Comparison of Hakwes Model and Compartmental Models for COVID-19

Sarah Masri

97415681

September 4, 2024

Abstract

1 Background

The occurrence of pandemics and the spread of infectious disease has induced the application of various epidemiological models to understand and predict the illnesses patterns and behaviour. Among these, the Hawkes model and compartmental models, such as the SIR (Susceptible-Infectious-Recovered) model, stand out as prominent approaches, each with unique characteristics, assumptions and conclusions. Within the application of modelling the spread of SARS-CoV-2, most approaches involve some extension of the classic SIR model [?].

1.1 The Hawkes Process

The Hawkes process is a self-exciting point process often used to model seismic behaviour, dynamics in crime, and infectious disease. The self-exciting behaviour of this process implies that a single event may influence the rate of future events, for some amount of time. We explore how the Hawkes process may be applied to understanding and modelling the spread of infectious diseases like COVID-19.

Definition 1 (Point Process). *A point process is a stochastic process with event times $t_1, t_2, \ldots$ where each t_j describes the arrival time of the j-th event [?].*

Definition 2 (Poisson Process). *A Poisson process is a stochastic model for a series of discrete events where the average time in between events is known, but the exact placement of each event is random [?].*

Definition 3 (Homogenous Poisson Process). *In a homogeneous Poisson process is a counting process with inter-arrival times $\tau_j = t_j - t_{j-1}$ that are iid exponentially distributed random variables with parameter λ [?].*

Definition 4 (Self-Exciting Process). *A self-exciting process is a generalized Poisson process whose intensity is increased after each arrival event for some period of time [?].*

Definition 5 (Hawkes Process). *The Hawkes process is a self-exciting process such that for each arrival time $t_j < t$, new events are generated at rate $\phi(t - t_j)$. The intensity (event rate) of the Hawkes process is characterized by a stochastic function depending on the horizon (previous arrival times)*

$$\lambda(t) = \mu + \sum_{t_j < t} \phi(t - t_j)$$

where μ is the background rate and $\phi(\cdot)$ is referred to as the (triggering) kernel [? ?].

New events either enter the system through the background rate, or are generated through previous events corresponding with the triggering kernel $\phi(\cdot)$.

Consider the application of infectious disease. The Hawkes point process is a popular choice as an epidemic model.

Definition 6 (HawkesN Process). *The HawkesN model is a generalization of the Hawkes process with the assumption of a finite population. This process has intensity*

$$\lambda^H(t) = \left(1 - \frac{N_t}{N}\right)\left[\mu + \sum_{t_j < t} \phi(t - t_j)\right]$$

where $\phi(t - t_j)$ is the triggering kernel, N_t is the counting process $1 - \frac{N_t}{N}$ scales the event rate at time t with the proportion of events that can occur after time t [?].

The Hawkes process is a special case of the HawkesN as $N \to \infty$.

1.2   Compartmental Model

Compartmental models are a common approach in epidemiology for understanding and modelling infectious diseases. These models partition the population into compartments based on their disease states which typically include: susceptible (S), exposed (E), infectious (I), and recovered (R). These models use differential equations to describe the transitions between the compartments which allows the simulation of dynamics of the diseases transmission over time [?].

Definition 7 (SIR Model). *The SIR model describes a class compartmental model with SIR population groups [?].*

$$\frac{dS}{dt} = -\beta\frac{IS}{N}, \quad \frac{dI}{dt} = \beta\frac{IS}{N} - \gamma I, \quad \frac{dR}{dt} = \gamma I$$

where β is the transmission rate constant, γ is the recovery rate constant, and $R_0 = \beta/\gamma$.

Definition 8 (SEIR Model). *The SEIR model describes a class compartmental model accounting for an 'exposed' compartment [?].*

$$\frac{dS}{dt} = -\beta\frac{IS}{N}, \quad \frac{dE}{dt} = \beta\frac{IS}{N} - aE, \quad \frac{dI}{dt} = aE - \gamma I, \quad \frac{dR}{dt} = \gamma I$$

where β is the transmission rate constant, γ is the recovery rate constant, and $R_0 = \beta/\gamma$.

Definition 9 (Stochastic SIR). *One representation of the stochastic SIR model is the bivariate point process [?]. Only two possible events may occur: infection and recovery. In this representation, the j-th infected individual in the SIR process gets infected at time $_i^I$ and recovers at time t_j^R.*

Let S_t, I_t, and R_t be discrete random variables. These are all stochastic counter parts to $S(t)$, $I(t)$, and $R(t)$ respectively such that

$$S(t) = \mathbb{E}_{\mathcal{H}_t}[S_t], \quad I(t) = \mathbb{E}_{\mathcal{H}_t}[I_t], \quad S(t) = \mathbb{E}_{\mathcal{H}_t}[R_t]$$

Define the time to recovery as $\tau_j = t_j^R - t_j^I$. These times are distributed exponentially with parameter γ and infections lasting on average $1/\gamma$ units of time.

Let C_t be the counting process associated with infections, and R_t be the counting process associated with recovers. Then we have that $C_t = N - S_t$.

Let $\mathcal{H}_t$ be the history of the bivariate epidemic process up to time t such that $\mathcal{H}_t = \{t_1^I, t_2^I, \ldots, t_1^R, t_2^R, \ldots\}$. Then, it can be shown that

$$\lambda^I(t) = \beta \frac{S_t}{N} I_t, \qquad \lambda^R(t) = \gamma I_t$$

1.3 Literature Review

Both the Hawkes and SIR models for SARS-CoV-2 have been explored in literature. Our particular interest is the comparison and connection between the two models. It is natural to compare and contrast the performance of each of these models, but analysis on their relationship and possible convergence is limited.

[?] discusses the relationship between the two models by mean of the reproduction number. The reproduction number notably used in the SIR model is naturally connected to the productivity constant K in Hawkes, as they can both be interpreted as the expected number of infected individuals. This paper also claims that assuming an exponential triggering kernel, the intensity of the Hawkes point process is a continuous time analog to the discrete stochastic SIR model. This claim is proved in [?]. Further extension on this claim is limited in literature.

Literature regarding the use of the Hawkes process and SIR model varies in approaches and application.

Scaling Hawkes processes to one million COVID-19 Cases Holbrook - Likelihood based inference requires $O(N^2)$ floating point operations for N cases - Applies two spatiotemporal Hawkes models to the analysis of one million COVID-19 cases in the US

Opinion Market Model: stemming far-right opinion spread using positive interventions Rizoiu - Demonstrate the convergence of proposed estimation scheme on a synthetic dataset

Stratified epidemic model using a latent marked Hawkes process Lamprinakou and Sandy - Extend unstructured homogeneously mixing epidemic model to a finite population stratified by age bands - Use latent marked Hawkes process - Apply KDPF

Briding the COVID-19 data and the epidemiological model using the time-varying parameter SIRD model Cakmakli and Simsek - Extends canonical model of epidemiology SIRD model to allow for time-varying parameters for real time measurement and prediction of COVID19

A stochastic model for the early stages of highly contagious epidemics by using a state dependent point process Casillas

Estimating COVID transmission time using Hawks point process Schoenberg

Lecture notes for computational modelling of contagion and epidemics: a companion to the rule of contagion Mohler

Some statistical problems involved in foforecasting and estimating the spread of COVID19 using the Hawkes point processes and SEIR models

2 Comparison and Connection

Understanding the connections and differences between the Hawkes and compartmental models for COVID-19 can help deepend our understanding of the spread of the virus. This connection may enhance our ability to predict observed and latent cases, estimate pandemic-related parameters, and help understand the underlying mechanisms that drive the spread of the disease.

Compartmental models such as SIR and SEIR are typically favoured to model COVID-19 for their physically plausible framework for the disease relative to the Hawkes model [?]. In particular, they model the random movement of individuals between compartments and allow for the estimation of various parameters relevant to policy-makers.

Conversely, the Hawkes model may be favoured due to its allowance of non-parametric estimation of the triggering function, spatial covariates, and ability to model how background events may trigger future events through its self-exciting properties. Additionally, this model is computationally less expensive with respect to parametric and non-parametric estimation of pandemic parameters.

2.1 Assumptions

When modelling the behaviour of infectious disease, different models incorporate various assumptions to capture the dynamics of disease transmission [? ?].

2.1.1 Hawkes Model

1. **Self-excitement**: New infection events increase the likelihood of subsequent infections.

2. **Infinite population size**: There are an infinite number of individuals in the system.

3. **Homogeneity**: Homogeneity of individuals within the population.

4. **Immunity to reinfection**: Individuals are immune to reinfection for some amount of time.

5. **Epidemic trigger**: The epidemic is triggered by a set of initial infections.

6. **Transition kernel is a PDF**: The transition kernel in the intensity of the process is a probability density function.

7. **Independence** The process is independent of it's time-constant parameters.

2.1.2 HawkesN Model

1. **Self-excitement**: New infection events increase the likelihood of subsequent infections.

2. **Fixed population size**: There is a finite number of individuals in the system.

3. **Homogeneity**: Homogeneity of individuals within the population.

4. **Immunity to reinfection**: Individuals are immune to reinfection for some amount of time.

5. **Epidemic trigger**: The epidemic is triggered by a set of initial infections. As stated by [?], it is assumed for the Maximum Likelihood estimation that the background rate is zero $\mu(t) = 0 \; \forall \; t > 0$.

6. **Transition kernel is a PDF**: The transition kernel in the intensity of the process is a probability density function.

7. **Independence** The process is independent of it's time-constant parameters.

2.1.3 SIR Model

1. **Fixed population size**: There is a finite number of individuals in the system.

2. **Homogeneity**: Homogeneity of individuals within the population.

3. **Homogeneous mixing**: The contact of individuals is randomly distributed among the infected and susceptible populations.

4. **Short time scale:** The time scale of the SIR model is short enough that births and deaths are negligible within the system. In particular, the number of deaths from the disease is small compared to the rest of the population.

2.2 Consequences

The HawkesN is completely defined by it's parameters $\{\kappa, \theta, N\}$. Let $\mathcal{L}(\kappa, \beta, c, \theta)$ be the log-likelihood of observing a set of events $\{(m_j, t_j); j = 1, \ldots, n\}$ in a non-homogeneous Poisson process with rate $\lambda^H(t)$ [?]. Note that we assume the background rate to be zero, $\mu(t) = 0 \; \forall \; t > 0$. Equivalently, we that apart from the first, each event is a result of the initial event. Then,

$$\mathcal{L}(\kappa, \beta, c, \theta) = \sum_{j=1}^{n} \log(\lambda^H(t_j^-)) - \int_0^{t_n} \lambda^H(\tau)d\tau$$

It is claimed that most compartmental models can be represented using a renewal equation [?]. This naturally bridges a connection be the two models as both the reproduction number in the renewal equation and productivity constant K in Hawkes can both be interpreted as the expected number of direct transmission per infected individual.

[?] further discusses connections between the two frameworks. They claim that the rate of events in an extended Hawkes model is identical to the rate of new infection in the SIR model.

Theorem 1. *Suppose the new infections in a stochastic SIR process of parameters $\{\beta, \gamma, N\}$ follow a point process of intensity $\lambda^I(t)$. Suppose also that the events in a HawkesN process with*

parameters $\{\mu, \kappa, \theta, N\}$ have intensity $\lambda^H(t)$.

Let $\tau = \{\tau_1, \tau_2, \ldots\}$ be the set of the times to recovery of the infected individuals in the SIR process. the expectation of $\lambda^I(t)$ over τ is equal $\lambda^H(t)$:

$$\mathbb{E}_\tau[\lambda^I(t)] = \lambda^H(t)$$

when $\mu = 0$, $\beta = \kappa\theta$, and $\gamma = \theta$ [?].

Proof. Recall that C_t denotes the infection process, and R_t denotes the recovery process. First, we write these using the sum of indicators functions. We find that

$$S_t = N - C_t = N - \sum_{j \geq 1} \mathbb{1}(t_j^I < t)$$

$$I_t = C_t - R_t = \sum_{j \geq 1} \mathbb{1}(t_j^I < t, t_j^R > t) = \sum_{t_j^I < t} \mathbb{1}(t_j^I + \tau_j > t).$$

Consider the point process consisting of only the infection events $\{t_j^I\}$. The event rate in this process is obtained by marginalizing out times of recovery.

$$\begin{aligned}
\mathbb{E}_\tau[\lambda^I(t)] &= \mathbb{E}_\tau\left[\beta \frac{S_t}{N} \sum_{t_j^I < t} \mathbb{1}(t_j^I + \tau_j > t)\right] \\
&= \sum_{t_j^I < t} \mathbb{E}_\tau\left[\beta \frac{S_t}{N} \mathbb{1}(t_j^I + \tau_j > t)\right] \\
&= \sum_{t_j^I < t} \int_0^\infty \beta \frac{S_t}{N} \mathbb{1}(t_j^I + \zeta > t) r(\zeta) d\zeta \\
&= \sum_{t_j^I < t} \beta \frac{S_t}{N} \int_{t-t_j^I}^\infty r(\zeta) d\zeta \\
&= \sum_{t_j^I < t} \beta \frac{S_t}{N} \int_{t-t_j^I}^\infty r(\zeta) d\zeta
\end{aligned}$$

where $r(\zeta)$ is the exponential probability distribution function for the time of recovery. Since $S_t = N - C_t$, we get

$$\mathbb{E}_\tau[\lambda^I(t)] = \left(1 - \frac{C_t}{N}\right) \sum_{t_j^I < t} \beta e^{-\gamma(t - t_j^I)}$$

$\square$

7

2.2.1 Exploration of Triggering Kernel

[?] throws equivalence in expectation between the intensities of the stochastic SIR and HawkesN models. We want to explore if any other choices in the triggering kernel $\phi(\cdot)$ of the HawkesN process yields more connection between the Hawkes model and other stochastic representations of other compartmental models.

First consider the SEIR model, in definition 8. Consider the stochastic representation of this model with random variables S_t, E_t, I_t, and R_t associated with their respective compartments. Suppose the new infections in this stochastic SEIR process of intensity $\lambda^I(t)$, as in definition 9.

Recall the C_t is the counting process of the infection process and R_t is the counting process of the recovery process. Then C_t is the aggregate number of occurred infections, including individuals that may still be infectious.

$$C_t = N - S_t$$

By definition, we have that $N = S_t + I_t + E_t + R_t$. It follows using the above that

$$I_t = C_t - E_t - R_t$$

Note that when considered C_t and $E_t + I_t$, the problem is reduced to Theorem 1.

$$\lambda$$

We aim to write S_t and I_t

3 Implementation

A Monte Carlo simulation with 20 repetitions are run, each with population size $N = 1300$, and theoretical parameters $I = 300$, $\gamma = 0.2$, and $\beta = 1$. These are adapted from code provided by [?]. It is important to note that the code is modified with the assumption that the total population size N is known. Hence, any estimation of N is omitted from the adapted code. The code produces synthetic data suitable for the models of interest. The SIR data is represented with a set of unique timepoints. Each timepoint t is associated with a number representing each compartment, S_t, I_t, R_t, and C_t, where $C_t = I_t + R_t$ is infection process.

The HawkesN data is transformed using the SIR data to be represented by the timepoints associated with each infection event. Moreover, by Theorem 1, the estimated parameters from the HawkesN model can be transformed to analogous SIR results using the following:

$$\gamma = \theta, \qquad \beta = \kappa\theta$$

We aim to estimate β and γ in the SIR model, as well as κ and θ in the HawkesN. We plot the likelihood for each parameter.
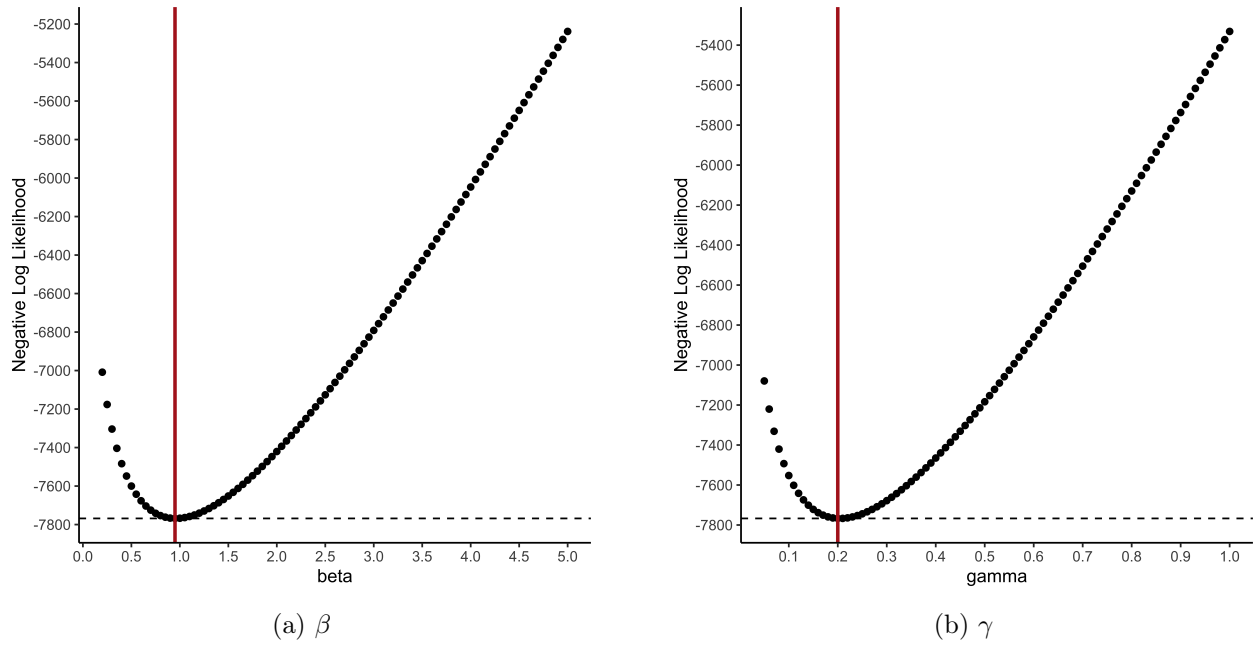
Figure 1: Negative log-likelihood plot of parameters for HawkesN model.

4 Results

Figures 1 and 2 reflect the negative log-likelihood of the parameters associated with the SIR model and HawkesN model respectively. Figure 3 shows a contour plot for each model between each pairs of parameters.

Table 1: Table Title

Variable Name	Regression 1	Mean	Std. Dev	Min	Max
text	text	text	text	text	text

Should be a caption

	I	γ	β	K	θ	c
theoretical	300	0.2	1	5	0.2	0.001
median	307	-4.3×10^{-10}	1.9	4.89	0.204	0.1
sd	11	0.2	1.1	0.52	0.026	0.000

5 Conclusions

6 Reflection

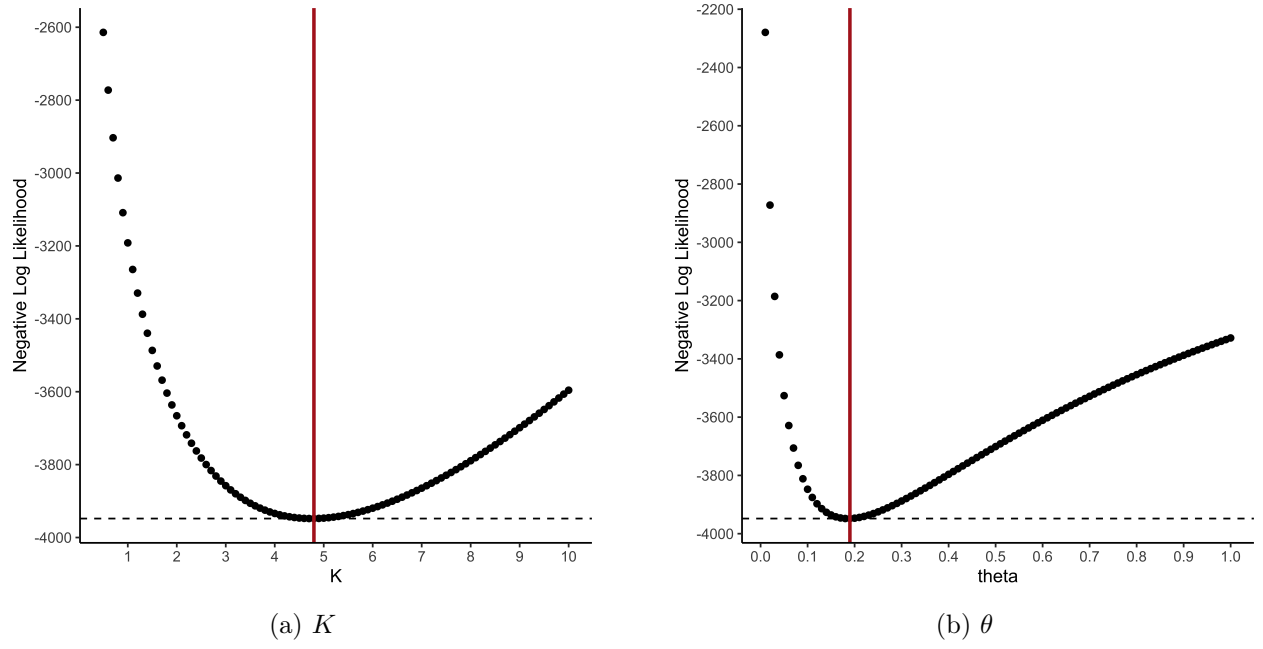(a) K (b) θ

Figure 2: Negative log-likelihood plot of parameters for HawkesN model.
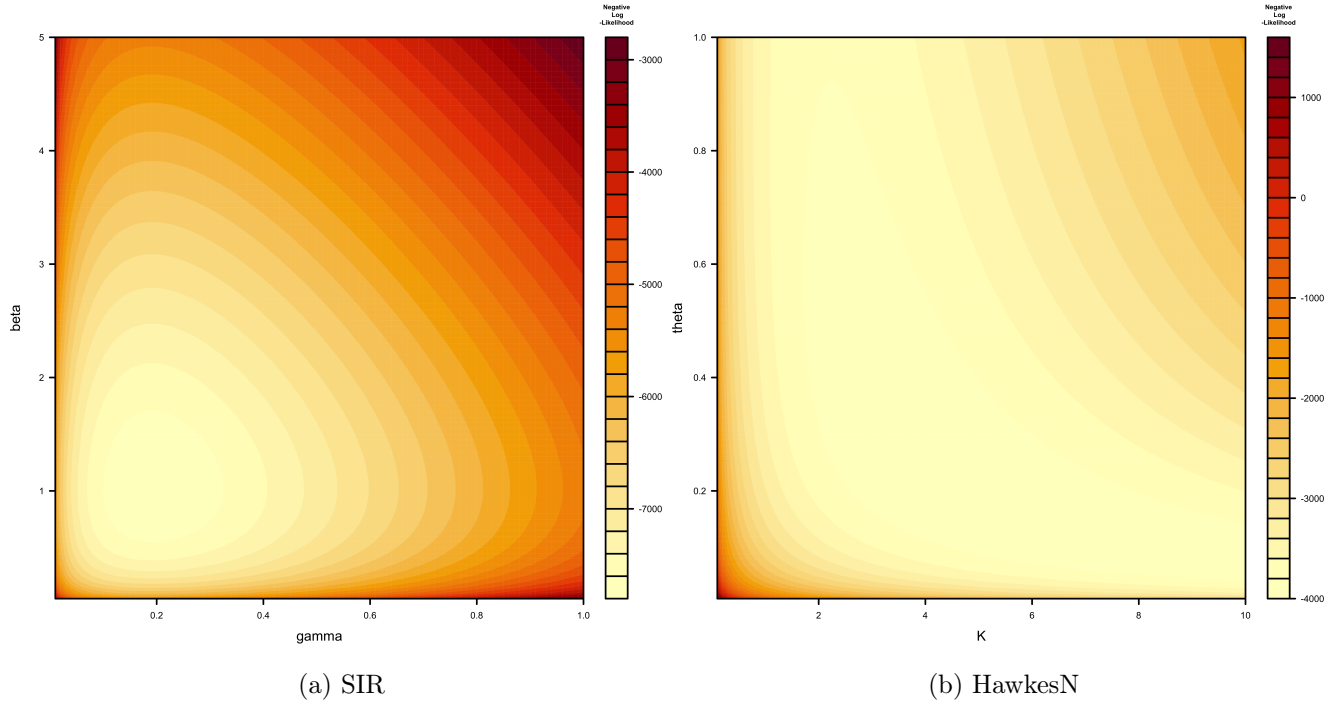

(a) SIR   (b) HawkesN

Figure 3: Contour plots of parameter negative log-likelihoods for HawkesN and SIR models.