

Analysis on the Distribution of COVID-19 Cases by Country — 2. Robust Power-Law Distribution Observed during the First Half of 2021 —

Hidetaka TOBITA*

(Received August 20, 2021)

Part 1 of this series^[1] analyzed the global cases of COVID-19 during April through November, 2020. In this second report, the data from November 14, 2020 to August 7, 2021 are used for the analysis. Four important findings on the rank-size relationship reported in Part 1 are reconfirmed. (1) The top countries approximately follow the power-law distribution, with $rank \propto size^{-\alpha}$. (2) The overall distribution can be represented by the lognormal distribution. (3) These two characteristics are reproduced reasonably well by the discrete-time stochastic model that employs the weekly reproduction rate μ . (4) The distribution of μ can be approximated by the gamma distribution. The μ -distribution determined for the time period of November 14, 2020 through May 22, 2021 predicts that the stationary distribution after a long time leads to form the power-law tail with $\alpha = 1.24$, while the actual data show $\alpha = 0.9$ during the whole period. This persistent characteristic is supported by the Monte Carlo simulation, showing the robustness of the power-law distribution once established. In order to keep the number average size finite, the magnitude of α must be larger than 1, which corresponds the condition with the average of μ smaller than 1. In addition to the cases by country, the distribution of the number of deaths by country is investigated to find that the characteristics described in item (1) and (2) apply.

Key Words : COVID-19, Ranking Plot, Power Law, Lognormal Distribution, Pandemic, Reproduction Rate, Gamma Distribution

1. Introduction

In Part 1 of this series,^[1] the global cases of COVID-19 reported on the website of Center for Systems Science and Engineering at Johns Hopkins University^[2] were used to analyze the distribution of the cases by country, i.e., the rank-size relationship. Because the rank is proportional to the number fraction of the countries whose number of cases is larger than a given size, the rank corresponds to the upper probability distribution $CN(x)$ as follows.

$$rank \propto CN(x) = \int_x^{\infty} N(x) dx. \quad (1)$$

In Eq. (1), $N(x)$ shows the number-based probability density function (pdf), and x is the number of cases in a country. The rank-size relationship directly shows the cumulative distribution function graphically.

It was found^[1] that the power-law relationship, $rank \propto x^{-\alpha}$ applies for the top countries, and α is approximately in the range of 0.7 – 0.9 during the time period of April through November, 2020. The power law for the COVID-19 distribution was also reported by the other researcher.^[3] When the power-law distribution $rank \propto x^{-\alpha}$ applies, Equation (1) leads to give $N(x)$ as follows.

$$N(x) \propto x^{-\alpha-1} \quad \text{for large } x\text{'s}. \quad (2)$$

The power-law distribution is sometimes referred to as scale free, because the average can go infinity. The scale-free power-law distribution is a pathway capable of connecting finite and infinite sizes smoothly, and appears frequently in the theory of phase transition.^[4,5] Because the pandemic is a phenomena that the infection prevails endlessly, it may be related with the scale-free power law distribution.

Although the upper tail probability distribution conforms to the power law, it was found^[1] that the whole distribution, including the lower-ranked countries, can be

* Department of Materials Science and Engineering

represented by the lognormal distribution rather well.

The following simple discrete-time model^[6] was used to examine the data. The increase of cases during a certain week is represented by Δx_t , and the increase in the next week is by Δx_{t+1} . The newly infected would be infectious, and let μ_t be the weekly reproduction rate, defined by:

$$\mu_t = \Delta x_{t+1} / \Delta x_t. \quad (3)$$

The weekly reproduction rate μ corresponds to an approximate representation of the effective reproduction number in epidemiology,^[7] sometimes referred to by using a symbol, R_t .

With this simple model, the number of cases x_t is represented by the following equations.

$$x_{t+1} = x_t + \Delta x_t. \quad (4)$$

$$\Delta x_t = \mu_{t-1} \Delta x_{t-1}. \quad (5)$$

During the time period of April 20 through November 11, 2020, the distribution of μ for the top 120 countries showed that the average of μ is 1.14, which is larger than 1, and the distribution cannot reach the stationary state.^[6] On the other hand, however, the Monte Carlo (MC) simulation showed^[1] that the distribution that conforms to the above-mentioned two different types distribution simultaneously can be formed during the transient period, i.e., the power-law for the upper tail distribution and the lognormal for the whole region.

In this subsequent report, the data were collected every week for the period of Nov. 14, 2020 (0th week, $t = 0$) through Aug. 7, 2021 (38th week, $t = 38$). From May 15, 2021 ($t = 26$), the numbers of deaths by country are also collected.

2. Results and Discussion

2.1 Cases by Country

Fig. 1 and 2 show the rank-size relationship on the designated dates, from $t = 0$ (Nov. 14, 2020) to 27 (May 22, 2021). Fig. 1 shows that the upper tail distribution conforms to the power law, while the whole distribution is represented by the lognormal distribution, as shown in Fig. 2. The power exponent, α was determined by the least square fit of the double logarithmic plot for the top 30 countries. Two characteristics found in Part 1^[1] are reconfirmed.

Fig. 3 shows the distribution of weekly reproduction

rate, μ obtained from the data between $t = 0$ and 27. The distribution can be approximated by the gamma distribution, represented by the following equation.

$$f(\mu) = \frac{\mu^{m-1}}{\Gamma(m)\eta^m} \exp\left(-\frac{\mu}{\eta}\right). \quad (6)$$

Two parameters, m and η are determined to be $m = 16.3$ and $\eta = 0.0609$.

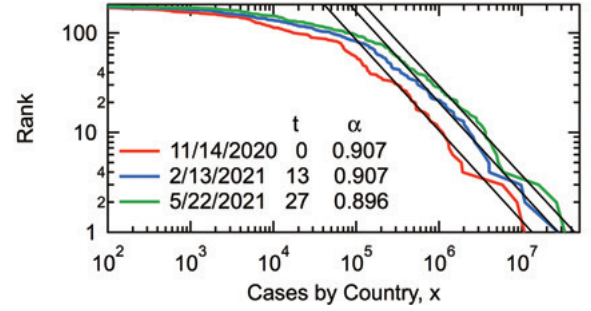


Fig. 1 Relationship between the rank of the country and the number of cases for the country. During this time period, the USA is ranked the first, and $rank = 1$ for the USA.

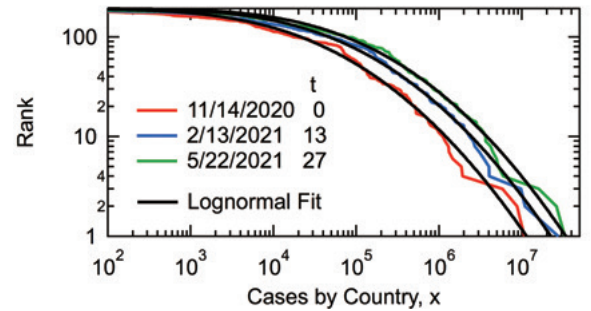


Fig. 2 Fitted curves by the lognormal distribution.

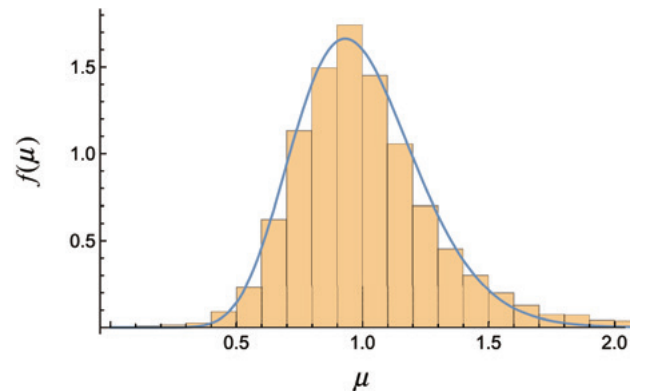


Fig. 3 Distribution of the weekly reproduction rate μ for the time period from $t = 0$ to 27. The solid curve is the fitted gamma distribution.

The average of the gamma distribution is given by

$$\bar{\mu} = m\eta, \quad (7)$$

which leads to $\bar{\mu} = 0.993$ and the distribution converges to the stationary state.

Yamamoto^[6] reported that the present discrete-time model leads to a stationary power-law tail with $rank \propto x^{-\alpha}$, if the following equation has a unique positive solution, α .

$$\int_0^{\infty} \mu^{\alpha} f(\mu) d\mu = 1. \quad (8)$$

Note that the left-hand side represents the expected value of μ^{α} .

For the gamma distribution, Equation (8) can be calculated to give

$$\int_0^{\infty} \mu^{\alpha} f(\mu) d\mu = \frac{\eta^{\alpha} \Gamma(m+\alpha)}{\Gamma(m)} = 1. \quad (9)$$

For the case with $m = 16.3$ and $\eta = 0.0609$, one obtains $\alpha = 1.24$.

To confirm if the stationary state α is equal to 1.24, the MC simulation was conducted. To determine the stationary state α accurately, the number of clusters (i.e., the total number of countries for the present problem) was set to be $N = 10^5$. The initial values used were $x_0 = 0$ and $\Delta x_1 = 1$ for all clusters. The random number that follows the gamma distribution with $m = 16.3$ and $\eta = 0.0609$ were used.

Fig. 4 shows the MC simulation results at $t = 10^3$ (blue) and $t = 10^4$ (red, broken). Note that the y-axis shows the upper probability distribution $CN(x)$, defined by Equation (1). Both curves overlap each other, and the distribution has already reached the stationary state at $t = 10^3$. It is confirmed that the upper tail follows the power-law distribution with $\alpha = 1.24$.

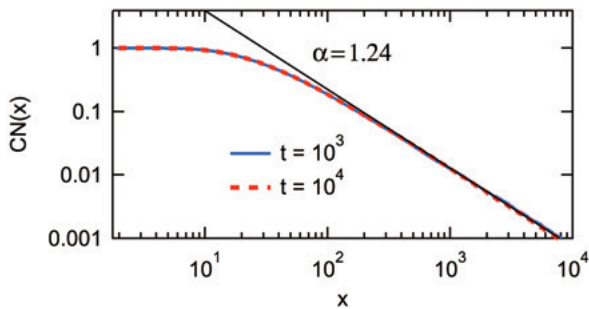


Fig. 4 MC simulation results to confirm the validity of $\alpha = 1.24$ at the stationary state.

In spite of the fact that the stationary distribution has a tail with $\alpha = 1.24$, the α -values shown in Fig. 1 are about 0.9 for all cases. Fig. 5 shows the determined values for each week up to the 27th week. The α -value is quite

stable and is about 0.9 throughout the time period. Does this discrepancy from the stationary state distribution means that the present discrete-time stochastic model cannot be applied? To examine this, another MC simulation was conducted.

The value of x_1 was set to be equal to the actual data at $t = 1$, and Δx_0 was determined from the difference of x_1 and x_0 for each country. The simulation was done for $N = 179$ countries where Δx_0 is larger than 0. The distribution at $t = 27$ was simulated.

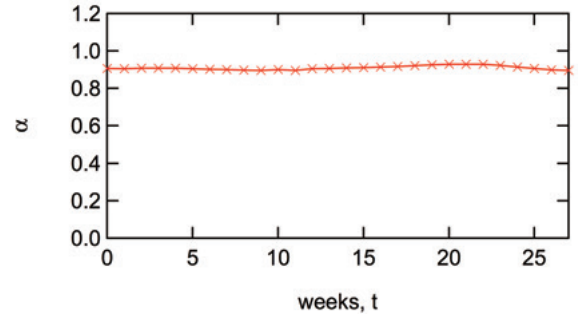


Fig. 5 Obtained α -values from $t = 0$ to 27.

The MC simulation was repeated for 100 times, and Fig. 6 shows all the simulated results, together with the actual data shown by the red line. The data (red) is within the range of 100 simulated results (colored dots).

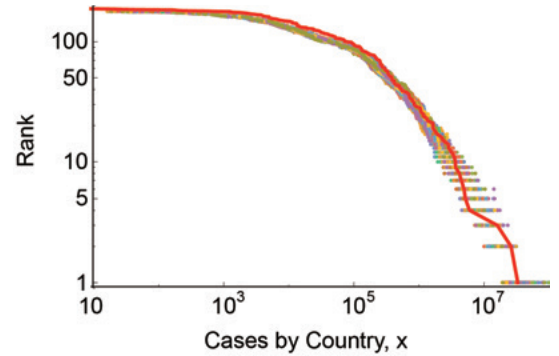


Fig. 6 Data at $t = 27$ (red) and the simulated results (colored dots).

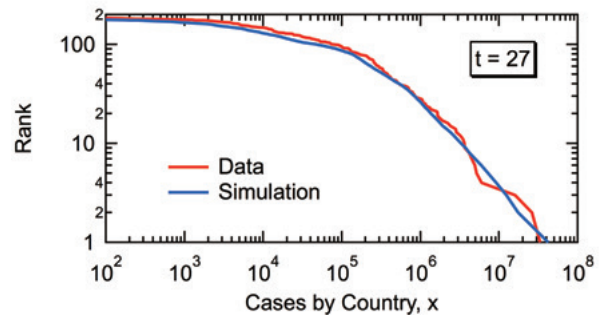


Fig. 7 Comparison between the data at $t = 27$ (red) and the average of simulated results (blue).

The blue line in Fig. 7 shows the average of the simulated results repeated 100 times, which agrees reasonably well with the actual data (red). The simulated results (blue) show two important characteristics: (1) the upper tail follows the power-law distribution with $\alpha = 0.9$, as shown in Fig. 8a, and (2) the whole distribution agrees well with the lognormal distribution, as shown in Fig. 8b.

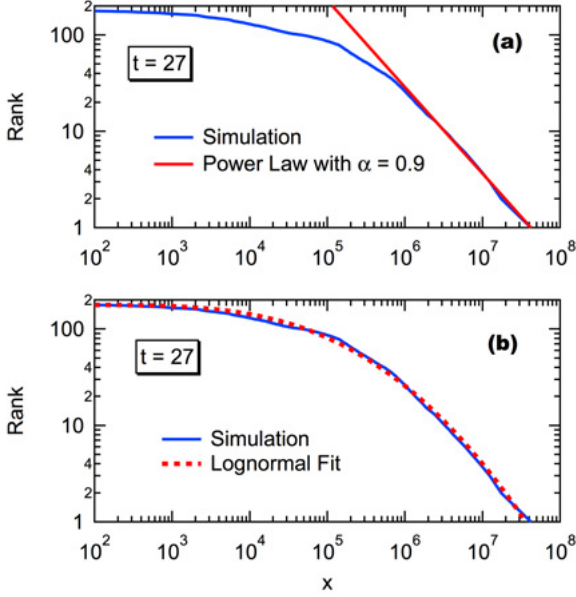


Fig. 8 MC simulation results (blue) at $t = 27$. Red line shows (a) the power law with $\alpha = 0.9$ and (b) the lognormal fit.

Although the convergent stationary distribution is the power-law distribution with $\alpha = 1.24$, as shown in Fig. 4, the MC simulation shows that a stable power-law distribution with $\alpha = 0.9$ holds throughout 27 weeks, which agrees with the actual data.

In order to further examine the robustness of the power law tail with $\alpha = 0.9$, the simulation for $t = 100$ was conducted. As shown in the green line in Fig. 9, the power law tail with $\alpha = 0.9$ is preserved even after 100 weeks. The power law tail is remarkably robust. As shown in Fig. 10, the overall lognormal distribution is also kept after 100 weeks.

Fig. 11 shows the actual data of α 's for the period from $t = 28$ to 38, and it is shown that $\alpha = 0.9$ persists. The power law distribution once formed seems to be rather robust and the power exponent α cannot be changed easily.

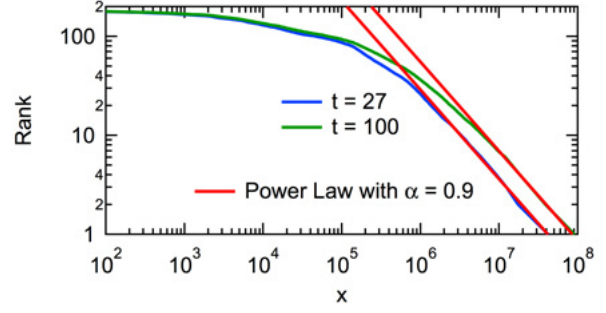


Fig. 9 Simulation results at $t = 27$ and 100. Both distributions possess the power law tail with $\alpha = 0.9$.

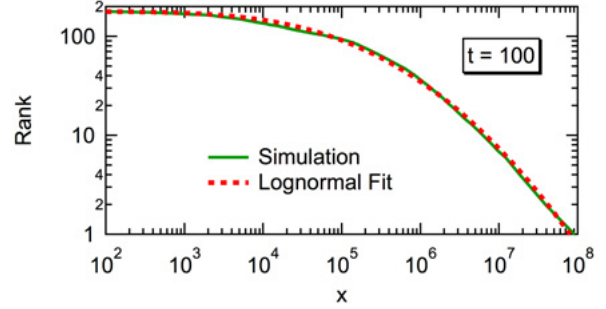


Fig. 10 Lognormal fit for the simulation results at $t = 100$.

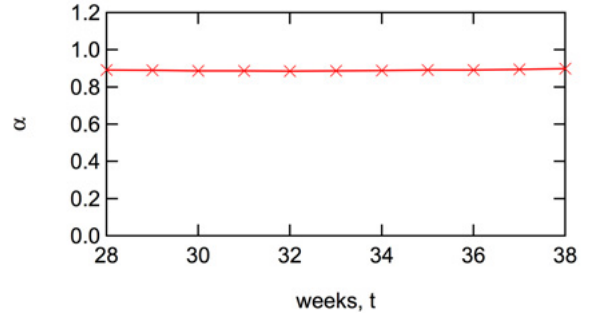


Fig. 11 Obtained α -value for the cases by country during the time period from $t = 28$ to 38.

2.2 Deaths by Country

Starting from May 15, 2021 ($t = 26$), the numbers of deaths by country were also collected. Fig. 12 shows the distribution of the number of deaths at $t = 38$. The distribution shows (1) the power-law tail and (2) good fit to the lognormal distribution. The fundamental characteristics are the same as for the cases by country. The power exponent determined from the top 30 countries is $\alpha = 0.85$ at $t = 38$.

Assuming that the fatality rate is almost the same for all countries, the number of deaths is expected to have the same power-law tail with $\alpha = 0.9$. Fig. 13 shows the power exponent for the deaths by country. The values are compared with those for the cases by country. The

α -values for the deaths ($\cong 0.85$) and the cases ($\cong 0.9$) are almost the same, but the deaths show slightly smaller α 's consistently. In general, smaller α indicates a wider disparity. This may imply the disparity of medical systems among countries.

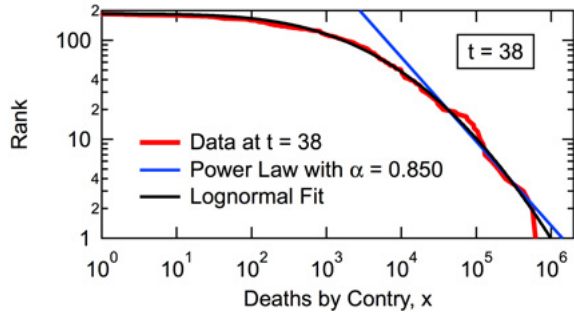


Fig. 12 Relationship between the rank of the country and the number of deaths for the country.

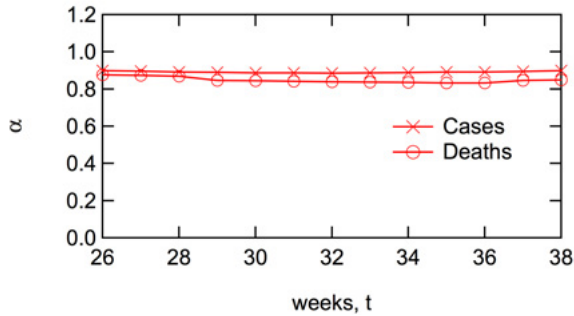


Fig. 13 Obtained α -value for the deaths by country, in comparison with cases by country, during the time period from $t = 26$ to 38.

Fig. 14 shows the fatality rate by country. Fatality rates of several countries are much higher than the average.

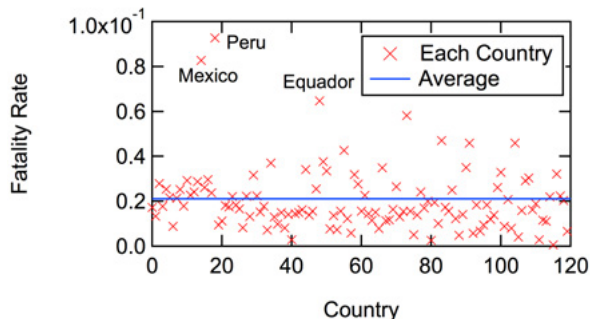


Fig. 14 Fatality rate by country as of August 7, 2021 ($t = 38$). The country order in the x-axis is the rank of the cases by country.

2.3 Power Exponent at the Critical Point

In the standard theory of gelation in polymer

science,^[8] the gel (critical) point is reached when the weight-average molecular weight goes to infinity. Various power-law relationships hold at the vicinity of gel point,^[4,5] which is known as the universality of the critical phenomena.

In order to cause polymeric gelation, the formation of cross-linkage that connects two different clusters is required.^[9] The molecular weight distribution at the gel point follows the power law, with $\alpha \leq 2$ at which the weight average goes to infinity.

On the other hand, in the branched structure of infection tree, no cross-linkage is formed. The elements are connected in one direction, and there exists only one start point in each tree. The expected size of the tree when one picks up such a start point randomly is the number-average size.^[10] For the infection tree, it is reasonable to consider that the critical point is reached when the number average goes to infinity. The number of cases in a country is the sum of a finite number of infection trees, and therefore, the critical point will be the same as for the tree. In the distribution having a power-law tail, the number average goes to infinity when the magnitude of power exponent is $\alpha \leq 1$.

Assuming that the μ -distribution follows the gamma distribution, Equation (8) reduces to Equation (9). The condition that yields $\alpha = 1$ is given by the following equation.

$$\eta m = \bar{\mu} = 1. \quad (10)$$

The number average goes to infinity when the average of μ reaches 1. In fact, Yamamoto^[6] showed that $\alpha = 1$, which is known as the Zipf's law, is obtained when the average of μ is 1, irrespective of the type of μ -distribution. In epidemiology, usual measure to suppress epidemic is to make the effective reproduction number, which corresponds to μ , smaller than 1.^[7] The condition to make $\alpha > 1$ conforms to this usual criterion.

In Part 1 of this series,^[1] the conditions to make the power exponent $\alpha > 2$ was considered, in order to suppress the pandemic. It is now shown that $\alpha > 1$ would be a better criterion for the suppression.

The μ data for $0 < t < 27$ give $m = 16.3$ and $\eta = 0.0609$, as shown in Fig. 3. In this case, $\bar{\mu} = 0.993$ and $\alpha = 1.24$. Because $\bar{\mu} < 1$, the pandemic is expected to be terminated if the present μ -distribution continues. The $\bar{\mu}$ -value ($= 0.993$) is very close to 1, but the α -value ($= 1.24$) of the stationary distribution is significantly larger than 1.

As could be recognized from Fig. 3 the values of $m =$

16.3 and $\eta = 0.0609$ are approximate. Assuming $m = 16.3$, it is straightforward to show that $\eta = 0.05955$ gives $\alpha = 2$ from Equation (9). The values of $m = 16.3$ and $\eta = 0.05955$ leads to $\eta m = \bar{\mu} = 0.971$.

Fig. 15 shows these two types of gamma distribution graphically. The difference is surprisingly small. It would be difficult to distinguish these two curves from the actual data. Considering small differences in the gamma distribution, the prevention criterion proposed in Part 1, based on $\alpha = 2$, might be still acceptable: (1) the average μ must be made smaller than 1 and (2) the time period for $\mu > 2$ must be made negligibly small. Safer condition of item (2) might be “the time period for $\mu > 1.5$ should be controlled to make as small as possible”.

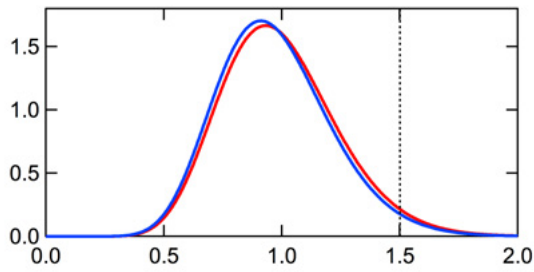


Fig. 15 Two types of gamma distribution. The red curve shows the case with $m = 16.3$ and $\eta = 0.0609$, and the blue curve is the case with $m = 16.3$ and $\eta = 0.05955$.

3. Conclusions

Following the analysis of the global cases of COVID-19 reported in Part 1 of this series,^[1] the data during November 14, 2020 (elapsed weeks, $t = 0$) through August 7, 2021 ($t = 38$) were used for the analysis in this report. Four important findings on the relationship between the rank and the number of cases by country reported in Part 1 are reconfirmed for the present time period. (1) The top countries approximately follow the power-law distribution, with $rank \propto size^{-\alpha}$. (2) The overall distribution is represented well by the lognormal distribution. (3) These two characteristics can be reproduced reasonably well by the discrete-time stochastic model that employs the weekly reproduction rate μ . (4) The distribution of μ can be approximated by the gamma distribution.

The μ -distribution determined for the time period from $t = 0$ to 27 predicts that the stationary distribution after a long time leads to have the power-law tail with $\alpha = 1.24$,

while the actual data show $\alpha = 0.9$ persistently throughout the investigated time period. This persistent α -value agrees with the MC simulation results. The power-law distribution once formed is quite stable, and the power exponent cannot be changed easily.

The number of deaths by country was investigated for the time period of $t = 26$ through 38. The relationship between the rank and the number of deaths shows the similar characteristics with the number of cases. (1) The distribution possesses the power-law tail and (2) the distribution is well fit to the lognormal distribution. The power exponent of the tail distribution was $\alpha = 0.85$, which is slightly smaller than that for the cases by country.

In order to keep the number average size finite, the magnitude of α must be larger than 1, which corresponds the condition with the average of μ smaller than 1. In order to suppress the pandemic, (1) the average μ must be made smaller than 1 and (2) the time period for $\mu > 1.5$ should be controlled to make as small as possible.

References

- [1] H. Tobita and M. Fujii: Mem. Fac. Eng. Univ. Fukui, **69**, 37-42 (2021).
- [2] COVID-19 Dashboard by the Center for Systems Science and Engineering at Johns Hopkins University, <<https://coronavirus.jhu.edu/map.html>>, (2020/04/11-2021/08/07).
- [3] B. Blasius: Chaos, **30**, 093123 (2020).
- [4] W. Gebhardt and U. Krey: Phasenübergänge und kritische Phänomene, Friedr. Vieweg & Sohn, Braunschweig/Wiesbaden, Germany (1982). (Japanese translation, ISBN: 4-8427-0242-7)
- [5] D. Stauffer: Introduction to Percolation Theory, Taylor & Francis, London, UK (1985).
- [6] K. Yamamoto: Phys. Rev. E, **89**, 042115 (2014).
- [7] R.M. Anderson and R.M. May: Infectious Diseases of Humans, Oxford University Press, UK (1991).
- [8] P.J. Flory: Principles of Polymer Chemistry, Cornell University Press, Ithaca, USA (1953).
- [9] H. Tobita: J. Polym. Sci., Part B, Polym. Phys., **36**, 2015-2018 (1998).
- [10] H. Tobita: Macromol. Theory Simul., **5**, 1167-1194 (1996).