Applications of Hawkes Process and SIR model in COVID-19

Ruihan Liu

Abstract

The first purpose of this paper is to use two methods to estimate the reproduction number $\mathcal{R}$ of COVID-19, which describes the average of cases directly generated by one infected individual in a population where all individuals are susceptible. The first method is to use SIR (Susceptible-Infected-Recovered) model, we estimate $\mathcal{R}$ by minimizing loss function. The second method is Hawkes process, we use maximal likelihood function to estimate $\mathcal{R}$. For further usage of Hawkes process in COVID-19, we use thinning algorithm to simulate synthetic data about conformed infection cases per day. Finally, we compare the simulation results of thinning algorithm with the actual data make some analysis. All data used in this paper from Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, which gives the increment of conformed infection cases per day in China, Italy, France, Germany, UK, US and Russia.

Keywords: Hawkes process, SIR model, Covid-19

1 Introduction

Coronavirus disease 2019 (COVID-19) was first reported in Wuhan, China in Dec 2019, it has caused serious health problems around the world. Chinese government has established restriction measures to stop the transmission of COVID-19 after Jan 20, 2020. These measures successfully controlled the infection in Mar 2020. However, many other countries did not take effective healthy restriction measures and their situations are still serious now. In this paper, we apply SIR model to reflect the effectiveness of healthy restriction policies of some countries.

The Susceptible-Infected-Recovered (SIR) model is widely used in infectious diseases. It contains three parameters, the susceptible rate β, recover rate γ and basic reproduction number $\mathcal{R} = \frac{\beta}{\gamma}$, which shows the transmission behavior of infectious disease. If $\mathcal{R} > 1$, the infection will spread in population as an exponential growth. If $\mathcal{R} < 1$, the infection will finally disappear. Here, we assume $\mathcal{R}$ as a function of time. Hence, by analyzing the variation of $\mathcal{R}$, we could see the effectiveness of restriction measures.

Hawkes process is a special point process (see §3.1), which is widely used in seismology

and crime prediction. It has property that the current events are triggered by former events, which fits disease transmission model well. On the other hand, the SIR model can also be regarded as a point process (see §3.2), so it is natural to connect SIR model and Hawkes process. From this perspective, we induce another method to estimate $\mathcal{R}$. In [1] and [4], it gives a non-parameters estimation by Hawkes process. In this paper, due to difficulty about obtaining the time sequence of events (see §3.3), we use discrete-time Hawkes process instead, see [2]. As a result, we induce log-likelihood function and apply Maximal Likelihood Estimation (MLE).

In §4, we give a simulation method about generating the synthetic data of conformed infection cases of each day by thinning algorithm, which can generate the time sequence of events governed by Hawkes process. Therefore, we fit the actual data about conformed infection cases with our simulation results. Here, we choose several countries as examples to test performance of our method. Finally, we some data analysis and induce conclusions in §5. All data used in this paper is gathered by Johns Hopkins University. These data provide daily counts of confirmed infection cases by country and region. Here, we choose China, Italy, France, Germany, UK, US and Russia as examples.

2 SIR model

2.1 Background

The Susceptible-Infected-Recovered (SIR) model contains three classes:

(i) $S(t)$: the number of *susceptible* individuals.

(ii) $I(t)$: the number of *infected* individuals.

(iii) $R(t)$: the number of *recovered* individuals.

Here, $N(t)$ is denoted by the number of population size. Since we suppose individuals who have recovered from the diseases will never be infected again, and the death individuals could also be regarded as class $R(t)$, we assume $N(t)$ is a constant and consider the relationship $N \equiv N(t) \equiv S(t) + I(t) + R(t)$.

Next, the SIR model assumes the following process.

(i) Every susceptible individual will become infected with average rate β.

(ii) Every infected individual will recover with average rate γ.

As a result, SIR model is governed by the following differential equations.

$$\begin{cases} \frac{dS(t)}{dt} = -\beta \frac{S(t)}{N} I(t) \\ \frac{dI(t)}{dt} = \beta \frac{S(t)}{N} I(t) - \gamma I(t) \\ \frac{dR(t)}{dt} = \gamma I(t) \end{cases} \tag{1}$$

2.2 Analytical Solution of SIR in Parametric Form

In general, the system of equations (1) is non-linear, but an exact analytical solution in parametric form is available. Letting $s(t) = S(t)/N$, $i(t) = I(t)/N$ and $r(t) = R(t)/N$, then we will have

$$\begin{cases} \frac{ds(t)}{dt} = -\beta s(t)i(t) \\ \frac{di(t)}{dt} = \beta s(t)i(t) - \gamma i(t) \\ \frac{dr(t)}{dt} = \gamma i(t) \end{cases} \tag{2}$$

The analytical solution of SIR in parametric form [6] is the following proposition.

Proposition 2.2.1. *Let $s(0) = s_0 > 0$, $i(0) = i_0 > 0$ and $r(0) = r_0 \geq 0$ be given, where $s_0 + i_0 + r_0 = 1$. The solution of equation is*

$$\begin{cases} s(t) = s_0 v \\ i(t) = \frac{1}{\mathcal{R}} \log(v) - s_0 v + s_0 + i_0 \\ r(t) = -\frac{1}{\mathcal{R}} \log(v) + r_0 \end{cases} \tag{3}$$

where

$$t = \int_v^1 \frac{d\xi}{\xi(\beta s_0(1-\xi) + \beta i_0 + \gamma \log \xi)} \tag{4}$$

The proof of Proposition 2.2.1 see [3]. Let's make some assumptions here. In the beginning, there is only one infected individual and other healthy individuals are all susceptible since COVID-19 was a new virus at that time. As a result, letting $S(0) = N - 1$, $I(0) = 1$ and $R(0) = 0$, i.e. $s_0 = \frac{N-1}{N}$, $i_0 = \frac{1}{N}$ and $r_0 = 0$.

Proposition 2.2.2. *With condition in Proposition 2.2.1 and $\beta s_0 > \gamma$, taking $s_0 = \frac{N-1}{N}$, $i_0 = \frac{1}{N}$ and $r_0 = 0$, the following conclusion is valid*

$$i_{max} = \frac{1}{\mathcal{R}} \log(\frac{1}{\mathcal{R} s_0}) - \frac{1}{\mathcal{R}} + 1 \tag{5}$$

Proof. Since $\beta s_0 > \gamma$, we conclude that $\frac{di(0)}{dt} = \beta s_0 i_0 - \gamma i_0 > 0$ and $i(t)$ attains its maximum when $\frac{di(t)}{dt} = 0$, i.e. $\beta s(t) = \gamma = s_0 v$. As a result,

$$i_{max} = \frac{1}{\mathcal{R}} \log(\frac{1}{\mathcal{R} s_0}) - \frac{1}{\mathcal{R}} + 1$$

□

2.3 Data analysis

In this part, we will estimate the reproduction number of several countries. According to survey, the recover time of COVID-19 is about 12 days, so we choose $\gamma = 1/12$. Then we only need to estimate the average infection rate β. On the other hand, this rate will change over time due to the public health policy. Hence, we regard β as a function of time and

define it by $\beta(t) = \sum_{k=1}^{N} b_k 1_{t\in I_k}$, where I_k are discrete time intervals and N is the number of total time intervals, see [1] and [4]. Now suppose we have a set of data $\{t_k, C(t_k)\}_{k=1}^{n}$, by using SIR model, the loss function will be constructed by

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} [\hat{C}(t_i) - C(t_i)]^2 \tag{6}$$

where $\hat{C}(t_k) = Nc(t_k) = N[i(t_k) + r(t_k)] = N - Ns_0v_k$ and $\boldsymbol{\beta} = (b_k)_{k=1}^{N}$. The main difficulty of minimizing this loss function is to calculate v_k such that

$$t_k = \int_{v_k}^{1} \frac{d\xi}{\xi(\beta s_0(1-\xi) + \beta i_0 + \gamma \log \xi)}$$

Consider the equation $\frac{ds(t)}{dt} = -\beta s(t)i(t) = -\beta s(t)[1 - s(t) - r(t)]$, the number of recover (death) individuals $R(t)$ is much less than the whole population of Wuhan, so we let $i(t) \approx 1 - s(t)$, see [6]. As a result, we obtain a new equation for $s(t)$, that is,

$$\frac{ds(t)}{dt} = -\beta s(t)[1 - s(t)] \tag{7}$$

The solution is

$$s(t) = \frac{s_0}{s_0 + (1 - s_0)e^{\beta t}} = s_0 v \tag{8}$$

So we obtain $v_k = \frac{1}{s_0+(1-s_0)e^{\beta t_k}}$. Now, we choose the time intervals I_k have length of 10 days, which does not overlap with each others, and the total number of time intervals N equals to 4. Based on data from CSSE, we estimate variation of $\mathcal{R} = \beta\gamma$ of China, Italy, France, Germany, UK, US and Russia. See Table 1.

2.4 Non-homogeneous Poisson process

In Poisson process, the arrival of events is triggered by a constant intensity λ, here, we consider $\lambda = \lambda(t)$ as a function of time t.

Definition 2.4.1 (non-homogeneous Poisson process)**.** *A point process $(N_t)_{t\geq 0}$ is called non-homogeneous Poisson process with intensity function $\lambda(t)$ if*

$$\lambda(t) = \lim_{h\to 0} \frac{P\{N_{t+h} - N_t = 1|\mathcal{H}_t\}}{h} \tag{9}$$

where $\mathcal{H}_t$ is all the events up to time t.

The following equations are equivalent to the above equation.

$$\begin{cases} P\{N_{t+h} - N_t = 1|\mathcal{H}_t\} = \lambda(t)h + o(h) \\ P\{N_{t+h} - N_t = 0|\mathcal{H}_t\} = 1 - \lambda(t)h + o(h) \\ P\{N_{t+h} - N_t > 1|\mathcal{H}_t\} = o(h) \end{cases} \tag{10}$$

where $o(h)$ means $\lim_{h\to 0} \frac{o(h)}{h} = 0$. In other words, we have

$$\lim_{h\to 0} \mathbb{E}[N_{t+h} - N_t|\mathcal{H}_t] = \lambda(t) \tag{11}$$

In fact, Hawkes process is a special kind of non-homogeneous Poisson process.

Definition 2.4.2 (Hawkes process)**.** *A Hawkes process is a non-homogeneous Poisson process with intensity function*

$$\lambda(t) = \mu(t) + \sum_{t>t_i} \phi(t - t_i) \tag{12}$$

where $\mu(t)$ is the background intensity and $\phi(\cdot)$ is the kernel function.

2.5 Hawkes process in SIR model

Another view of SIR is regarding the number of infected and recover individuals as a point process, see [5]. Suppose the number of infected individuals $I(t)$ and recover individuals $R(t)$ is i and r respectively at time t, i.e. $\{I(t) = i, R(t) = r\}$. Then choose time interval δt such that at most one event happens, there are total three possible states, that is $\{I(t+\delta t) = i+1, R(t+\delta t) = r\}$, $\{I(t+\delta t) = i-1, R(t+\delta t) = r+1\}$ and $\{I(t+\delta t) = i, R(t+\delta t) = r\}$. Here, we denote S_t, I_t and R_t as point process in order to distinguish $S(t)$, $I(t)$ and $R(t)$ in SIR model. Furthermore, letting $\{t_i^I\}_{i=1}^{N^I}$ and $\{t_i^R\}_{i=1}^{N^R}$ to be the time sequence of I_t and R_t, that is, t_i^I is the time when i-th infected individual appears as well as t_i^R.

Next, denote $C_t = I_t + R_t$, which could be regarded as a Hawkes process with intensity function $\lambda^C(t)$, according to the definition of intensity, we conclude

$$\lambda^C(t) = \lim_{\delta t \to 0} \frac{P\{C_{t+\delta t} - C_t = 1 | \mathcal{H}_t\}}{\delta t} \tag{13}$$

where $\mathcal{H}_t = \{t_1^I, t_2^I, ..., t_1^R, t_2^R, ...\}$ for all $t_i^I, t_i^R < t$. Moreover, according to equation 1, it concludes $\frac{dC(t)}{dt} = \beta S(t) I(t)$. We could assume the following equations.

$$\begin{cases} P\{C_{t+\delta t} - C_t = 1 | \mathcal{H}_t\} = \beta \frac{S_t}{N} I_t \delta t + o(\delta_t) \\ P\{C_{t+\delta t} - C_t = 0 | \mathcal{H}_t\} = 1 - \beta \frac{S_t}{N} I_t \delta t + o(\delta_t) \\ P\{C_{t+\delta t} - C_t > 1 | \mathcal{H}_t\} = o(\delta_t) \end{cases} \tag{14}$$

From above argument, $\lambda^C(t) = \beta \frac{S_t}{N} I_t$, we need to adapt equation (12) into the following form.

$$\lambda^C(t) = \beta[\mu(t) + \sum_{t>t_i} \phi(t - t_i)] \tag{15}$$

2.6 More adaptations in Hawkes process

In last subsection, we conclude that $C(t)$ is a Hawkes process. However, since we only know the number of infection per day from CSSE, the time sequence of infected individual $\{t_k^I\}_{k=1}^{N_I}$ and $\{t_i^R\}_{i=1}^{N^R}$ are not available. On the other hand, if we compute $\lambda(t)$ in Definition 2.4.2, we need the information about $\{t_k^I\}_{k=1}^{N_I}$. As a result, we use discrete-time Hawkes process as replacement. In detail, we consider the following.

Definition 2.6.1 (Discrete Hawkes process). *Suppose $(N_t)_{t\in\mathbb{N}}$ is a discrete-time Hawkes process, the intensity function $\lambda(t)$ defined as the following*

$$\lambda(t) = \mathbb{E}[N_t - N_{t-1}|\mathcal{H}_{t-1}] \tag{16}$$

Notice that Definition 2.6.1 is similar as (11). Hence, we let $\lambda^C(t) = \mathbb{E}[C_t - C_{t-1}|\mathcal{H}_{t-1}]$. In general, Hawkes process has self-exciting property, which is suitable for infectious diseases transmission model. However, not only the previous infected individuals, other factors such as location and the number of infected population also play an important role in diseases transmission. It is necessary to make some adaptations on Hawkes process, as we have done in §3.2. For example, in [5], it introduces HawkesN process as the following.

Definition 2.6.2 (HawkesN process). *A point process is called HawkesN process if its intensity function has the following form*

$$\lambda^N(t) = (1 - \frac{N_t}{N})[\mu(t) + \sum_{t>t_i} \phi(t - t_i)] \tag{17}$$

where N is the total number of population in our model.

However, in our cases, since $\frac{C_t}{N} = 1 - \frac{I_t+R_t}{N}$ and the total number of infected individuals is no more than 100000, which is much less than the whole population of Hubei province, we have an approximation $\frac{C_t}{N} \approx 1$. Therefore, we just use standard Hawkes process as an approximation of HawkesN.

From another perspective, we adapt the intensity function of $(C_t)_{t\in\mathbb{N}}$ as the following

$$\lambda^C(t) = \mathbb{E}[C_t - C_{t-1}|\mathcal{H}_{t-1}] = \beta[\mu(t) + \sum_{t>t_i} m_{t_i}\phi(t - t_i)] \tag{18}$$

where m_s is the number of events in day $s \in \mathbb{N}$ and $t_i \in \mathbb{N}$, i.e. the number of conformed infected individuals before s. In this way, we can see the number of infected population also contributes the future infection.

2.7 Estimation of $\mathcal{R}$

In paper [1], it uses a non-parameters estimation of $\mathcal{R}$. We will use Maximal Likelihood Estimation (MLE), to estimate $\mathcal{R}$. First, we need to choose proper kernel function $\phi(\cdot)$. For simpleness, we let the external rate μ be a constant and the kernel function $\phi(x) = \alpha e^{-\omega x}$ be exponential, hence we have

$$\lambda^C(t) = \beta[\mu + \alpha \sum_{t>t_i} m_{t_i} e^{-\omega(t-t_i-1)}] \tag{19}$$

In addition, for $t \in \mathbb{N}$, the probability of k events happened in time interval $[t, t+1]$ is

$$P(C_{t+1} - C_t = k|\mathcal{H}_t) = \frac{e^{-\lambda(t)}\lambda(t)^k}{k!} \tag{20}$$

Hence, the log-likelihood function will be

$$\mathcal{L}(\beta, \mu, \alpha, \omega) = \sum_{n=1}^{N-1} \Delta C_n \log[\lambda(n)] - \lambda(n) \tag{21}$$

where $\Delta C_n = C_{n+1} - C_n$ and N is the total number of days in our data.

In order to calculate the maximal value of $\mathcal{L}(\beta, \mu, \alpha, \omega)$, we use Monte Carlo method here. Since for some algorithms like Expectation-Maximal and gradient descent method, the initial value will greatly influence the estimation results. In other words, we may obtain the local maximal value of $\mathcal{L}(\beta, \mu, \alpha, \omega)$ instead of overall maximal value. For this reason, using Monte Carlo method is a reasonable option. Here, we let

$$\mu, \alpha, \omega \sim U(0, \infty) \; and \; \beta \sim U(0, 1) \tag{22}$$

where $U(a, b)$ is the uniform distribution on interval (a, b). Next, we generate parameter pairs $(\beta_k, \mu_k, \alpha_k, \omega_k)_{k=1}^N$ for N times based on (22), then calculate $\mathcal{L}_k := \mathcal{L}(\beta_k, \mu_k, \alpha_k, \omega_k)$ for each time. Suppose $(\beta^*, \mu^*, \alpha^*, \omega^*)$ satisfies

$$\mathcal{L}(\beta^*, \mu^*, \alpha^*, \omega^*) = \max_{1 \le k \le N} \mathcal{L}_k$$

We choose β^* as our estimation result. Now, we just need to repeat the above process for M times and obtain the estimation results $(\beta_l^*)_{l=1}^M$. Finally, we can obtain the 90% confidence of $(\beta_l^*)_{l=1}^M$. Similar as §2.4, we estimate $\mathcal{R}$ of China, Italy, France, Germany, UK, US and Russia based on CSSE, then calculate the 90% confidence of these countries. All results are shown in Table 1.

3 Further application of Hawkes process in COVID-19

In this part, we will use Hawkes process to simulate synthetic data about the conformed infection cases per day. In detail, we regard $C(t)$ defined in §3.2 as a standard Hawkes process here, whose intensity function defined by (12). After generating synthetic data, we compare our results with the actual data from CSSE. In addition, since we will not use discrete-time Hawkes process as in §3.3, one difficulty related to the time sequence of events should be considered. In the next subsection, we introduce some adaptation to overcome this problem.

3.1 Preliminary preparation for simulation

In this subsection, we use Hawkes process defined by (12) to generate synthetic data of the total conformed infected cases and then compare our simulation results with actual data. For purpose of simpleness, we choose the background intensity $\mu(t)$ to be constant and kernel function $\phi(t)$ to be Weibull distribution, that is

$$\lambda(t) = \mu + \sum_{t > t_i} \frac{k}{\lambda} (\frac{t - t_i}{\lambda})^{k-1} e^{-(\frac{t-t_i}{\lambda})^k} \tag{23}$$

Country	**model**	**start date**	$\mathcal{R}_1$	$\mathcal{R}_2$	$\mathcal{R}_3$	$\mathcal{R}_4$
China	Hawkes	Jan 22	(1.26,1.88)	(1.25,1.61)	(1.24,1.59)	(1.22,1.32)
China	SIR	Jan 22	1.81	1.66	1.34	0.93
Italy	Hawkes	Feb 21	(1.68,2.26)	(1.59,2.14)	(1.70,2.12)	(1.41,1.92)
Italy	SIR	Feb 21	2.34	2.38	2.17	1.80
France	Hawkes	Mar 1	(1.48,1.90)	(1.63,2.12)	(1.51,2.00)	(1.32,1.93)
France	SIR	Mar 1	1.62	1.63	1.60	1.34
Germany	Hawkes	Mar 1	(1.25,1.96)	(1.42,2.04)	(1.50,2.06)	(1.27,1.80)
Germany	SIR	Mar 1	1.62	1.86	1.75	1.43
US	Hawkes	Mar 1	(1.42,1.91)	(1.57,2.38)	(1.57,2.53)	(1.63,2.55)
US	SIR	Mar 1	1.65	1.90	2.05	2.00
UK	Hawkes	Mar 1	(1.21,1.72)	(1.25,1.96)	(1.27,1.85)	(1.48,1.74)
UK	SIR	Mar 1	1.86	1.82	1.77	1.45
Russia	Hawkes	Mar 11	(1.24,1.93)	(1.56,2.30)	(1.42,2.06)	(1.51,1.90)
Russia	SIR	Mar 11	1.90	2.04	2.01	1.95

Table 1: Results from SIR and Hawkes process. Here, $\mathcal{R}_1$ is the value of the basic reproduction number in the first ten days after the start date, as well as other $\mathcal{R}_k$. The brackets in Hawkes line means the 90% confidence intervals of estimation results from §3.4.

However, in order to estimate the value of parameters (μ, λ, k), we need the time sequence of infection $\{t_i\}_{i=1}^n$, where t_i is the time of i-th infected individual appeared. This difficulty has been discussed in §3.3. And we use discrete-time Hawkes process as a replacement to solve this difficulty.

Here, we use another method to estimate (μ, λ, k). We define $m(t) = \int_0^t \lambda(s)ds$ for $t \in \mathbb{N}$ and the probability of n infection cases happened in time interval $[t, t+1]$ is

$$P(C_{t+1} - C_t = n|\mathcal{H}_t) = \frac{e^{-[m(t+1)-m(t)]}[m(t+1) - m(t)]^n}{n!} \tag{24}$$

Hence, we conclude $\mathbb{E}[C_{t+1} - C_t|\mathcal{H}_t] = m(t+1) - m(t)$. In addition, suppose $\{t_i^*\}_{i=1}^n$ is the time sequence of infection cases happened in $[t, t+1]$ and $\{t_j\}_{j=1}^{N_t}$ is the time sequence of infection cases happened in $[0, t]$, we have

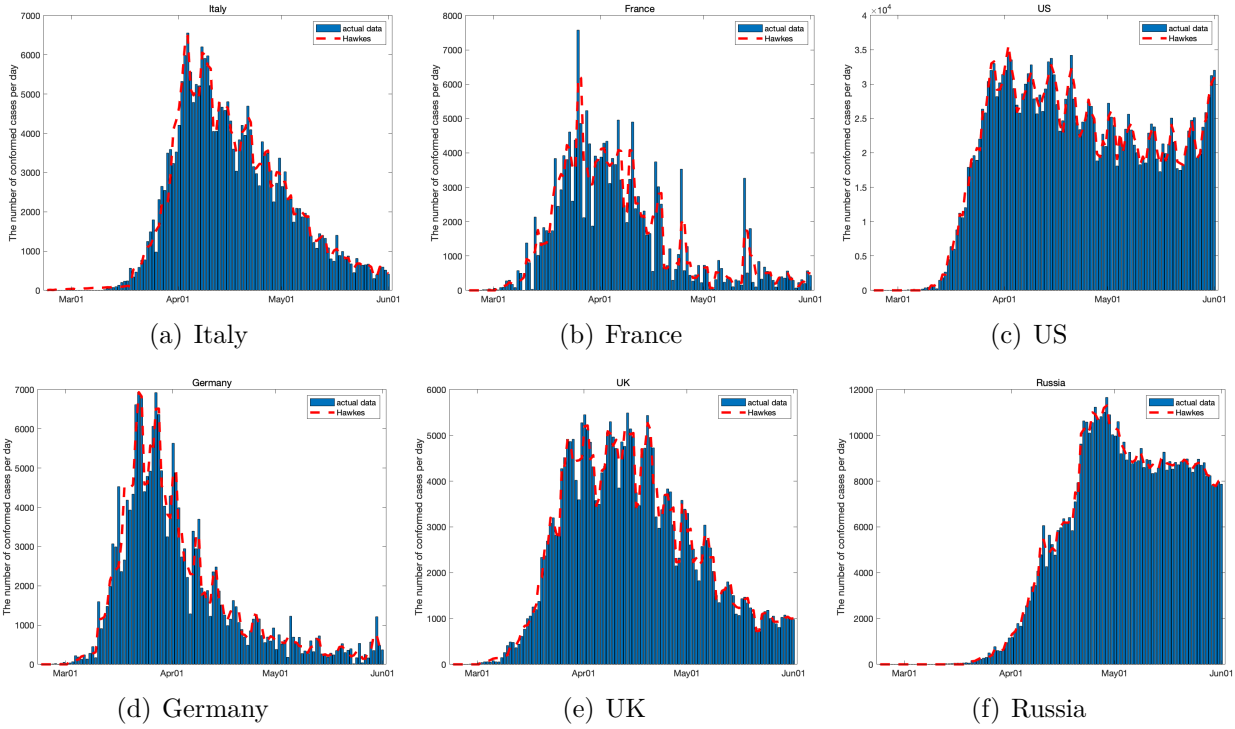
$$m(t+1) - m(t) = \mu + n - \sum_{i=1}^{n} e^{-(\frac{t+1-t_i^*}{\lambda})^k} + \sum_{j=1}^{N_t}[e^{-(\frac{t-t_j}{\lambda})^k} - e^{-(\frac{t+1-t_j}{\lambda})^k}] \tag{25}$$

Here we assume $\lambda >> t$ and get an approximation $m(t+1) - m(t) \approx \mu$. On the other hand, if we simulate a Hawkes process defined in (23) by thinning algorithm (see §4.2), we will obtain the number of events happened in $[t, t+1]$. In detail, we assume $\widetilde{C}_{t+1}^l - \widetilde{C}_t^l = n_l$ for $l \in \{1, 2, ..., M\}$, that is, we repeat simulation process for M times and $\widetilde{C}^l$ is the l-th

simulation result. Then by *Large Number law*, we conclude

$$\lim_{M\to\infty}\frac{\sum_{l=1}^{M} n_l}{M} = \lim_{M\to\infty}\frac{\widetilde{C}_{t+1}^{l} - \widetilde{C}_{t}^{l}}{M} = \mathbb{E}[C_{t+1} - C_t|\mathcal{H}_t] \approx \mu \tag{26}$$

From (26), we have the expected number of events in $[t, t+1]$ only depends on μ when $\lambda >> t$. In other words, we can fix the value of λ and k, then only estimate the value of μ.



(a) Italy (b) France (c) US

(d) Germany (e) UK (f) Russia

Figure 1: The blue bar shows the number of conformed infection cases per day of above countries, and the red line shows our fitting results by Hawkes process.

3.2 Simulation process

In this subsection, we first introduce the thinning algorithm, which is used to simulate Hawkes Process. It has the following steps, see [5].

1. Input: μ, λ, k, $\mathcal{R}(\cdot)$, initial time $t = 0$, event conter $i = 1$ and stopping time T.
2. While $t < T$
   (a) Compute $\lambda^* = \lambda(t)$ by using (19).
   (b) Generate $r \sim U(0,1)$ by uniform distribution and $\tau = -\frac{\ln(u)}{\lambda^*}$.
   (c) $t = t + \tau$ and $s \sim U(0,1)$.
   (d) If $s \leq \frac{\lambda(t)}{\lambda^*}$, denote $t_i = t$; otherwise return (a).

Now, we choose $k = 4$ and $\lambda = 50$. From §4.2, if the increment of infection cases on day t is n, the optimal estimation of μ should equal to n. According to (26), we hope $|n_l - n|$ as small as possible for $l \in \{1, 2, ..., M\}$, hence we conclude

$$\lim_{M\to\infty} \frac{\sum_{l=1}^{M} n_l}{M} \approx n$$

In conclusion, we have $\mu \approx n$. Finally, we simulate the increment of conformed infection cases per day by thinning algorithm for 10000 times, then find the best fitting results. Here, we choose Italy, US, Germany and France as examples based on data CSSE. All results are shown in Figure 1.

4 Conclusion

In Table 1, we estimate the reproduction number $\mathcal{R}$ of several countries after the outbreak of COVID-19. Due to many external conditions, such as the public health policy, $\mathcal{R}$ will change over time. If $\mathcal{R}$ is decreasing and finally less than 1, that means COVID-19 is nearly controlled. On the other hand, from Table 1, the reproduction number of all countries except China is increasing or keeping at a comparatively high level (more than 1). This result shows that all these countries did not take effective measures to stop the transmission of COVID-19. In fact, according to our model, the reproduction number of China is less than 1 in the beginning of March 2020. This result shows the public health policies of Chinese government are successful. However, for other countries, the situation is serious until now. In conclusion, we recommend all countries should take high-level restriction in order to stop COVID-19.

As for fitting results in Figure 1, we clearly see our synthetic data is quite close to actual data. Although the simulation process in §4.2 will lead some randomness, the fluctuation caused by our algorithm is comparatively small. In summary, we make sure our method has well performance in fitting the number of infection cases. From Figure 1, we also notice some European countries, such as Italy, UK, France and Germany, have relatively controlled the transmission of COVID-19 in Jun, 2020. But we can not conform these countries have totally succeed in stopping the virus, since their health restriction was relaxed due to temporary good results. In addition, for country like US and Russia, their situation is still serious but their public health policies are ineffective. Hence, external transmission from these countries will destroy temporary safe condition of Italy, France, UK, Germany or other European countries. In conclusion, high-level public health restriction should be hold on until all countries in the world successfully control COVID-19.

References

[1] Andrea L Bertozzi, Elisa Franco, George Mohler, Martin B Short, and Daniel Sledge. The challenges of modeling and forecasting the spread of covid-19. *Proceedings of the National Academy of Sciences*, 117(29):16732–16738, 2020.

[2] Raiha Browning, Deborah Sulem, Kerrie Mengersen, Vincent Rivoirard, and Judith Rousseau. Simple discrete-time self-exciting models can describe complex dynamic processes: A case study of covid-19. *PloS one*, 16(4):e0250015, 2021.

[3] Tiberiu Harko, Francisco SN Lobo, and MK Mak. Exact analytical solutions of the susceptible-infected-recovered (sir) epidemic model and of the sir model with equal death and birth rates. *Applied Mathematics and Computation*, 236:184–194, 2014.

[4] George Mohler, Frederic Schoenberg, Martin B Short, and Daniel Sledge. Analyzing the world-wide impact of public health interventions on the transmission dynamics of covid-19. *arXiv preprint arXiv:2004.01714*, 2020.

[5] Marian-Andrei Rizoiu, Young Lee, Swapnil Mishra, and Lexing Xie. A tutorial on hawkes processes for events in social media. *arXiv preprint arXiv:1708.06401*, 2017.

[6] Alexis Akira Toda. Susceptible-infected-recovered (sir) dynamics of covid-19 and economic impact. *arXiv preprint arXiv:2003.11221*, 2020.