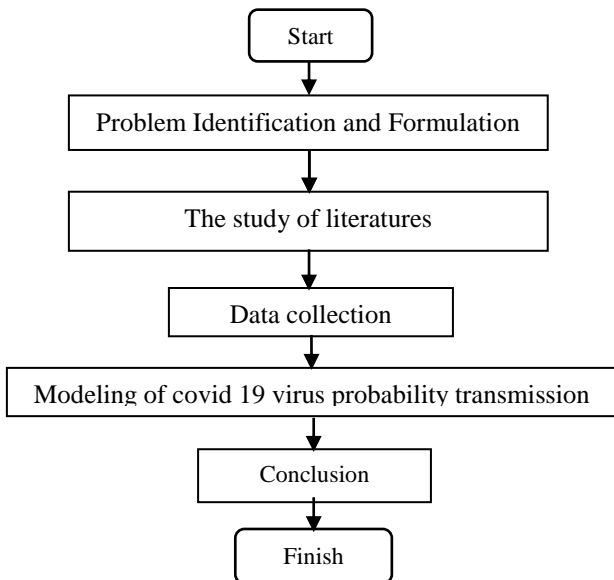


Fig. 1. Flowchart of research stages.

A. Problem Identification and Formulation

The background of the research is the initial reference of the stages of identifying the problems that occur, namely in

modeling the transmission of the COVID-19 virus. Problem identification is intended to find out the core problems that occur so that they are formulated into several points which are the goals or targets of the research to be conducted.

B. Study of Literature

Literature studies are intended to obtain various kinds of references from various sources such as books, journal papers, or browsing the internet to support the completion of this research. A summary of basic concepts and appropriate methods is used as a reference for this research. Also, this stage is carried out to support the achievement of the goal of this study and problem solving with a theoretical approach that fits the research topic. A literature study includes a literature study and a review of previous research.

C. Data Collection

Data collection techniques are done using secondary data that is the data collected and obtained through second parties in the form of documents. The technique used in data collection is through internet searching or direct contact with relevant parties. What is meant by the population in this study are all people who are exposed to the coronavirus in Indonesia. And the sample is some people in the population. This study uses a sampling design that is taking data samples of all people exposed to the coronavirus in Indonesia and the analysis process using statistical software.

D. Modeling of the Transmission of Coronavirus

Modeling of coronavirus transmission is carried out with several steps, namely concept the model used, collecting the data of infected patients in the period in the affected areas, analyzing data, determining the model that matches the data to be applied for modeling and the final result is the model (distribution) of the number of people infected by the coronavirus (See Figure 2).

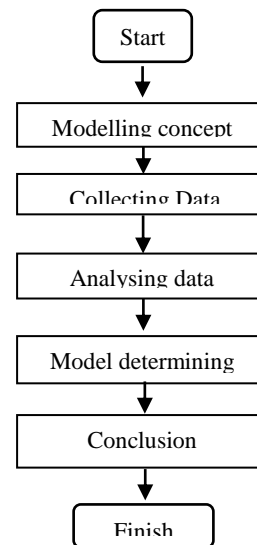


Fig. 2. Modeling of coronavirus transmission steps.

IV. DATA ANALYSIS AND SIMULATION RESULTS

A. Data Collection

The data needed in this study is data on the number of people who are infected by the COVID-19 virus in each province obtained from data released by the Indonesian government which is recorded in (www.covid-19.go.id) which began on March 1th to June 30th, 2020.

B. Data Analysis

Data processing is done using easyfit-software by inputting the data of people who have contracted with coronavirus in Indonesia and looking for the model that best matches the data.

1) *Data and descriptive statistics:* The Last Data Recorded in 123 days of observation, from march to june 2020, the total number of people get infected by the virus was 56385 people shown by figure 3, and descriptive statistic of data which have been collected can be seen in table 1.

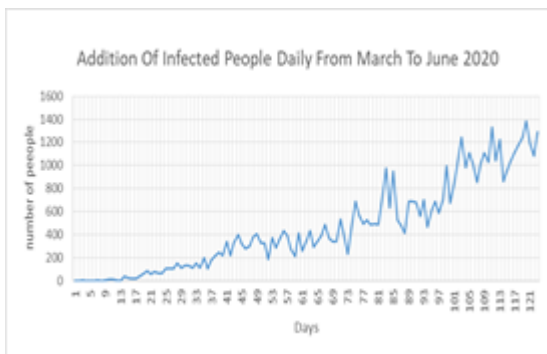


Fig. 3. Data recorded in 122 days of observation, from March to June 2020.

TABLE I. DESCRIPTIVE STATISTICS TABLE OF THE NUMBER OF PEOPLE PER DAY INFECTED WITH THE CORONAVIRUS (COVID-19)

Descriptive Statistics	
Statistic	Value
Sample Size	122
Range	1385
Mean	462,17
Variance	1,4658E+5
Std. Deviation	382,86
Coef. Of Variation	0,82839
Std. Error	34,662
Skewness	0,70412
Excess Kurtosis	-0,56799
Percentile	Value
Min	0
5%	0,3
10%	14,2
25% (Q1)	125,25
50% (Median)	377,5
75% (Q2)	687,5
90%	1072,7
95%	1221,8
Max	1385

2) *Probability distribution modelling:* To find out the number of people who are infected by the virus from a population, it is necessary to know the data about people who have been confirmed positive from the virus and the policies

implemented set by the Indonesian government in the handling of coronavirus outbreaks.

To find out the distribution pattern of the number of infected people daily, Easyfit 5.2 software is used. The results show that the distribution of the number of people infected daily in Indonesia is Johnson SB distribution in which the histogram plot and probability distribution are shown in figures 4 and 5.

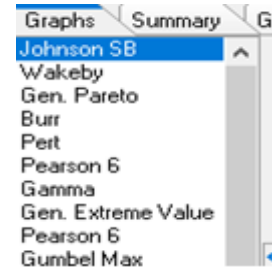


Fig. 4. Rank of the model created by easyfit software.

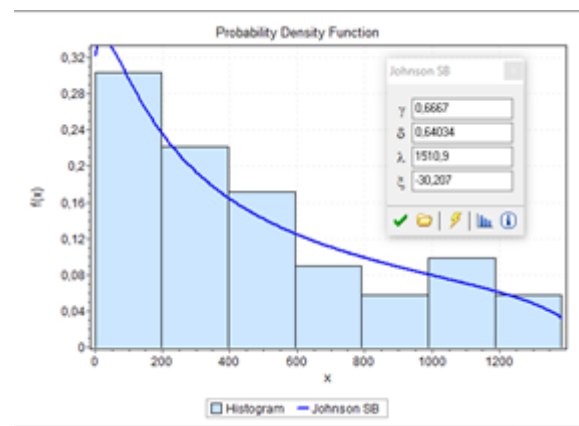


Fig. 5. Johnson SB distribution histogram and its probability distribution.

By using Kolmogorov-Smirnov and Anderson darling fitting test method the result shows that Johnson SB distribution with parameters $\gamma=0,67026$, $\delta=0,63814$, $\lambda=1512,6$, $\xi=-32,765$, is the best rank in the distribution fitting shown by the software.

V. RESULTS AND DISCUSSION

The result shows that the Johnson-SB Distribution is the best model to fit the data taken from the Period of March to June 2020. The model of infection probability from human to human in Indonesia caused by Corona Virus during the period mentioned earlier is Johnson-Sb Distribution but other candidates (wakeby, gen Pareto, Burr, Pert, Pearson 6, and Gamma) still have the chance to be the best model to suit the situation happening on the real situation. So, further discussion is still going on to find the perfect model for what happens in the field not far from the data we are trying to model.

VI. CONCLUSION

The result of this study shows that the best model found is Johnson SB distribution. This model hopefully can contribute to handling the pandemic such as helping the government to predict the possible additional number of people get infected daily by the virus and make anticipation steps to cure them. Besides, this study is also aimed to introduce another way to predict the number of people get infected by the COVID-19 that has been introduced by previous researches.

ACKNOWLEDGMENT

I Would Like To Express My Special Thanks Of Gratitude To the State Polytechnics Of Bengkalis by Making This Research Realized.

REFERENCES

- [1] Y. Gao, C. Shi, Y. Chen, P. Shi, J. Liu, Y. Xiao, ... and E. Chen, "A cluster of the Corona Virus Disease 2019 caused by incubation period transmission in Wuxi, China," *Journal of Infection*, vol. 80, no. (6), pp. 666-670, 2020.
- [2] I. Chakraborty and P. Maity, "COVID-19 outbreak: Migration, effects on society, global environment and prevention," *Science of the Total Environment*, vol. 728, p. 138882, 2020.
- [3] D.K. Mamo, "Model the transmission dynamics of COVID-19 propagation with public health intervention," *Results in Applied Mathematics*, vol. 7, p. 100123, 2020.
- [4] D. Wang, M. Zhou, X. Nie, W. Qiu, M. Yang, X. Wang, ... and W. Chen, "Epidemiological characteristics and transmission model of Corona Virus Disease 2019 in China," *The Journal of infection*, vol. 80, no. (5), p. e25, 2020.
- [5] F.A. Cássaro and L.F. Pires, "Can we predict the occurrence of COVID-19 cases? Considerations using a simple model of growth," *Science of the Total Environment*, vol. 728, p. 138834, 2020.
- [6] B. Malavika, S. Marimuthu, M. Joy, A. Nadaraj, E.S. Asirvatham, and L. Jeyaseelan, "Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models," *Clinical Epidemiology and Global Health*, vol. 9, pp. 26-33, 2021.
- [7] K. Mizumoto and G. Chowell, "Transmission potential of the novel coronavirus (COVID-19) onboard the diamond Princess Cruises Ship, 2020," *Infectious Disease Modelling*, vol. 5, pp. 264-270, 2020.
- [8] N. Feroze, "Forecasting the patterns of COVID-19 and causal impacts of lockdown in top five affected countries using Bayesian Structural Time Series Models," *Chaos, Solitons & Fractals*, vol. 140, p. 110196, 2020.
- [9] M. Maleki, M.R. Mahmoudi, D. Wraith, and K.H. Pho, "Time series modelling to forecast the confirmed and recovered cases of COVID-19," *Travel medicine and infectious disease*, vol. 37, p. 101742, 2020.
- [10] B. Ivorra, M.R. Ferrández, M. Vela-Pérez, and A.M. Ramos, "Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China," *Communications in nonlinear science and numerical simulation*, vol. 88, p. 105303, 2020.
- [11] S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, and D. Chen, "Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis," *International journal of infectious diseases*, vol. 93, pp. 201-204, 2020.
- [12] P. Samui, J. Mondal, and S. Khajanchi, "A mathematical model for COVID-19 transmission dynamics with a case study of India," *Chaos, Solitons & Fractals*, vol. 140, p. 110173, 2020.