Maximum entropy method for estimating the reproduction number: An investigation for COVID-19 in China

Yong Tao[†]

College of Economics and Management, Southwest University, Chongqing, China

Department of Management, Technology and Economics, ETH Zurich, Switzerland

Abstract: The key parameter that characterizes the transmissibility of a disease is the reproduction number R. If it exceeds 1, the number of incident cases will inevitably grow over time, and a large epidemic is possible. To prevent the expansion of an epidemic, R must be reduced to a level below 1. To estimate the reproduction number, the probability distribution function of the generation interval of an infectious disease is required to be available; however, this distribution is often unknown. In this letter, given the incomplete information for the generation interval, we propose a maximum entropy method to estimate the reproduction number. Based on this method, given the mean value and variance of the generation interval, we first determine its probability distribution function and in turn estimate the real-time values of reproduction number of COVID-19 in China. By applying these estimated reproduction numbers into the susceptible-infectious-removed epidemic model, we simulate the evolutionary track of the epidemic in China, which is well in accordance with that of the real incident cases. The simulation results predict that China's epidemic will gradually tend to disappear by May 2020 if the quarantine measures can continue to be executed.

Keywords: Maximum entropy; Reproduction number; Generation interval; COVID-19; Incomplete information

---

[†] Correspondence to: taoyingyong@yahoo.com

In December, 2019, a cluster of pneumonia cases in Wuhan, China was caused by a novel coronavirus, the COVID-19 [1-4]. At first the local governments did not take effective measures which leaded to local people not paying enough attention to the risk. However, with the epidemic in Wuhan further expanding, the Chinese Government started to take emergency actions to lock down the Wuhan city on January 23, 2020. Despite this, the epidemic still spread throughout the entire country. By March 12, 2020, controlling the spread of the epidemic has become a global challenge. One of the key parameters in epidemic models is the basic reproduction number R_0, defined as the number of secondary infections that arise from a typical primary case in a completely susceptible population [5]. As an infection is spreading through a population, it is more convenient to work with an effective reproduction number R_t, which is defined as the number of secondary infections that arise from a typical primary case [5]. The magnitude of R_t is a useful indicator for evaluating the risk of an infectious disease and the validity of controlling the epidemic. If R_t exceeds 1, the number of incident cases will inevitably grow over time, and a large epidemic is possible. To prevent the expansion of an epidemic, R_t must be reduced to a level below 1. Using the parameter R_t, one can establish the susceptible-infectious-removed (SIR) epidemic model as below [6-8]:

$$\frac{dS(t)}{dt} = -\frac{S(t)}{N}\frac{R_t}{\tau}I(t), \tag{1}$$

$$\frac{dI(t)}{dt} = \frac{S(t)}{N}\frac{R_t}{\tau}I(t) - \frac{I(t)}{\tau}, \tag{2}$$

$$\frac{d\mathcal{R}(t)}{dt} = \frac{I(t)}{\tau}. \tag{3}$$

where $S(t)$, $I(t)$ and $\mathcal{R}(t)$ are the number of susceptible, infectious, and removed (including recovered and death) individuals at time t; τ denotes the generation interval that is the time from infection of an individual to the infection of a secondary case by that individual, namely, the "contagion period" of an infection [5]. The generation interval τ should be a random variable. If we denote the number of populations by N, we have:

$$N = S(t) + I(t) + \mathcal{R}(t). \tag{4}$$

If R_t and τ of an epidemic are known, one can employ the SIR model (1)-(4) to

simulate the evolutionary track of this epidemic. To estimate the reproduction number R_t, the probability distribution function of the generation interval of an infectious disease, $p(\tau)$, is required to be available [5, 9-12]; however, this distribution is often unknown. In the existing literature, many scholars used exponential distribution [5], normal distribution [5], Weibull distribution [9, 10], and Gamma distribution [3, 12] to approximate $p(\tau)$. Theoretically, to use these distributions to approximate $p(\tau)$, one needs to know enough information about symptom onsets of all cases, namely, large sample cases for τ. Regarding the incomplete information, one also applied the Monte-Carlo method [4] and Bayesian statistical inference [11] to estimating $p(\tau)$. However, thus far, there is scant literature to discuss the potential application of the maximum entropy method (MaxEnt) [13-15] in estimating the reproduction number. Our letter fills this gap. In the statistical inference, MaxEnt is a powerful tool of predicting probability distributions. The main idea of MaxEnt is to estimate a target probability distribution by finding the probability distribution of maximum entropy, subject to a set of constraints that represent our incomplete information for the target distribution [13]. Due to the advanced predictive capacity, MaxEnt has been widely applied in thermodynamics [13], economics [16-19], artificial intelligence [20-21], and ecology [22-26].

In this letter, we apply the MaxEnt to determining the function shape of $p(\tau)$. Before doing so, we first introduce the relationship between R_t and $p(\tau)$. Here, we adopt Wallinga and Lipsitch's method [5] for deriving the reproduction number. By both authors' method, the number of infectious individuals at time t can be written as [5]:

$$I(t) = \int_0^{\tau_{max}} I(t-\tau)n(\tau,t)d\tau, \tag{5}$$

where $n(\tau,t)$ denotes the number of cases infected by a τ-day infectious individual at time t. Here τ_{max} denotes the maximum symptom duration. Wallinga and Lipsitch assumed [5] $\tau_{max} = +\infty$. To make the model more realistic, we assume that τ_{max} is a finite number. Thus, the reproduction number R_t can be defined as [5]:

$$R_t = \int_0^{\tau_{max}} n(\tau,t)d\tau. \tag{6}$$

Let us order

$$p(\tau) = \frac{n(\tau,t)}{R_t}. \qquad (7)$$

Substituting equation (7) into equation (6) yields:

$$\int_0^{\tau_{max}} p(\tau)d\tau = 1, \qquad (8)$$

where $p(\tau)$ is the probability distribution function of the generation interval τ [5].

Using equation (8), the mean value of the generation interval can be written as:

$$\int_0^{\tau_{max}} p(\tau) \cdot \tau \; d\tau = \bar{\tau}. \qquad (9)$$

If the mean value $\bar{\tau}$ is known, then by using equation (8) one can obtain the variance of the generation interval:

$$\int_0^{\tau_{max}} p(\tau) \cdot (\tau - \bar{\tau})^2 \; d\tau = \sigma^2. \qquad (10)$$

Substituting equation (7) into equation (5) we finally obtain:

$$R_t = I(t) / \int_0^{\tau_{max}} I(t - \tau)p(\tau)d\tau . \qquad (11)$$

Equation (11) is the basic formula for calculating the reproduction number. If $I(t)$ and $p(\tau)$ are known, one can calculate the reproduction number by using equation (11). Generally speaking, the number of infectious individuals $I(t)$ is reported for each day, while the function shape of $p(\tau)$ is unknown. Therefore, many scholars used exponential distribution [5], normal distribution [5], Weibull distribution [9, 10] and Gamma distribution [3, 12] to approximate $p(\tau)$. To do so, one needs to collect enough information of τ, which requires examining a large number of cases. From a practical point of view, it is easier to collect a sample set of cases (at least 30 samples) to calculate the approximate estimates of the mean value $\bar{\tau}$ and the variance σ^2. Given the approximate estimates of $\bar{\tau}$ and σ^2 as the prior information, we maximize the information entropy of the generation interval τ to infer the function shape of the probability distribution $p(\tau)$. This is the basic idea of the MaxEnt, which agrees with everything that is known, but avoids assuming anything that is not known [15]. The resulting statistical inference gives the least biased predictions of the shapes of probability distributions consistent with prior knowledge [23].

Now we apply the MaxEnt to determining the probability distribution function

$p(\tau)$. To this end, we assume that the mean value $\bar{\tau}$ and the variance σ^2 of the generation interval τ are known. By equation (8), we define the information entropy of the generation interval τ as:

$$S = -\int_0^{\tau_{max}} p(\tau) \cdot \ln p(\tau) d\tau. \tag{12}$$

Because we only know the mean value $\bar{\tau}$ and the variance σ^2, maximizing the information entropy (12) should yield:

$$\begin{cases} \max S[p(\tau)] = -\int_0^{\tau_{max}} p(\tau) \cdot \ln p(\tau) \, d\tau \\ s.t. \quad \int_0^{\tau_{max}} p(\tau) d\tau = 1 \\ \quad \int_0^{\tau_{max}} p(\tau) \cdot \tau \, d\tau = \bar{\tau} \\ \quad \int_0^{\tau_{max}} p(\tau) \cdot (\tau - \bar{\tau})^2 \, d\tau = \sigma^2 \end{cases} . \tag{13}$$

To solve the optimal problem (13), we construct the Lagrange function:

$$\mathcal{L}[p(\tau)] = -\int_0^{\tau_{max}} p(\tau) \cdot \ln p(\tau) \, d\tau - \alpha'\left(\int_0^{\tau_{max}} p(\tau) d\tau - 1\right) - \beta\left(\int_0^{\tau_{max}} p(\tau) \cdot\right.$$

$$\left.\tau \, dx - \bar{\tau}\right) - \gamma\left(\int_0^{\tau_{max}} p(\tau) \cdot (\tau - \bar{\tau})^2 \, d\tau - \sigma^2\right), \tag{14}$$

where α', β, and γ are Lagrange multipliers.

Plugging equation (14) into the functional derivative $\delta\mathcal{L}[p(\tau)]/\delta p(\tau) = 0$ we get the optimal solution:

$$p(\tau) = p(\tau|\alpha, \beta, \gamma) = \alpha \cdot \exp[-\beta\tau - \gamma(\tau - \bar{\tau})^2], \tag{15}$$

where $\alpha = \exp(-\alpha' - 1)$.

Theoretically, substituting equation (15) into equations (8), (9), and (10) one can calculate the values of α, β, and γ. However, it is difficult to obtain the analytic results of integrals (8), (9) and (10). To do numerical calculation for equations (8), (9), and (10), we assume that $p(\tau)$ quickly tends to 0 as $\tau \gg 1$. The validity of this assumption can be justified by checking equation (15); therefore, equations (8), (9) and (10) can be written as:

$$\int_0^{\tau_{max}} p(\tau) d\tau \approx \int_0^{+\infty} p(\tau) d\tau, \tag{16}$$

$$\int_0^{\tau_{max}} p(\tau) \cdot \tau \, d\tau \approx \int_0^{+\infty} p(\tau) \cdot \tau \, d\tau, \tag{17}$$

$$\int_0^{\tau_{max}} p(\tau) \cdot (\tau - \bar{\tau})^2 \, d\tau \approx \int_0^{+\infty} p(\tau) \cdot (\tau - \bar{\tau})^2 \, d\tau. \tag{18}$$

Based on equations (16), (17) and (18), we propose a numerical method to calculate

the approximate values of α, β, and γ. To this end, let us order:

$$H(\alpha,\beta,\gamma) = \int_0^{+\infty} p(\tau|\alpha,\beta,\gamma)\, d\tau. \tag{19}$$

The partial derivatives of equation (19) with respect to β and γ yield:

$$\frac{\partial H(\alpha,\beta,\gamma)}{\partial \beta} = -\int_0^{+\infty} p(\tau|\alpha,\beta,\gamma) \cdot \tau\, d\tau, \tag{20}$$

$$\frac{\partial H(\alpha,\beta,\gamma)}{\partial \gamma} = -\int_0^{+\infty} p(\tau|\alpha,\beta,\gamma) \cdot (\tau - \bar{\tau})^2\, d\tau. \tag{21}$$

Using equations (19), (20) and (21), equations (8), (9) and (10) can be rewritten in the form:

$$\begin{cases} H(\alpha,\beta,\gamma) = 1 \\ \frac{\partial H(\alpha,\beta,\gamma)}{\partial \beta} = -\bar{\tau} \\ \frac{\partial H(\alpha,\beta,\gamma)}{\partial \gamma} = -\sigma^2 \end{cases}, \tag{22}$$

where we have used the approximations (16), (17) and (18).

By solving equation (22), one can obtain the approximate values of α, β, and γ. Substituting equation (15) into equation (19) we have

$$H(\alpha,\beta,\gamma) = \frac{\alpha}{2}\sqrt{\frac{\pi}{\gamma}}\,\mathrm{erfc}\left(\frac{\beta}{2\sqrt{\gamma}} - \bar{\tau}\right)\exp\left(\frac{\beta^2}{4\gamma} - \bar{\tau}\beta\right), \tag{23}$$

where $\mathrm{erfc}(\xi) = \frac{2}{\sqrt{\pi}}\int_{\xi}^{\infty}\exp(-x^2)\,dx$ denotes the error function.

By equation (23) it is easy to get:

$$\frac{\partial H(\alpha,\beta,\gamma)}{\partial \beta} = \frac{\alpha}{2}\sqrt{\frac{\pi}{\gamma}}\,\exp(-\gamma\bar{\tau}^2)\left[\mathrm{erfc}\left(\frac{\beta-2\bar{\tau}\gamma}{2\sqrt{\gamma}}\right)\exp\left(\frac{(\beta-2\bar{\tau}\gamma)^2}{4\gamma}\right)\frac{\beta-2\bar{\tau}\gamma}{2\gamma} - \frac{1}{\sqrt{\pi\gamma}}\right], \tag{24}$$

$$\frac{\partial H(\alpha,\beta,\gamma)}{\partial \gamma} = \frac{\alpha}{2}\sqrt{\frac{\pi}{\gamma}}\,\exp(-\gamma\bar{\tau}^2)\left[\frac{\beta+2\bar{\tau}\gamma}{2\gamma\sqrt{\pi\gamma}} - \mathrm{erfc}\left(\frac{\beta-2\bar{\tau}\gamma}{2\sqrt{\gamma}}\right)\exp\left(\frac{(\beta-2\bar{\tau}\gamma)^2}{4\gamma}\right)\frac{\beta^2+2\gamma}{4\gamma^2}\right]. \tag{25}$$

Solving equation (22) is equivalent to minimizing the following function:

$$e[\alpha,\beta,\gamma] = (H(\alpha,\beta,\gamma) - 1)^2 + \left(\frac{\partial H(\alpha,\beta,\gamma)}{\partial \beta} + \bar{\tau}\right)^2 + \left(\frac{\partial H(\alpha,\beta,\gamma)}{\partial \gamma} + \sigma^2\right)^2. \tag{26}$$

Here we employ the Matlab software to depict $e[\alpha,\beta,\gamma]$ as a $100 \times 100 \times 100$ lattice-point matrix, where the lattice spacing is 0.01. Given the accuracy of 0.01, by inputting the observed values of $\bar{\tau}$ and σ^2, we calculate α, β, and γ. We first apply equation (26) to the SARS epidemic in Singapore in 2003, where $\bar{\tau} = 8.4$ (days) and

$\sigma = 3.8$ (days) [9]. Substituting both observed values into equation (26), we seek the lattice-point minimizing equation (26) as below:

$$\begin{cases} \alpha \approx 0.27 \\ \beta \approx 0.11. \\ \gamma \approx 0.04 \end{cases} \qquad (27)$$

Substituting equation (27) into equation (15) we have:

$$p(\tau) = 0.27 \cdot \exp[-0.11\tau - 0.04(\tau - 8.4)^2]. \qquad (28)$$

The shape of equation (28) is showed in the Figure 1, see the blue curve. It is well in accordance with the sample data of generation interval of the SARS in Singapore in 2003, see the red histogram in Figure 1. This result supports the validity of the MaxEnt. The latest clinical research [1] showed that the generation interval of COVID-19 is similar to that of SARS. Therefore, we assume that the generation interval of COVID-19 shares the same probability function shape as that of SARS. This assumption was also adopted by Wu et al [4]. Based on this assumption, we apply equation (28) to estimating the reproduction number of COVID-19 in China. To this end, substituting equation (28) into equation (11) we have:

$$R_t = I(t) / \int_0^{\tau_{max}} I(t-\tau) \cdot p(\tau|0.27,0.11,0.04)d\tau. \qquad (29)$$

Because the observed data of $I(t)$ was reported for each day, we denote the unit of time t by "day". To calculate R_t by using equation (29), we need to rewrite the integral (29) as a summation formula. To do so, by Figure 1 we observe $p(\tau = 14) \approx 0$. By equations (16)-(18), this means $\tau_{max} \approx 14$. Therefore, the maximum generation interval of COVID-19 can be approximately denoted by 14. Based on this setting, equation (29) can be rewritten in the form:

$$R_{t=a\tau_{max}} \approx I(14a) / \left(\sum_{\tau=0}^{14} I(14a-\tau) \, p(\tau|0.27,0.11,0.04) \right), \qquad (30)$$

where $a = 1,2,...$ denote the ordinal number of the period and $\tau_{max} = 14$. Without loss of generality, in equation (30) we have approximately identified the reproduction number of the last day of a period as the reproduction number of this period. From the perspective of entire time span of an epidemic, this approximation satisfies the spirit of statistical mean-field method.

By using equation (30) we can report the reproduction number every 14 days.

Before doing so, we first determine the starting point of each contagion period for COVID-19 in China. According to the report of China CDC [2], January 8, 2020 was considered as the last day of a contagion period in Wuhan, China; therefore, we mark January 9, 2020 as the first day of the next contagion period. Based on this setting, the first period we report is from January 9, 2020 to January 22, 2020. In fact, this setting agrees with the date of locking down the Wuhan city, January 23, 2020. Here, we have collected the national-level data of the accumulative infected, recovered, and death cases of China's epidemic from January 10, 2020 to March 4, 2020, see Figure 2. By using the data in Figure 2, it is easy to calculate the number of real-time infected cases, $I(t)$, in China for each day. This result has been shown in Figure 3. Using the data in Figure 3, we can report the reproduction number every 14 days by using equation (30). The results have been listed in Table 1. The first period (from January 9, 2020 to January 22, 2020) can be regarded as a free propagation stage of COVID-19 because most Chinese people were aware of the outbreak of COVID-19 after January 21, 2020 and local governments did not take effective measures to control the epidemic during this period. Unfortunately, the data in the first period is very incomplete. By contrast, the reported infected cases on January 10 (41 cases) and January 22 (571 cases) can be roughly used. Consider that the first period is a free propagation stage, we use both data to approximately restore real-time data of this period by the exponential growth formula $571 = 41 \cdot \exp(12 \cdot r)$, where r denotes the growth rate. Using the restored data, the estimated value of the reproduction number for the first period is calculated to be 3.7069, see Table 1, which implies that the intensity of free transmission of COVID-19 is quite high. The World Health Organization announced [27], up to March 12, 2020, the COVID-19 had spread to 118 countries. The rapid worldwide spread of COVID-19 is an evidence for supporting our calculation. After January 22, 2020, Chinese Government started to take emergency actions to lock down the Wuhan city, and quickly performed different quarantine measures in every provinces. The powerful quarantine measures substantially reduce the contagion probability among individuals. Therefore, the subsequent periods no longer belong to free transmission. For these periods, the results of the reproduction number have been listed in Table 1. Due to the

quarantine measures, the reproduction number for the second period (from January 23, 2020 to February 5, 2020) has been reduced to 3.122 with the reduction amplitude being 15.78%. For the third period (from February 6, 2020 to February 19, 2020), the reproduction number has remarkably been reduced to 1.2114, which is close to 1. This implies that the epidemic has been effectively controlled. It should be pointed out that the last day of the third period (February 19) is just the turning point of the epidemic, refer to Figure 3; therefore, the real data supports our calculations for the reproduction number. As the epidemic come to the fourth period (from February 20, 2020 to March 4, 2020), the reproduction number is eventually reduced to 0.6028, a level below 1. In this sense, China's quarantine measures have obtained a preliminary success.

To further test the validity of the reproduction numbers in Table 1, we substitute them into the SIR model (1)-(4) for simulating the evolution of China's epidemic. To this end, let us first check the scope of application of the reproduction number formula (11), which is derived by equation (5). By the mean value theorem of integrals, equation (5) can be rewritten in the form:

$$I(t) = I(t - \tau_c) \int_0^{\tau_{max}} n(\tau, t) d\tau, \tag{31}$$

where $0 \leq \tau_c \leq \tau_{max}$.

By equation (6) and (31) we have:

$$I(\tau_c) = I(0) \int_0^{\tau_{max}} n(\tau, \tau_c) d\tau = I(0) \cdot R_{\tau_c}. \tag{32}$$

On the other hand, if we assume that R_t is a constant (step function) for each period, by equations (1)-(4) it is easy to get:

$$I(t) = I(0) + S(0) - S(t) + \frac{N}{R_t} \ln \frac{S(t)}{S(0)}. \tag{33}$$

Let us order:

$$\epsilon \approx \frac{S(0) - S(t)}{S(0)} \ll 1. \tag{34}$$

By using equation (34), equation (33) can be approximately written as:

$$I(t) \approx I(0) + S(0) - S(t) + \frac{N}{R_t} \left(\frac{S(t) - S(0)}{S(0)} \right) \approx I(0) + I(t) - \frac{N}{R_t} \frac{I(t)}{S(0)}, \tag{35}$$

which implies

$$I(\tau_c) \approx \frac{S(0)}{N} \cdot I(0) \cdot R_{\tau_c} \approx I(0) \cdot R_{\tau_c}, \tag{36}$$

9

where equation (35) is derived by using the approximations $|R_t - 1| \cdot \mathcal{R}(t) \ll I(t)$ and $S(0) \approx N$.

Comparing equations (32) and (36), we find that the reproduction number formula (11) can be applied to the SIR model (1)-(4) if the approximation (34) holds. Equation (34) implies $\left(N - S(t)\right)/N \ll 1$. The approximation obviously holds for China, where $N \approx 1.4 \times 10^9$ and $N - S(t) \approx 1 \times 10^5$. Therefore, we substitute the reproduction numbers in Table 1 into the SIR model (1)-(4) to simulate the evolution of China's epidemic. The result is shown by Figure 5, where the evolutionary track (red circles) of the epidemic is well in accordance with that (black circles) of the real incident cases in China. The simulation result requires $\tau \approx 8$, which agrees with our previous setting $\tau = 8.4 \pm 3.8$ [9]. Furthermore, we find that the reproduction numbers of quarantine periods in Table 1 can be fitted by an exponential function with $R^2 = 0.9924$, see Figure 4. Therefore, we apply this exponential function to predicting the reproduction numbers for the next seven periods (from March 5, 2020 to June 10, 2020). The results have been listed in Table 2, where we also present the predicted values of the number of real-time infected cases for the last day in each period. These predicted values imply that China's epidemic will gradually tend to disappear by May 2020, see the blue circles in Figure 5.

In conclusion, to estimate the reproduction number, the probability distribution function of the generation interval of an infectious disease is required to be available; however, this distribution is often unknown. In the existing literature, many scholars used exponential distribution, normal distribution, Weibull distribution, and Gamma distribution to approximate the generation interval distribution. To do so, one needs to collect enough information about symptom onsets of all cases, which requires examining a large number of cases. By contrast, the maximum entropy method has more advantage of predicting probability distributions given the incomplete information. In this letter, we argue that, given the mean value and variance of the generation interval, one can determine its probability distribution function by using maximum entropy method. Because the overall data (population) of the generation

interval is always absent, the maximum entropy method is a more convenient approach for estimating the probability distribution function of generation interval. By the maximum entropy method we first determine the probability distribution function of generation interval of COVID-19 and further apply it to estimating the real-time values of reproduction numbers of China's epidemic. Plugging these estimated reproduction numbers into the susceptible-infectious-removed epidemic model, we simulate the evolutionary track of the epidemic in China, which is well in accordance with that of the real incident cases.

References:

[1]. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet (2020): https://doi.org/10.1016/S0140-6736(20)30183-5

[2]. Li, Q. et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. The New England Journal of Medicine (2020):
DOI: 10.1056/NEJMoa2001316

[3]. Zhao, S. et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. International Journal of Infectious Diseases (2020): https://doi.org/10.1016/j.ijid.2020.01.050

[4]. Wu, J. T. et al. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet (2020): https://doi.org/10.1016/S0140-6736(20)30260-9

[5]. Wallinga, J. and Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proceedings of the Royal Society B **274** (2007) 599-604

[6]. Kenah, E. and Robins, J. Network-based analysis of stochastic SIR epidemic models with random and proportionate mixing. Journal of Theoretical Biology **249** (2007) 706-722

[7]. Kenah, E. and Robins, J. Second look at the spread of epidemics on networks. Physical Review E **76** (2007) 036113

[8]. Cauchemez, S. and Ferguson, N. M. Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London. Journal of the Royal Society Interface **5** (2008) 885-897

[9]. Lipsitch, M. et al. Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. Science **300** (2003) 1966

[10]. Wallinga, J. and Teunis, P. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. American Journal of Epidemiology **160** (2004) 509-516

[11]. Simon Cauchemez, S. Real-time Estimates in Early Detection of SARS. Emerging Infectious Diseases **12** (2006) 110–113.

[12]. Donnelly, C. A. et al. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. The Lancet **361** (2003) 1761-1766

[13]. Jaynes, E. T. Information Theory and Statistical Mechanics. Phys. Rev. **106** (1957) 620-630

[14]. Jaynes, E. T. On the rationale of maximum-entropy methods. Proceedings of the IEEE **70** (1982) 939-952

[15]. Jaynes, E.T. Notes on present status and future prospects. In: Grandy Jr., W.T., Schick, L.H. (Eds.), Maximum Entropy and Bayesian Methods. Kluwer, Dordrecht, The Netherlands, (1990) 1-13.

[16]. Judge, G. and Miller, D. Maximum Entropy Econometrics: Robust Estimation with Limited Data, John Wiley (1996)

[17]. Tao, Y. Competitive market for multiple firms and economic crisis. Physical Review E **82** (2010) 036118

[18]. Tao, Y. (2016), Spontaneous economic order. Journal of Evolutionary Economics **26** (2016) 467-500

[19]. Tao, Y. et al., Exponential structure of income inequality: evidence from 67 countries. Journal of Economic Interaction and Coordination **14** (2019) 345-376.

[20]. Wissner-Gross A. D. and Freer, C. E. Causal entropic forces. Physical Review Letters **110** (2013) 168702

[21]. Tao, Y. Self-referential Boltzmann machine". Physica A 123775 (2020):

https://doi.org/10.1016/j.physa.2019.123775

[22]. Phillips, S. J. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190** (2006) 231–259

[23]. Harte, J. et al. Maximum entropy and the state-variable approach to macroecology. Ecology **89** (2008) 2700–2711

[24]. Dewar, R. C. and Porte, A. Statistical mechanics unifies different ecological patterns. Journal of Theoretical Biology **251** (2008) 389-403

[25]. Frank, S. A. The common patterns of nature. Journal of Evolutionary Biology. **22** (2009) 1563-1585

[26]. Harte, J. and Newman, E. A. Maximum information entropy: a foundation for ecological theory. Trends in Ecology & Evolution **29** (2014) 384-389

[27]. https://www.who.int/redirect-pages/page/novel-coronavirus-(covid-19)-situation-dashboard
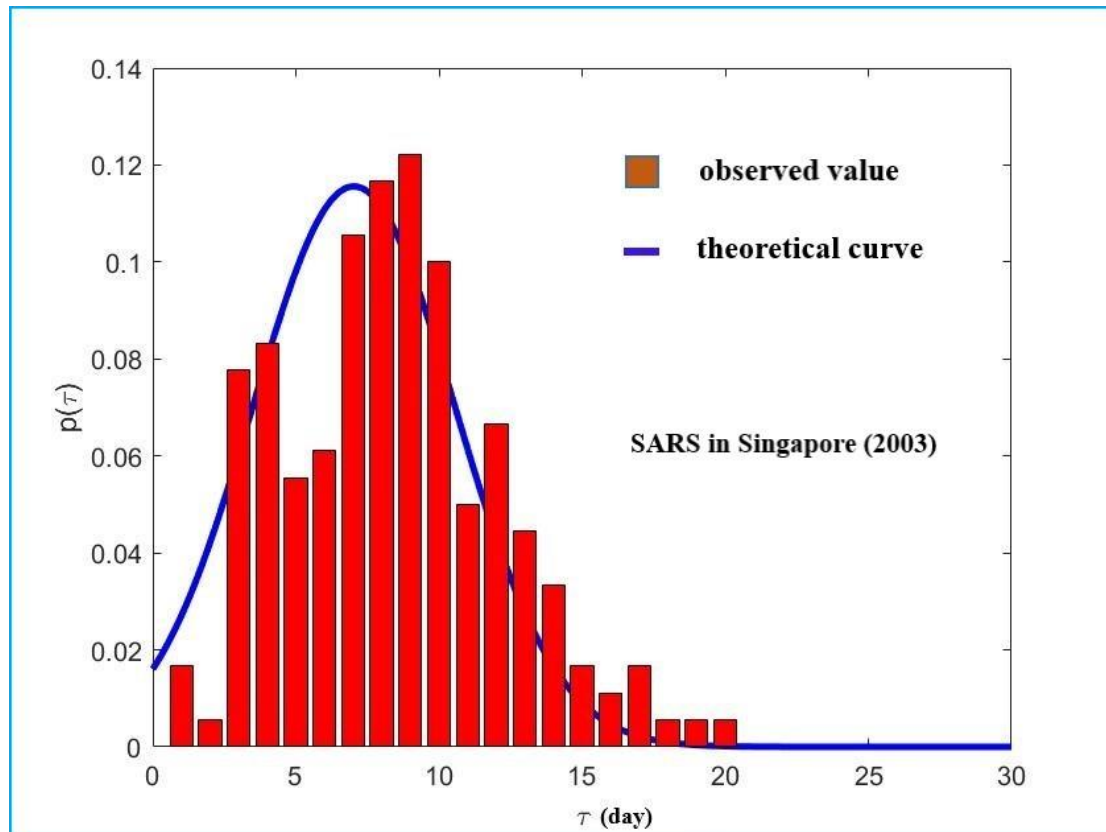
Figure 1. The function shape of equation (28) is showed by a blue curve. The sample data of generation interval of the SARS in Singapore (2003) is showed by a red histogram, and the data resource refers to [9].
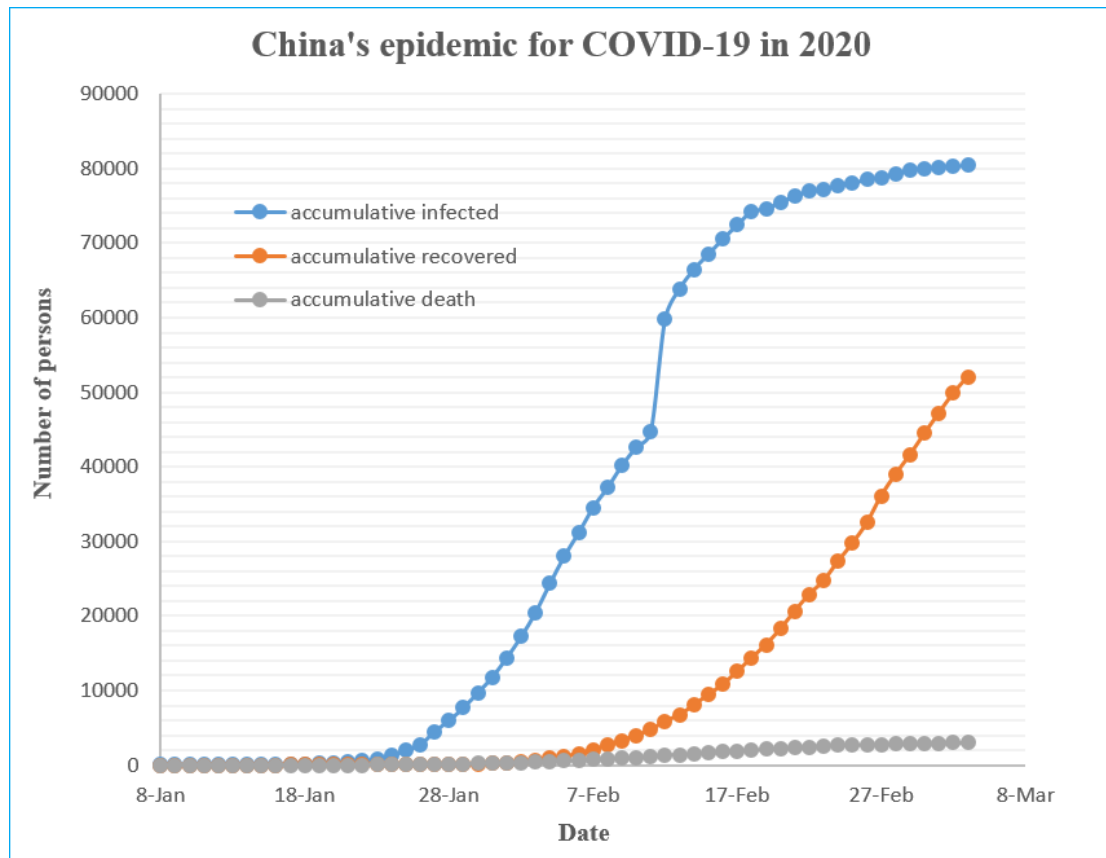
15

Figure 2. National-level data of the accumulative infected, recovered, and death cases of China's epidemic from January 10, 2020 to March 4, 2020. The data of January 8 and 9 is simply assumed to be same as that of January 10.

Data resource:

https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari_pc_3
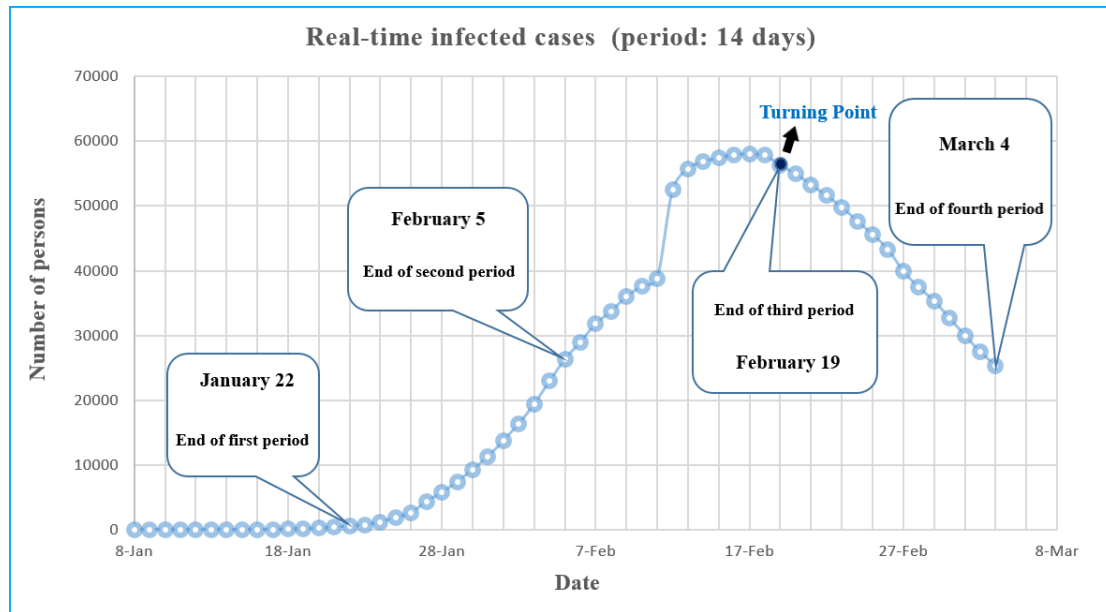
Figure 3. The number of real-time infected cases for each day = the number of accumulative infected cases for each day − the number of accumulative recovered cases for each day − the number of accumulative death cases for each day. The data comes from Figure 2.
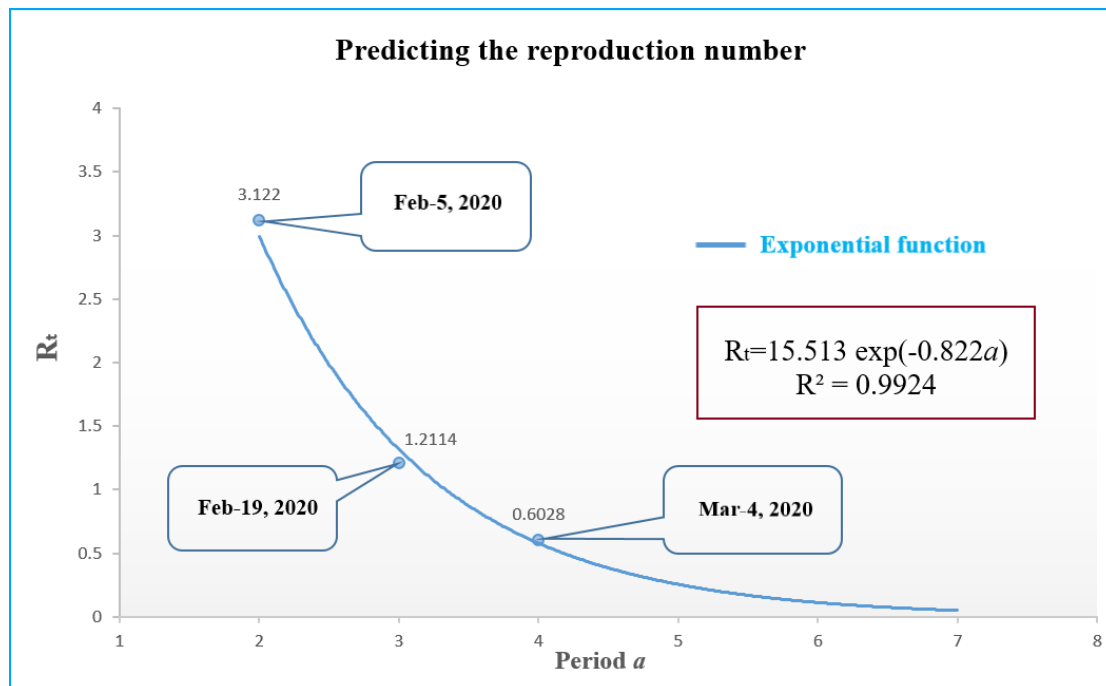
Figure 4. The reproduction numbers of quarantine periods in Table 1 are fitted by an exponential function.
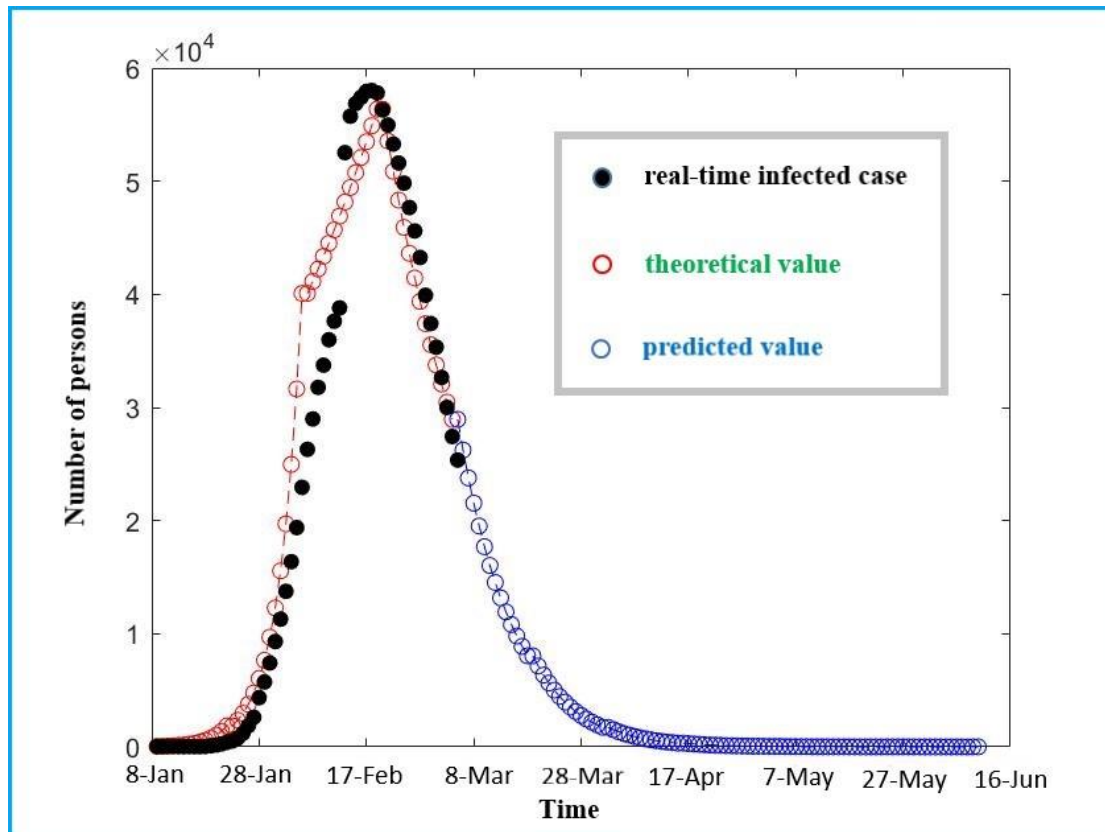
Figure 5. The SIR simulation result by using the reproduction numbers in Table 1 is showed by red circles and the SIR simulation result by using the reproduction numbers in Table 2 is showed by blue circles, where $N \approx 1.4 \times 10^9$ and $\tau \approx 8$. The real-time infected cases in Figure 3 are showed by black circles.

Table 1: The reproduction number for each period (2020)

Period number a	Period interval	R_t	Amplitude reduction
1	9-Jan to 22-Jan	3.7069	—
2	23-Jan to 5-Feb	3.1220	15.78%
3	6-Feb to 19-Feb	1.2114	61.20%
4	20-Feb to 4-Mar	0.6028	50.24%

Table 2: Predicted values (2020)

Period number a	Period interval	R_t	infected case number
5	5-Mar to 18-Mar	0.2545	8047 (last day)
6	19-Mar to 1-Apr	0.1119	1519 (last day)
7	2-Apr to 15-Apr	0.0492	253 (last day)
8	16-Apr to 29-Apr	0.0216	40 (last day)
9	30-Apr to 13-May	0.0095	6 (last day)
10	14-May to 27-May	0.0042	1 (last day)
11	28-May to 10-Jun	0.0018	0 (June 1)