Modeling COVID-19 and Parameters Estimation of Hawkes Process

By

Ayman Abbas Najm
College of Administration and Economics
University of Al-Qadisiyah
admin.stat21.8@qu.edu.iq

Dr. Muhannad Fayez Kazem
College of Administration and Economics
University of Al-Qadisiyah
muhannad.alsaadony@qu.edu.iq

Abstract

This paper deals with the study of stochastic self-exciting processes called Hawkes processes, where many accidents during their occurrence usually form data over time represented by the so-called cluster events, meaning that the occurrence of the event is represented by cluster samples in which the occurrence of each event stimulates the occurrence of another event at an accelerated rate is similar to a cluster. Hawkes processes are a type of stochastic processes that can be classified in many types of data that are characterized by their occurrence followed by the occurrence of accidents in an accelerated manner, such as the occurrence of aftershocks after certain earthquake or the occurrence of trading operations in a stock market or stock market after a certain jump in trading as a result of a certain circumstance, which drives market traders to tend towards speculations, whether by buying or selling. That is, Hawkes processes are stochastic processes that depend in their analysis of accidents on the effects resulting from the occurrence of previous accidents, and this is what is called the effect of the self-exciting of the event. Real data analysis was conducted to study the behavior of the behavior of the data of incidents of the number of people infected with the COVID-19 virus via Hawkes process. The estimation methods, DNMLL and AEMA, have used to estimate the Hawkes process parameters. The bias criterion and standard deviation have used to assess the performance of the estimation methods in terms of the quality of the estimators. The result shows the performance of the DNMLL method in estimation. In addition the Kolmogorov-Smirnov test have used ensure that the data follows the standard exponential distribution (Hawkes process), as well as testing the stationary of the Hawkes process b using the branching ration test.

Keywords: Hawkes process, exponential decay, DNMLL, AEMA, COVID-19.

1 Introduction

The global pandemic disease Covid-19 has spread out in most countries in this world, industrialized countries, consuming countries, developing and capitalist countries. In the end of 2019, this respiratory disease has infected millions of people around the world causing thousands of death cases [8]. The global economy in most of the infected countries faces a big challenge, as well as the global health situation, where there is no vaccine available. Consequently, this required a global effort to stop the spread of this pandemic [7]. This paper investigates studying the behavior of this pandemic through using a mathematical model that may be useful to local and state governments and health care [6]. The Covid-19 rapidly spreads and transmission by human-to- human. This motivates the scientific community to understand the behaviors of this pandemic, so we will attempt to understand the characteristics of this pandemic form statistical modeling perspective. So, this paper introduced the well know mathematical model named Self-exciting process that proposed by [1], this process also known as Hawkes process. The studying of this mathematical model aims to develop guidelines intervention to prepare the health care sector to take precautions for reducing critical care and mortality. Many counties have resorted to strict orders, such as, closing businesses, disrupting official working hours, and staying at home (Shelter-in-place) which negatively affected the economy of these countries. Many mathematical models have used to study the behavior or modeling the infectious diseases, among which, the stochastic process models are commonly used to study the behavior of events that are observed clustery in nature over the time [4]. Mathematical model (Hawkes process) formulation and forecasting are playing a pivotal role in the outbreak of Covid-2019 disease. So, we can say that the Hawkes process is an accurate mathematical model that tracks the behavior of Covid-2019 transmission in one dimension. Also, the goal of this paper is to show that self-exciting process Hawkes process commonly used in the field of machine learning approaches to model contagion behaviors in event data, also is to show that self-exciting process are an appropriate for modeling Covid-2019 outbreak, as well as is a suited for casting the future cases with mortality.

2 Preliminary and Basic Concepts

To begin, in order to understand the self- exciting process in a precise, we have to define the point process, count process, and non-homogeneous Poisson process as follows:

Definition 1: (point process) [8]

The stochastic process of sequence of non-negative random variables $\{T_i ; i \in N\}$ which satisfy:

1) $P(0 \leq T_0 \leq T_1 \leq \cdots \} = 1$
2) $P(T_i < T_{i+1}, T_i < \infty) = P(T_i < \infty), \forall\ i \geq 1$
3) the number of points (T_i) in a bounded region is finite a.s.

The we call $\{T_i ; i \in N\}$ is a (simple) point process, the following definition is with help of definition 1.

Definition 2: (Counting process) [10]

The point stochastic process of sequence of random variables that satisfy $\{T_i , i \in N\}$, then the following self-exciting with continuous right jumps:

$$N(t) = \sum_{i \in N} I_{\{T_i \leq t\}} ,$$

where

$$I_{\{Ti \leq t\}} = \begin{cases} 1 & if\ \ T_i \ \leq t \\ 0 & if\ \ T_i > \ t \end{cases}$$

is called counting process associated with the point process $\{T_i , i \in N\}$.

Definition 3: (Nonhomogeneous Poisson process) [2] [19]

Suppose that $\{N(t), t \geq 0\}$ is a counting process the poisons process is called nonhomogeneous Poisson process with intensity function $\lambda(t), t \geq 0$, if satisfy:

1) $N(t = 0) = 0$
2) the process $\{N(t), t \geq 0\}$ has independent increments,
3) $P(N(t + h) - N(t) \geq 2) = 0(h)$,

4) $P(N(t+h) - N(t) = 1) = \lambda(t)h + 0(h)$.

Def.4: Conditional Intensity Function [5] [13]

Suppose the count process $N(.)$ and $H(.)$ is stochastic process of History data for $N(.)$, the if $\lambda^*(t)$ exist function that satisfy:

$$\lambda^*(t) = \lim_{h \to 0} \frac{E[N(t+h) - N(t)|H(t)]}{h}$$

Where $\lambda^*(t)$ is the expected rate of arrivals based on the history function and is called the conditional intensity function for $N(.)$.

Now, we can define the self-exciting (Hawkes process) as follows:

Definition 5: Hawkes (Self-exciting) process [3] [4]

Suppose that $\{N(t), t \in R\}$ is a counting process, let $H(t)$ is the history process $\{H(t), t \in R\}$ than $\{N(t), t \in R\}$ is called Hawkes Process if satisfy:

1) $P(N(t+h) - N(t) = 1|H(t) = \lambda^*(t)h + 0(h)$,
2) $P\big(N(t+h) - N(t) > 1\big|H(t)\big) = 0(h)$,
3) $P\big(N(t+h) - N(t) = 0\big|H(t)\big) = 1 - \lambda^*(t)h + 0(h)$.

Where $\lambda^*(t) = \lambda_0 + \int_{-\infty}^{t} \mu(t-u)dN(u)$ is the conditional intensity function, $\lambda \in R^+$, $\mu: R^+ \longrightarrow R^+ \cup \{0\}$. λ_0 is called the initial(base) intensity, and μ is the excitation function.

3. One-dimensional Hawkes process with exponential decay function) [13] [3]

The very popular excitation function $M(.)$ that used in Hawkes process is the exponential decay function that defined as follows :

$$M(t) = \alpha e^{-\beta t}\,; \quad \alpha, \beta > 0 \qquad (1)$$

So by substituting this decay function in the function of intensity, we have:

$$\lambda^*(t) = \lambda_0 + \int_0^t \alpha e^{\beta(t-T_i)}\ dN(T_i)$$

$$= \lambda_0 + \alpha \sum_{T_i < t} e^{-\beta(t-T_i)} \qquad (2)$$

The stationary condition of One-dimensional Hawkes process should holds the following condition,

$$\widehat{M} = \int_0^{\infty} M(T_i)\, dT_i$$

$$= \int_0^{\infty} \alpha\, e^{-\beta T_i}\, dT_i = \frac{\alpha}{\beta} < 1 \iff \alpha < \beta \qquad (3)$$

Then under some initial value of $\lambda^*(0) = \lambda_0$, we can intensity function as follows:

$$\lambda^*(t) = e^{-\beta t}(\lambda_0 - \lambda) + \lambda + \int_0^t \alpha e^{-\beta(t-T_i)}\, dN(T_i) \qquad (4)$$

The function (2) can be seen as an extension for the function (1), intensity function explained that the process $(\lambda^*(t), N(t))$ is a continuous time Markov process, see (2) for more details.

4. Goodness of fit test [17] [14]

In this paper we will use the simplest way to test the goodness of fit for the Hawkes process by proposing that the residuals of the estimated parameters follows the standard exponential distribution through using the kolmogorov-Smirnov test (K-S) that used the following statistics:

$$\mathrm{D}_n = \sup_x |\mathrm{F}_\mathrm{n}(x) - \mathrm{F}(x)|$$

Her F(x) is the c.d.f of standard exponential distribution and $F_n(x)$ is empirical c.d.f, where

$$\mathrm{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{[0,\mathrm{x}]}(e_i)$$

Here $\mathrm{I}_{[0,\mathrm{X}]}$ is an indicator function and e_i is the residuals, the k-s test will apply for the following hypotheses:

H_0: residuals~exp (1)

H_1: residuals$\nsim$exp (1)

5. One-dimensional Hawkes process parameter estimation[5] [9]

In this section, we will briefly explain two methods that we used for estimation the one-dimensional Hawkes process, the first method discussed in (16) and (11) which is called the direct numerical Maximization log-likelihood (DNMLL), and the Secord estimation of parameters method called the approximation expectation Maximization algorithm (AEMA) that discussed in works of (1) and (7). So, in order to fit the exponential decay function in Hawkes process we will illustrate the mechanism of the above estimation methods that depends on the Maximum likelihood estimation.

5.1 DNMLL Method [12] [15]

For the One-dimensional Hawkes process (2) and for the point process $\{X_t\}_{t\in(0,T]}$, we supposed that the likelihood function for the Hawkes process is defined by :

$$L(\theta) = \prod_{i=1}^{N(T)} \lambda_{t_i} \exp(-\int_0^T \lambda_u \, du) \ , \qquad (5)$$

Based on the likelihood function ( ) the log-likelihood function is defined as follows:

$$\log L(\theta) = \sum_{i=1}^{n} log\lambda^*(T_i) - \int_0^{t_n} \lambda^*(u)du$$

Where the last integral can be partitioned into set of integrals with different; interval supports;

$$\Lambda(t_n) = \int_0^{t_n} \lambda^*(u)du$$

$$= \int_0^{t_1} \lambda^*(u)du + \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} \lambda^*(u)du$$

Now, as we assumed that the function of intensity will follows the exponential decay function then, we have

$$\Lambda(t) = \int_0^{t_1} \lambda du + \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} \lambda + \sum_{t_{i<u}} \alpha e^{-\beta(u-t_i)} du$$

$$= \lambda t_n + \alpha \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} \sum_{j=1}^{i} e^{-\beta(u-t_i)} du$$

$$= \lambda t_n + \alpha \sum_{i=1}^{n-1} \sum_{j=1}^{i} \int_{t_i}^{t_{i+1}} e^{-\beta(u-t_i)} du$$

$$= \lambda t_n - \frac{\alpha}{\beta} \sum_{i=1}^{n-1} \sum_{j=1}^{i} [e^{-\beta(t_{i+1}-t_j)} - e^{-\beta(t_i-t_j)}]$$

$$= \lambda t_n - \frac{\alpha}{\beta} \sum_{i=1}^{n} [e^{-\beta(t_n-t_i)} - 1]$$

The calculations form last equation is not easy because there are two summations with different subscripts, (11) stated that the last equation of Λ(t) can be addressed by the following procedure,

$$A(i) = \sum_{j=1}^{i-1} e^{-\beta(t_i-t_j)}$$

$$= e^{-\beta t_i + \beta t_{i-1}} \; e^{\beta t_j + \beta t_{i-1}} \sum_{j=1}^{i-1} e^{-\beta t_i + \beta t_j}$$

$$= e^{-\beta t_i + \beta t_{i-1}} \sum_{j=1}^{i-1} e^{-\beta t_{i-1} + \beta t_j}$$

$$= e^{-\beta(t_i - t_{i-1})}[1 + \sum_{j=1}^{i-2} e^{-\beta(t_{i-1}-t_j)}]$$

$$= e^{-\beta(t_i-t_{i-1})}[1 + A(i-1)]$$

Consequently, there is no closed form for Maximum likelihood estimators of (λ, α, β), therefore we have to use the numerical Maximization by using R programing in order to find the solution of (λ, α, β).

5.2 Approximate Expectation Maximization Algorithm (AEMA) method [18] [15]

The AEMA method is an alternative computing method for estimating the maximum likelihood estimators Also the AEMA reduces the computation step in EM algorithm , where there are two steps, the first one is the expectation step (E-step) and the second one is the maximization step (M-step). In 2008 (16) discussed the using of EM algorithm to estimate the parameters under unobservable structure (latent variables). However, (7) discussed that AEMA with the exponential decay Hawkes process gives closed form for the parameters of Hawkes process. Consequently, the AEMA solution is as follows:

$$\lambda_{\infty}^{(K+1)} = \frac{\sum_{i=1}^{N_T} \Pr\left\{ui = i \middle| F_{t_1}, \theta^{(K)}, t_i\right\}}{T}$$

$$\alpha^{(K+1)} = \frac{\beta^{(K+1)} \sum_{i=2}^{NT} \sum_{i=1}^{i-1} \Pr\left\{u_i = j \middle| F_{t_1}, \theta^{(K)}, t_i\right\}}{N_T}$$

$$\beta^{(K+1)} = \frac{\sum_{i=2}^{N_T} \sum_{i=1}^{i-1} \Pr\left\{u_i = j \middle| F_{t_1}, \theta^{(K)}, t_i\right\}}{\sum_{i=2}^{N_T} \sum_{j=1}^{i-1} \left(t_i - t_j\right) \Pr\left\{u_i = j \middle| F_{t_1}, \theta^{(K)}, t_i\right\}}$$

6. Real Data Analysis

In this section we first explain the data that representing the incidence of this epidemic in the Holy Karbala Governorate, which is officially published on the website of the Iraqi Ministry of Health, by taking the number of infections on a daily basis for the months (April, May, June, July, August, September) of the year 2021, we relied on each day, as the total number of injuries during the aforementioned period where there are (27,694) cases. We will model the Covid-19 effect numbers by the exponential decay Hawkes process. Also, K-S goodness of fit test has used to ensure that the number of effects in Covid-19 individuals follows standard exponential distribution, after that we will employ the DNMLL and AEMA methods for estimating the Hawkes process. The following Figure show the infections number of COVID-19 for the above period.

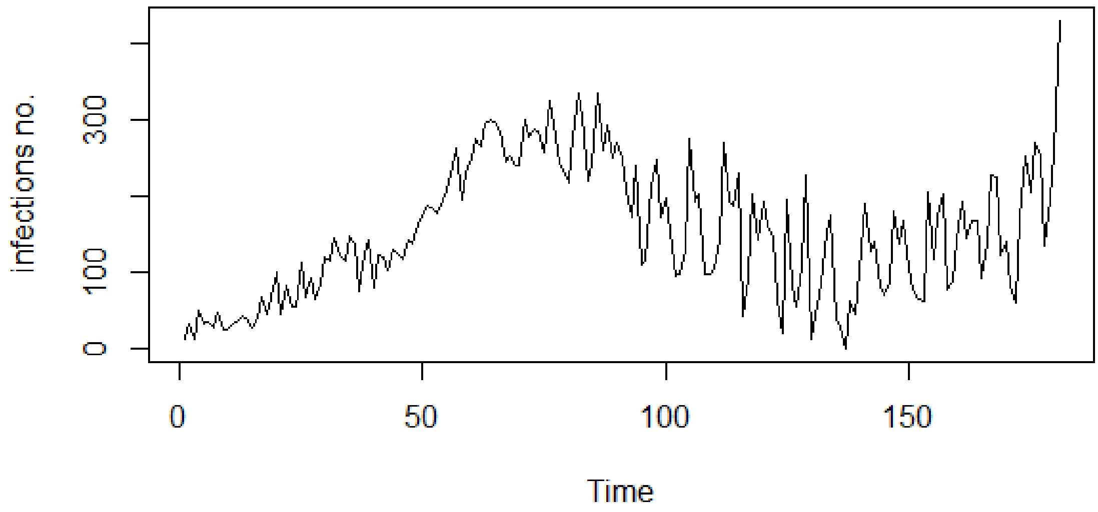

Figure (1): Number of infections of Covid-19

We can see from Figure (1) the peaks indicate that the infection numbers have occurred in order of cluster points. So, we conclude that the behavior of the infection number can be model though Hawkes process. Now, we will employ the methods DNMLL AND AEMA to estimates the parameter of the univariate Hawkes process in (2). The following Table gives the parameters estimates.

Table 1: Estimation results of univariate Hawkes process

Method	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}/\hat{\beta}$	$E(\lambda)$
DNMLL	0.0586	0.0057	0.0113	0.5044	0.1182
AEMA	0.0793	0.0048	0.0171	0.2807	0.1102

Results in Table 1 shows the parameters estimates of Hawkes process, in addition to the estimated branching ration and the expected number of the infections (0.1068) per minutes which is very closed number to expected numbers of infections that estimated in DNMLL and AEMA methods (0.1182) and (0.1102). Also, the branching ration indicates that the

Hawkes process is stationary process. Thus, after finding the parameters estimates, a simple method to assess the goodness of fit of the Hawkes model is to calculate the residuals with estimated parameters, and then compare the distribution to the standard exponential distribution by using K-S test.

6.1 K-S test

We used the available data of Covid-19 effects to perform the K-S test to ensure that these data follow standard exponential distribution under $\alpha = 1\%$ the following table shows the K-S test results.

Table 2: K-S test results

K-S Statistics	P-Value	Decision
1.560	0.015	did not reject H_0

The K-S test statistics in Table 2 shows that we does not reject the null hypothesis, thus the Hawkes process seems proper to fit the COVID-19 data.

7 Conclusions

In this paper, we show how the One-dimensional Hawkes process can be employed to accurately model the Covid-19 pandemic. Two parameter estimation methods have used (DNMLL) and (AEMA) to estimate the values of unknown parameters of Hawkes process. This parameter estimation leads to statistical inference about this parameter to understand the behavior of this pandemic. First, we ensured that the infection numbers of Covid-19 has follows the standard exponential distribution which implies that data follows the Hawkes process. The second step we estimates the unknown parameters and calculate the branching ratio to capture the stationary of the Hawkes process and then we estimates the expected number of infection per-minute. Consequently, the expected number of infection coincides with the actual number, which indicates the well performing of Hawkes process for this kind of data.

References

[1] Erik Lewis and George Mohler. “A nonparametric EM algorithm for multiscale Hawkes processes”. In: preprint (2011), pp. 1–16.

[2] Gao, X., Zhou, X., & Zhu, L. (2018). Transform analysis for Hawkes processes with applications in dark pool trading. Quantitative Finance, 18(2), 265-282.

[3] Haghdan, M. (2017). Hawkes Process Models for Unsupervised Learning on Uncertain Event Data (Doctoral dissertation, University of Toledo).

[4] Hawkes A. G., Spectra of some self-exciting and mutually exciting point Processes, Biometrika, Vol. 58, No.1, 1971a ,pp.83-90.

[5] Hawkes A. G. Point spectra of some mutually exciting point processes, Journal of the Royal Statistical Society, Series B (Methodological), Vol.33, No. 3, 1971b, pp.438-443.

[6] Hawkes A. G., Oakes, D., A cluster process representation of a self-exciting process. Journal of Applied Probability, Vol. 11, No. 3, 1974, pp.493-503.

[7] Jamie F Olson and Kathleen M Carley. “Exact and approximate EM estimation of mutually exciting hawkes processes”. In: Statistical Inference for Stochastic Processes 16.1 (2013), pp. 63–80.

[8] Laub, P. J., Taimre, T., & Pollett, P. K. (2015). Hawkes processes. arXiv preprint arXiv:1507.02822.

[9] Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R.,Leung, K. S.M., Lau, E.H.Y., et al., Early transmission dynamics in Wuhan, China of novel coronavirus infected pneumonia. New Engl. J. Med.,382,2020,pp.11991207.

[10] Obral, K. (2016). Simulation, estimation and applications of Hawkes processes (Doctoral dissertation, Master’s thesis, University of Minnesota).

[11] Yoshiko Ogata. “The asymptotic behaviour of maximum likelihood estimators for stationary point processes”. In: Annals of the Institute of Statistical Mathematics 30.1 (1978), pp. 243–261.

[12] Ozaki T., Maximum likelihood estimation of Hawkes' self-exciting point Processes, Annals of the Institute of Statistical Mathematics, Vol.31, No. 1, 1979, pp. 145-155.

[13] Peter F. H., Paul D. B., Modelling dyadic interaction with Hawkes Processes, Psychometrika, Vol.78, No. 4, 2013, pp.793-814.

[14] Rodriguez.A.B.(2019). Hawkes processes in finance. thesis. university of Barcelona. http://diposit.ub.edu.

[15] Saroj K. B. , Nasir U. A., Mathematical modeling and optimal intervention of COVID-19 outbreak, Quantitative Biology, Vol.9, No. 1, 2021, pp. 8492.

[16] Veen, A., & Schoenberg, F. P. (2008). Estimation of space–time branching process models in seismology using an em-type algorithm. Journal of the American Statistical Association, 103(482), 614-624.

[17] Vere-Jones, D. (1978). Earthquake prediction-a statistician's view. Journal of Physics of the Earth, 26(2), 129-146.

[18] Wang.Q.(2015).Applications of Hawkes process in Finance. Master thesis. Tilburg University. http://arno.uvt.nl.

[19] Zhao S., Chen H. , Modeling the epidemic dynamics and control of COVID-19 outbreak in China, Quantitative Biology, Vol.8, 2020, pp.1119.