

Modeling of COVID-19 Epidemic Growth Curve in Indonesia

Muhammad Fajar¹, Wahyudi²

^{1,2}Badan Pusat Statistik-Statistics Indonesia, Indonesia

Article history:

Received Jun 26, 2020

Revised Mar 8, 2021

Accepted May 30, 2021

Kata Kunci:

Pemodelan, Logistik,
Gompertz, Epidemik,
COVID-19

Keywords:

Modeling, Logistic,
Gompertz, Epidemic,
COVID-19

Abstrak. Tujuan studi ini adalah untuk melakukan pemodelan parametrik terhadap kurva pertumbuhan epidemi COVID-19 sehingga dapat diperoleh nilai maksimum dan waktu pada titik tersebut dari kumulatif kasus COVID-19. Sumber data yang digunakan adalah data jumlah kumulatif kasus terkonfirmasi positif COVID-19 yang berasal dari <https://covid19.go.id/>. Metode yang digunakan dalam penelitian ini adalah *fitting* data dengan model Logistik dan Gompertz. Hasil penelitian ini adalah (1) model Logistik dan Gompertz memiliki tingkat akurasi tinggi (konteks *fitting* data) dalam memodelkan kurva pertumbuhan epidemi COVID-19, diindikasikan dari nilai R^2 (koefisien determinasi) yang mencapai lebih dari 99%; (2) Dari model Logistik diperoleh bahwa estimasi jumlah kasus kumulatif maksimum pada akhir epidemi COVID-19 adalah 7.714 kasus konfirmasi positif yang dicapai dalam waktu sekitar 82 hari (22 Mei 2020) dari tanggal 2 Maret 2020 ketika kasus positif COVID-19 pertama diumumkan oleh pemerintah; dan (3) Dari model Gompertz diperoleh bahwa estimasi jumlah kasus kumulatif maksimum pada akhir epidemi COVID-19 adalah 33.975 kasus konfirmasi positif yang dicapai dalam waktu sekitar 152 hari (30 Juli 2020) dari tanggal 2 Maret 2020. Temuan ini dapat dijadikan bahan input pembuatan langkah pemerintah dalam mengendalikan penyebaran COVID-19.

Abstract. Aim of this study is to make parametric modeling of the COVID-19 epidemic growth curve so that the maximum value and time at that point can be obtained from the cumulative cases of COVID-19. The data used in this study is the cumulative number of positive confirmed cases of COVID-19 from <https://covid19.go.id/>. The method used in this study is fitting data with the Logistic and Gompertz models. Result of this study are (1) the Logistic and Gompertz models are very fit in modeling the COVID-19 epidemic growth curve, indicated from the value of R^2 (coefficient of determination) which reaches more than 99%; (2) From the Logistics model it is obtained that the estimated amount of the maximum cumulative case at the end of the COVID-19 epidemic is 7,714 positive confirmed cases, achieved in about 82 days (May 22, 2020) from Mar 2, 2020, when the first positive COVID-19 case was announced by the government; and (3) From the Gompertz model, it is obtained that the estimated maximum cumulative case at the end of the COVID-19 epidemic is 33,975 positive confirmed cases, achieved in about 152 days (Jul 30, 2020) from Mar 2, 2020. The results of this study can be used as input to the government to take steps in controlling the spread of COVID-19.

How to cite:

M. Fajar and Wahyudi, "Modelling of Covid-19 Epidemic Growth Curve in Indonesia", *J. Mat. Mantik*, vol. 7, no. 1, pp. 67-73, May 2021.

CONTACT:

Muhammad Fajar,  mfajar@bps.go.id  Badan Pusat Statistik-Statistics Indonesia, Jakarta 10710, Indonesia



1. Introduction

Novel Coronavirus 2019 (COVID-19) is an infectious disease caused by SARS-CoV-2, one of the coronavirus types [5]. This virus can cause Pneumonia, which is inflammation of the lung tissue causing interference with oxygen exchange, resulting from breathing, becoming congested. Patients with COVID-19 can experience fever, dry cough, and breathing difficulties.

In Indonesia, COVID-19 has spread to 32 provinces. As of Apr 12, 2020, the cumulative number of positive confirmed cases was recorded at 4,241 cases, with the details of as many as 3,509 patients in the treatment, 359 patients recovered, and 373 patients died [13]. A total of 80.22% of COVID-19 positive cases in Indonesia came from Java Island. DKI Jakarta is the province with the most COVID-19 case found, with 2,044 cases of positive confirmed cases, 1,707 patients in the treatment, 142 patients cured, and 195 died [13].

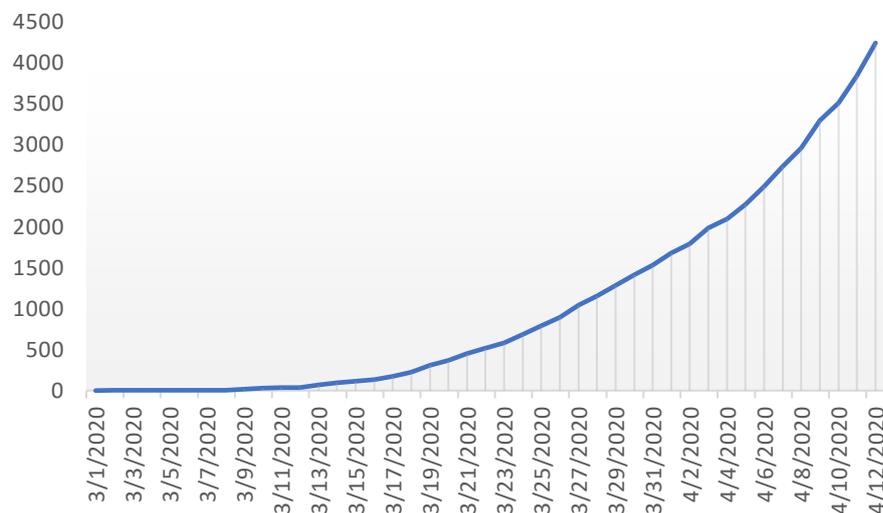


Figure 1. Number of cumulative cases of COVID-19 in Indonesia, as of Apr 12 2020

Figure 1 visually presents the cumulative cases of COVID-19 or also called the epidemic curve. When we look at Figure 1, the cumulative cases of COVID-19 continue to increase over time, and the pattern formed is nonlinear. Generally, on the epidemic growth curve, the movement of the data moves nonlinearly to reach a point where the cumulative case number will not increase again (Stable data movement) which is visually shaped in a horizontal line pattern.

To explore more detailed information on these curves, modeling is needed. Modeling in epidemiology [1,2,4,7,9,11,14,15,16] is commonly used to estimate parameters, such as how large the maximum cumulative case can be achieved by an epidemic [2,10] and the epidemic's growth rate and the point of time when the maximum cumulative case is reached. Estimations of these parameters can be used as an input for the formation of a mitigation plan for the Government.

Therefore, in this study, the authors modeled the COVID-19 epidemic curve with a parametric model (analogy based on references [11,12,17]), namely the Logistics and Gompertz functions. Considerations of using these parametric models include: (1) can predict the time point when reaching the maximum asymptotic peak point in the cumulative

data of a disease (in this case is COVID-19), (2) can predict how large the maximum cumulative COVID-19 case is achieved, (3) Ease and calculation speed in parameters estimation process is easier and faster than Generalized Logistic model, and (4) unlike the SIR model (Susceptible, Infectious, Recovered) which is commonly used for studies of disease transmission (epidemics) [5, 14] and requires assumptions, among others [8]: (a) the population is constant, (b) the birth rate is equal to the death rate in the population, (c) changes in susceptible and infected individuals are proportional to the population, (d) infected individuals assumed to be able to recover with constant chance over time, (e) constant speed in terms of transmission and recovery, and (f) it is assumed that once the person has been infected and then has recovered, then that person will not be re-infected due to strong immunity, the Logistic and Gompertz models do not require these assumptions. The difference in the length of time in reaching the inflection points of the two models can be used as the basis for an optimistic and pessimistic scenario in controlling COVID-19 with the assumption that there is no intervention, meaning that the growth process of the COVID-19 epidemic runs naturally without intervention. The implication is for stakeholders to manage the risk of an ongoing epidemic so that they can make the right steps.

2. Methods

2.1 Data Source

The data used in this study is daily data on the cumulative number of COVID-19 positive confirmed cases in Indonesia as of Apr 12 2020, sourced from www.covid19.go.id [13].

2.2 Growth Model

The cumulative growth of COVID-19 cases (by analogy as referenced [11, 12, 17]) indicates a phase, in which the growth rate starts from zero at $t = 0$, and then accelerates to μ_m within a certain time interval of λ . When heading to the final phase of the epidemic, growth rates are decreasing to reaching zero (it means the rate of addition of new cases is very low so that the number is approaching zero even reaches zero, the implications of the cumulative case data plot will be horizontally shaped), and the number of final cumulative cases is achieved asymptotically (maximum, a).

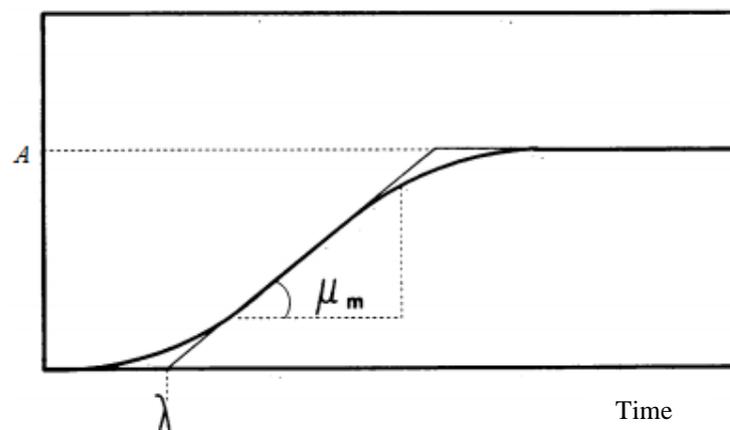


Figure 1. Epidemic Growth Curve

In plotting¹ data on the cumulative number of COVID-19 cases, the authors used the parametric model as follows:

a. Logistics Model [11,12,17]

$$y_t = \frac{A}{1 + \exp\left(\frac{4\mu_m}{A}(\lambda - t) + 2\right)}, \quad (1)$$

where y_t is the cumulative number of COVID-19 cases at time t , A is the maximum cumulative case at the end of the epidemic, μ_m is the maximum growth rate, λ is the time point when it reaches μ_m . When $y_t = A/2$, the time needed is based on equation (1):

$$t_{\frac{A}{2},logistik} = \lambda + \frac{A}{2\mu_m} \quad (2)$$

From equation (2) it can be obtained the time needed to reach A asymptotically is:

$$t_{A,logistik} = 2t_{\frac{A}{2},logistik} \quad (3)$$

b. Model Gompertz [11,12,17]

$$y_t = A \exp\left(-\exp\left(\frac{\mu_m e}{A}(\lambda - t) + 1\right)\right), \quad (4)$$

where e is a natural number (2.718...). When $y_t = A/2$, the time needed is based on equation (4):

$$t_{\frac{A}{2},gompertz} = \lambda - \frac{A(\ln(\ln 2) - 1)}{\mu_m e} \quad (5)$$

From equation (5) it can be obtained the time needed to reach A asymptotically, is:

$$t_{A,gompertz} = 2t_{\frac{A}{2},gompertz} \quad (6)$$

In these two models, there are parameters A , μ_m , and λ which must be estimated. In the estimation process, this study uses R software with a nonlinear square estimation algorithm [3].

3. Results and Discussions

The following Figure 2 is a visual presentation of the results of data plotting with the model proposed by the author in the previous section.

¹ Plotting is done based on the origin O.

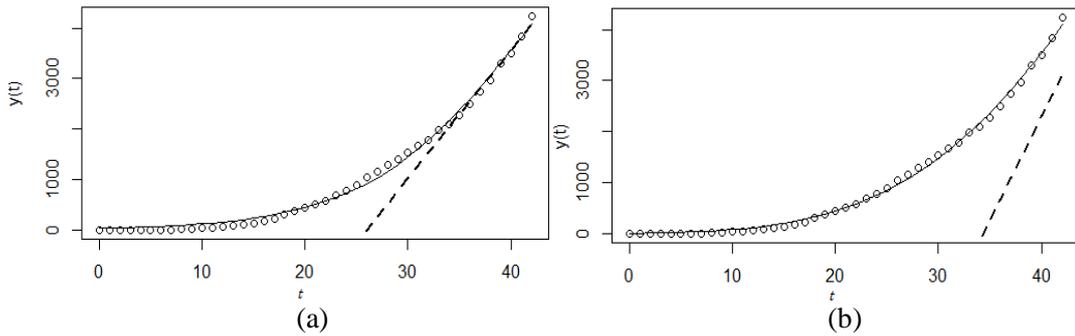


Figure 2. (a) Fitting the plot data with the Logistic model, (b) Fitting the plot data with the Gompertz model

Notes: o states the actual data, — states the fitting model, and - - - states the slope of the model

From Figure 2, visually, both models are accurate in plotting data patterns. This is indicated by the value of R^2 (coefficient of determination) generated by the Logistic and Gompertz models which are 99.571% and 99.789%, respectively. That is, variations of response variables (in this case the cumulative COVID-19 case) can be explained by both models by more than 99%, while the remaining less than 1% are explained by other variables outside the model.

By using the nonlinear least square estimation mechanism, the estimation results obtained are presented in table 1, as follows,

Table 1. The estimated results of the Logistics Model and Gompertz which are visualized in Figure 2.

No	Model	R^2	A	μ_m	λ	$\frac{t_A}{2}$	t_A
1	Logistic	99.571%	7,713.719 [6125.587, 9301.852]	253.860 [220.735, 286.984]	25.961 [24.213, 27.709]	41.154	82.307
2	Gompertz	99.789%	33,975.144 [16247.870, 51702.420]	409.535 [269.645, 549.426]	34.356 [28.378, 40.335]	76.062	152.123

Note: 95% confidence interval [lower limit, upper limit]

Based on Table 1, there are few information that can be obtained, namely: (1) From the Logistics model it is found that the estimated maximum cumulative cases at the end of the COVID-19 epidemic is 7,714 positive confirmed cases achieved in about 82 days (May 22, 2020) from Mar 2, 2020, when the first positive COVID-19 case was announced by the Government and the maximum growth rate achieved was the slope of the model, which was 253.860% (it means when the $\lambda = 25.961$ days from the start of the COVID-19 epidemic began, then the maximum growth rate reached by the epidemic was 253.860%). (2) From the Gompertz model, it is found that the estimated maximum cumulative cases at the end of the COVID-19 epidemic is 33,975 positive confirmed cases achieved in about 152 days (Jul 30, 2020) from Mar 2, 2020 when the first positive COVID-19 case was announced by the government and the maximum growth rate achieved is the slope of the model, which is 409.535% (that is, when the time $\lambda = 34.356$ days from the start of the COVID-19 epidemic began, the maximum growth rate achieved by the epidemic is 409.535%).

Based on the results of the estimated maximum cumulative number and time² needed to reach the maximum cumulative case from Table 1, it can be considered that the Logistic model is an optimistic scenario. In contrast, the Gompertz model is a pessimistic scenario, because: (1) the estimated maximum cumulative case achieved by COVID-19 from the Logistic model is lower than the results of the Gompertz model, and (2) the estimated time from the Logistic model to reach the maximum cumulative case is almost twice as fast as the time obtained from the Gompertz model.

Therefore, this information can be used as input for the COVID-19 control policy. For example, according to an optimistic scenario (Logistics model), it is predicted that the peak of the COVID-19 epidemic occurred on May 22, 2020 (about 1 week before Idul Fitri) with a cumulative number of 7,714 cases, so the Government has to do a lockdown policy in major cities throughout provinces in Java (not only in DKI Jakarta) by prohibiting people returning home to celebrate Lebaran and encouraging people to stay at home to reduce the number of positive cases does not reach the maximum point. If government authorities assume that the growth of the COVID-19 epidemic follows the Gompertz model, then the Government only needs to implement social distancing instead of lockdown or quarantine.

4. Conclusions

Based on the results of the discussion, it can be concluded that: (1) Logistics and Gompertz models are very fit in modeling the COVID-19 epidemic growth curve, indicated from the value of R^2 (coefficient of determination) which reaches more than 99%, (2) From the Logistic model it is obtained that the estimation the maximum cumulative cases at the end of the COVID-19 epidemic is 7,714 positive confirmed cases which are achieved in about 82 days (May 22, 2020) from Mar 2, 2020, when the Government announced the first positive COVID-19 case. (3) From the Gompertz model, it is obtained that the estimated maximum cumulative cases at the end of the COVID-19 epidemic is 33,975 positive confirmed cases which are achieved in about 152 days (Jul 30, 2020) from Mar 2, 2020, when Government announced the first positive COVID-19 case. From this finding, the results of this study can be used as input to the Government to take steps in controlling the spread of COVID-19.

References

- [1] Ahmadi A, Fadai Y, Shirani M, Rahmani F. Modeling and forecasting trend of COVID-19 epidemic in Iran. 2020. DOI: <https://doi.org/10.1101/2020.03.17.20037671>.
- [2] Arino J, Brauer F, Van Den Driessche P, Watmough J, Wu J. A final size relation for epidemic models. *Math Biosci Eng.* 2007;4:159. DOI: 10.3934/mbe.2007.4.159.
- [3] Bates D M, Watts D G. *Nonlinear Regression Analysis and Its Applications*, Wiley. 1988. DOI: 10.1002/9780470316757.

² The maximum cumulative number of cases and the length of time to reach the cumulative number of cases can be used as a proxy indicator of whether an attempt to control an outbreak is effective or ineffective.

- [4] Datolli G, Di Palma E, Licciardi S, Sabia E. A note on the Evolution of Covid-19 in Italy. 2020. <https://arxiv.org/pdf/2003.08684.pdf>.
- [5] Gorbalenya A E, Baker S C, Baric R S, de Groot R J, Drosten C, et al. 2020. Severe acute respiratory syndrome-related coronavirus: The species and its viruses - a statement of the Coronavirus Study Group. *bioRxiv*. 2020. DOI: <https://doi.org/10.1101/2020.02.07.937862>.
- [6] Gao D, Lou Y, He D, Porco TC, Kuang Y, Chowell G, et al. Prevention and control of Zika as a mosquito-borne and sexually transmitted disease: a mathematical modeling analysis. *Sci Rep*. 2016;6:28070. DOI: 10.1038/srep28070.
- [7] He D, Gao D, Lou Y, Zhao S, Ruan S. A comparison study of Zika virus outbreaks in French Polynesia, Colombia and the State of Bahia in Brazil. *Sci Rep*. 2017;7:273. DOI: 10.1038/s41598-017-00253-1.
- [8] Jones J. H. 2007. Notes on R_0 . <https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf>.
- [9] Jia L, Li K, Jiang Y, Guo X, Zhao T. Prediction and analysis of Coronavirus Disease 2019. 2020. <https://arxiv.org/ftp/arxiv/papers/2003/2003.05447.pdf>.
- [10] Ma J, Earn DJ. The generality of the final size formula for an epidemic of a newly invading infectious disease. *Bull Math Biol*. 2006;68:679–702. DOI: 10.1007/s11538-005-9047-7.
- [11] Razzak, W.A. Modelling New Zealand COVID-19 infection rate, and the efficacy of social distancing policy. *Economics discussion paper 20.04*. 2020. <http://econfin.massey.ac.nz/school/publications/discuss/2020/DP2004.pdf>.
- [12] Tjørve KMC, Tjørve E. The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. *PLoS ONE*. 2017; 12(6): e0178691. <https://doi.org/10.1371/journal.pone.0178691>.
- [13] www.covid19.go.id. Diakses 12 April 2020 pukul 18:00 WIB.
- [14] Zhang Q, Sun K, Chinazzi M, Pastore y Piontti A, Dean NE, Rojas DP, et al. Spread of Zika virus in the Americas. *Proc Natl Acad Sci USA*. 2017;114:E4334–43. DOI: 10.1073/pnas.1620161114.
- [15] Zhao S, Chen H. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. *Quant Bio*. 2020, 8(1): 11–19. DOI: <https://doi.org/10.1007/s40484-020-0199-0>.
- [16] Zhao S, Stone L, Gao D, He D. Modelling the large-scale yellow fever outbreak in Luanda, Angola, and the impact of vaccination. *PLoS Negl Trop Dis*. 2018;12:e0006158. DOI: 10.1371/journal.pntd.0006158.
- [17] Zwietering M, Jongenburger I, Rombouts F, van 't Riet K. 1990). Modeling of the Bacterial Growth Curve. *Appl & Env Micr*. 1990; 56:1875. DOI: 10.1128/AEM.56.6.1875-1881.1990.