A Self-limiting Hawkes Process: Interpretation, Estimation, and Use in Crime Modeling

Jack Olinde
School of Mathematics
Georgia Institute of Technology
Atlanta, GA, USA
jolinde@gatech.edu

Martin B. Short
School of Mathematics
Georgia Institute of Technology
Atlanta, GA, USA
mbshort@math.gatech.edu

***Abstract*—We introduce a self-limiting Hawkes process, building off of the standard Hawkes process and self-correcting point process in the literature. The self-limiting Hawkes process is intended to model situations in which competing forces – one driven by self-excitation and one by the tendency to block events from occurring – can be effectively captured by considering a single stochastic intensity. The example presented here is urban crime. We show that under some circumstances the parameters of the self-limiting Hawkes process can be well-estimated by adapting existing techniques. We apply these techniques to crime data, showing that the self-limiting Hawkes process is a plausible alternative to the standard Hawkes process for modeling crime.**

***Index Terms*—Crime, expectation-maximization algorithms, Hawkes process, maximum likelihood estimation, stochastic point processes**

I. Introduction

Many real-world stochastic systems appear to exhibit a self-exciting tendency, a phenomenon whereby the occurrence of these stochastic events seems to cause an increase in the rate of occurrence of subsequent events, at least locally in time (and potentially space). Some examples include earthquakes [1], financial markets [5], and various forms of communication [4]. One common model used to describe these systems is the Hawkes process [6], a linear model that is particularly amenable to fitting to potentially self-exciting datasets.

Another self-exciting system that has been modeled by the Hawkes process is urban crime. Various criminological theories and studies [9], [10] note the existence of "repeat victimization", whereby criminals have a tendency to commit their crimes at or against places or people who have previously been victimized. In [13], this basic phenomenon was cast in the form of a Hawkes process to describe repeat victimization in burglary data; other studies have followed [12].

But in the case of crime, there is another factor at play - the actions of police, one of whose goal is to prevent crimes from occurring in the first place. Indeed, in [12], the Hawkes process fit to up-to-date crime data was used in conjunction with police forces to inform police patrols, with measurable success. However, there is a subtle issue involved here that has not previously been addressed: given that past crime data was presumably influenced in some way by the past actions of the police, but the Hawkes process model does not explicitly capture this interaction, any estimates of the Hawkes process using past crime data will also implicitly include prior police effects. Using these fits to inform future police actions is therefore questionable, even if we have a good model for how police might influence true crime rates, as our estimates of the stochastic crime rates already include in some unknown way the effect of police. This effect is compounded by the fact that police actions themselves are typically influenced by those crimes that do occur, such that a feedback loop exists in the crime-police system, which should alter the estimated Hawkes process in some non-trivial way.

Within the point-process literature, there are models variously termed as self-correcting [7]. These models differ from a standard self-exciting Hawkes process in that events are typically modeled as decreasing the intensity of the process via multiplication by some positive factor less than unity. A common feature of these models is an exogenous rate of increase of the stochastic intensity over time, to offset the intensity decreases accompanying the events themselves. While these models have the flavor of what we want to capture in our crime example – a police-like effect limiting growth of the event rate – they don't explicitly capture the tension between self-excitation and self-correction that we believe the crime-police system ought to exhibit.

For this reason, we introduce here what we refer to as a self-limiting Hawkes process. The specific motivation is, as discussed, the crime-police system, but the model, and the methods we show to simulate it and estimate it from data, could be of potential interest in other domains whereby control is often exercised or desired over the occurrence of self-exciting events. The paper is structured as follows. In section II we provide some background on the mathematics of the Hawkes process and methods of its estimation. In section III we introduce the self-limiting Hawkes process model and describe methods of simulation. In section IV we provide a method for fitting the self-limiting Hawkes process to data, testing on simulated data to verify these methods. Finally, in section V we test the model on real crime data, finding that in certain cases, the self-limiting Hawkes model out-performs the standard Hawkes model.

II. Background

Before we define a Hawkes process, we will first define the conditional intensity of a point process as

$$\lambda(t) = \lim_{dt \to 0} \frac{\mathbb{E}[N(t, t+dt) \mid H_t)]}{dt},$$

where H_t is the history of the process up until time t and N counts the number of points in the interval $[t, t+dt)$ given H_t. This leads us to the definition of a Hawkes process, which has conditional intensity

$$\lambda(t) = \mu(t) + \sum_{i:t_i<t} g(t-t_i), \tag{1}$$

where $\mu(t)$ is the background intensity of the process and g is a function that describes the self-exciting property of the process [6]. For the purpose of this paper, we will use $\mu(t) = \mu \in \mathbb{R}^+$ and the function $g(t-t_i) = k\omega e^{-\omega(t-t_i)}$ for $t > t_i$, $g(t-t_i) = 0$ for $t \leq t_i$.

One way to conceptualize a Hawkes process over an interval of time $[0, T]$ is as a sum of individual Poisson processes: $\lambda_0(t) = \mu(t), \lambda_1(t) = g(t-t_1), \ldots, \lambda_n(t) = g(t-t_n)$. Each Poisson process creates a generation of points upon which the following Poisson processes are based. Intensity $\lambda_0 = \mu(t)$ has no conditions, so it defines a Poisson process on the whole interval $[0, T]$. Events that arise from intensity λ_0 are referred to as background events. Intensities $\lambda_i,\ i > 0$ do not activate until $t > t_i$, where t_i is the i^{th} point in the overall process. So λ_i defines a Poisson process on the interval $[t_i, T]$. Events arising from intensity $\lambda_i,\ i > 0$ are called daughter events, and the parent event of each of these daughters is event i.

A. Simulating a Hawkes Process

One way to simulate a Hawkes process is using what is known as the thinning method. This method was first proposed as a way to simulate non-homogeneous point processes [11], but has since been modified to simulate Hawkes processes [14]; it is especially useful if the excited kernel is not exponential. The algorithm is described in Fig. 1.

When applied to the Hawkes process, the thinning method first simulates background events from intensity $\mu(t)$. Then, it simulates all of the direct offspring of the background events through the various g kernels that the background events produce. Then another excited kernel is simulated for each of these daughters, etc. This is repeated until no new points are simulated for the last round of excited kernels.

Though this method for simulating a Hawkes process is intuitive and can be implemented relatively easily, there are more efficient methods of simulation available, especially when the excited kernel g is an exponential. One particular example is the method of Dassios and Zhao [3]; this algorithm will be discussed in more detail in Section III.

B. Estimating the Parameters of a Hawkes Process

Here we present a review of the Expectation-Maximization (E-M) method [15] to estimate the parameters μ, k, and ω

Input: the intensity function $\lambda(t)$
Output: a realization of the point process, $\{t_1, \ldots, t_n\}$

1) Define $M = \max\{\lambda(t) : t \in [0, T]\}$ and let $N \sim \text{Pois}(MT)$.
2) Place N points uniformly at random in the interval $[0, T]$.
3) For $i = 1, \ldots, N$, delete point t_i with probability $1 - \frac{\lambda(t_i)}{M}$.

Fig. 1. Thinning Method for Simulating a Non-homogeneous Point Process

of a Hawkes process from data. First, the log-likelihood of a process given intensity $\lambda(t)$ and data $\{t_1, \ldots, t_n\}$ is

$$\mathscr{L} = -\int_0^T \lambda(t)dt + \sum_i \ln(\lambda(t_i)). \tag{2}$$

To use this formula for a Hawkes process, first suppose we knew the true branching structure of the process: which events were background events and which were daughters, along with which event was the parent of each daughter. Then we could rewrite $\mathscr{L}$ as

$$\mathscr{L} = \sum_{i \in B} \ln(\mu) - \int_0^T \mu dt + \sum_{i \in D} \ln\left(k\omega e^{-\omega(t_i - t_{p(i)})}\right) - \int_0^T \sum_{i:t_i<t} k\omega e^{-\omega(t-t_i)} dt,$$

where B and D are the sets of background and daughter events, respectively, and $p(i)$ is defined as the parent event of event of i.

Though we generally don't know the true branching structure of the process, we will assume we can still generate a probabilistic branching structure P, where

$$P_{ij} = \begin{cases} \text{prob. that } i \text{ is a background event} & , i = j \\ \text{prob. that } i \text{ is a daughter of } j & , j < i \end{cases}. \tag{3}$$

Taking the expectation of $\mathscr{L}$ with respect to P gives us what is called the complete data log-likelihood [15]:

$$\mathbb{E}[\mathscr{L}] = \ln(\mu) \sum_i P_{ii} + \ln(k\omega) \sum_{j<i} P_{ij} - \omega \sum_{j<i} P_{i,j}(t_i - t_j) - \mu T - k \sum_i \left(1 - e^{-\omega(T-t_i)}\right). \tag{4}$$

One can then maximize $\mathbb{E}[\mathscr{L}]$ with respect to μ, k, and ω in a variety of ways, including by taking the respective partial derivatives with respect to each variable and setting them equal to 0 to obtain analytical formulas for the optimum values; for more details, see [15]. Of course, we must still specify P_{ij} in order to use (4). But, since a Hawkes Process can be thought of as a sum of Poisson processes, we have

$$P_{ij} = \begin{cases} \frac{\mu}{\lambda(t_i)} & i = j \\ \frac{k\omega e^{-\omega(t_i - t_j)}}{\lambda(t_i)} & j < i \end{cases}. \tag{5}$$

Input: An initial guess for μ, k, ω and a tolerance ϵ
Output: μ, k, ω

1) For each event pair $j \leq i$, calculate P_{ij} using the current values of μ, k, and ω using (5). This is the Expectation step of the E-M algorithm.
2) Update our values of μ, k, and ω using these P_{ij} by maximizing (4). This is the Maximization step of the E-M algorithm.
3) Repeat steps 2 and 3 until some measure of convergence, given the desired tolerance ϵ, is achieved.

Fig. 2. E-M Algorithm for Estimating Hawkes Process Parameters

From here, we can see that we need μ, k, and ω to calculate P_{ij} and P_{ij} to calculate μ, k, and ω. This leads us to the iterative E-M (Expectation-Maximization) method given in Fig. 2.

III. A Self-limiting Hawkes Process

We now turn to the development of our model for a self-limiting Hawkes process. Recall that the overall goal is to model a stochastic process with two competing properties: 1) the process should have self-excitation, for which a Hawkes process can serve as a baseline and 2) the model should incorporate a mechanism by which the intensity of the process can also be reduced by the occurrence of events, to represent potentially exogenous influences such as police activity. To capture this second, self-limiting effect, we introduce two new parameters, α and β, and define $N(\alpha, t)$, which counts the number of events that occurred through the process on interval $[t-\alpha, t)$. Then our model of a self-limiting Hawkes process intensity is

$$\lambda(t) = \left(\mu + k\omega \sum_{i:t_i<t} e^{-\omega(t-t_i)}\right) e^{-\beta N(\alpha,t)}. \qquad (6)$$

Parameter β therefore represents the strength of self-limiting, with greater values decreasing the intensity more than smaller values, and α represents a time-window over which any given event can contribute to self-limiting of the overall process. If the process is being used to model criminal events, then we can think of α as the memory of the police and β as the increase in police deterrent activity for each additional crime that occurs in the interval $[t-\alpha, t)$.

Equation (6) can be interpreted in the following mechanistic way, which aids in simulating. In the absence of self-limiting, the process would behave as a standard Hawkes process, and would generate some sequence of hypothetical events. However, the self-limiting effect is such that each event t_i that does in fact occur via the process causes every subsequent hypothetical event within the period $(t_i, t_i+\alpha]$ to be probabilistically "blocked" from occurring, with probability $p = 1-e^{-\beta}$. If we assume that multiple overlapping blockings of a single hypothetical event are probabilistically independent, then the probability of a hypothetical event at time t_j not being blocked is $e^{-\beta N(\alpha,t_j)}$. Hence, the intensity of (6) tells us that events occur only when the underlying Hawkes process would hypothetically cause them to occur, and only if they are not probabilistically blocked by some of the prior events that did in fact occur (weren't blocked themselves).

Using this interpretation, one could create a straightforward Poisson-thinning type algorithm to simulate the self-limiting Hawkes process. Specifically, first simulate the underlying Hawkes process by itself, without any self-limiting effect, being sure to retain the true branching structure of the process. Then, starting with the first event and working sequentially, retain each event with probability $e^{-\beta N(\alpha,t_j)}$. If an event t_i is retained, continue to the next event. If event t_i is not retained, remove it from the list of event times and also remove all subsequent events that are descendants (either directly or indirectly) of t_i in the branching process, then proceed to the next event.

While the above process is straighforward to describe, it is not very computationally efficient. Hence, we also provide a more efficient algorithm, which is a modified version of the algorithm of Dassios and Zhao [3], and which incorporates the preventative action right into the generation of the Hawkes process. This algorithm does have the drawback of only being valid for the exponential excited kernel g that we are using here. Rather than simulate the process layer-by-layer as described in section II-A and above, the method given by Dassios and Zhao starts at time $t = 0$ and jumps forward by simulating the time interval Δt until the next event occurs. Each such Δt is found by randomly generating two possible values: one from the background rate μ and one from the full summation of the excited kernels, which is itself simply a decaying exponential. The smaller of these two times is then chosen as Δt, time is incremented by this value, and the fully summed excited kernel is updated via exponential decay and an increase by k and the algorithm continues. To modify this method to account for the self-limiting aspect, we simply add a step where each event to be added is only added with probability $e^{-\beta N(\alpha,t_j)}$, where t_j is the time of the potential new event. If t_j is added, the algorithm continues just as in the Dassios and Zhao method. If t_j is not added, increment time by Δt and factor in the exponential decay of the excited kernel, but do not increment the excited kernel as one normally would.

We note here that, unlike a standard Hawkes process, our self-limiting process can still remain bounded even if $k > 1$. In a standard Hawkes process, $k > 1$ means that each event on average gives rise to more than one daughter event, generally causing the intensity to grow exponentially in time. However, the self-limiting process avoids this through the $e^{-\beta N(\alpha,t)}$ term. If $k > 1$ starts to cause λ to grow very large, then the number of events N will also grow, and the exponential dampening will force the value of λ back down. In Fig. 3, we illustrate two realizations of a self-limiting Hawkes process, one of which has $k > 1$.

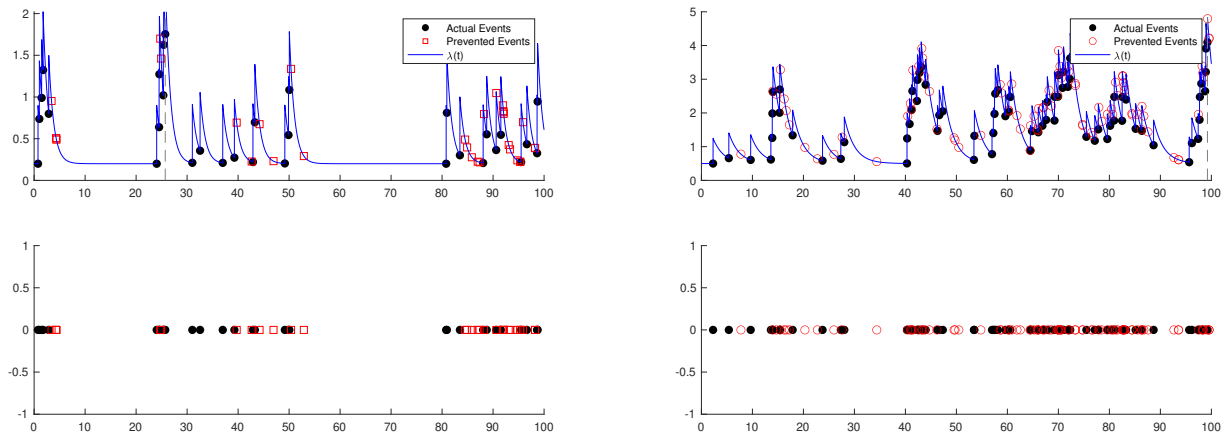

Fig. 3. Examples of simulated data from a self-limiting Hawkes model and the intensity function $\lambda(t)$. For the subfigure on the left, the parameters were $\mu = 0.15$, $k = 0.6$, $\omega = 1$, $\alpha = 5$, and $\beta = 0.3$. For the subfigure on the right, the parameters were $\mu = 0.5$, $k = 1.5$, $\omega = 0.5$, $\alpha = 10$, and $\beta = 0.1$.

IV. Estimating the Parameters of a Self-Limiting Hawkes Process

To estimate the parameters of a self-limiting Hawkes process, we modify the Expectation-Maximization procedure described above. First, we note that on a fixed time interval from $[0, T]$, $N(\alpha, t)$ is just a piece-wise constant function

$$N(\alpha, t) = \begin{cases} n_1 & t \in [x_0 = 0, x_1] \\ n_2 & t \in (x_1, x_2] \\ \vdots & \vdots \\ n_l & t \in (x_{l-1}, x_l = T]. \end{cases}$$

The pairs (n_i, x_i) can be easily computed by realizing that $N(\alpha, t)$ increases by exactly one at each event time t_i and decreases by exactly one at each time $t_i + \alpha$. The set $\{x_i\}$ is therefore constructed by taking the union of the two sets $\{t_i\}$ and $\{t_i + \alpha\}$, sorting it, and removing any entries with values greater than T. Then we can write the log-likelihood function by plugging the intensity from (6) into (2) and then taking the expectation with respect to the probabilistic branching structure P to obtain

$$\begin{aligned} \mathbb{E}[\mathscr{L}] = {} & \ln(\mu) \sum_i P_{ii} - \beta \sum_i P_{ii} n_{s(t_i)} + \ln(k) \sum_{j<i} P_{ij} \\ & + \ln(\omega) \sum_{j<i} P_{ij} - \omega \sum_{j<i} P_{ij}(t_i - t_j) - \beta \sum_{j<i} P_{ij} n_{s(t_i)} \\ & - \mu \sum_{i=1}^{l} e^{-\beta n_i}(x_i - x_{i-1}) \\ & + k \sum_{i=1}^{n} \sum_{j=1}^{l} e^{-\beta n_j} \left[e^{-\omega(x_j - t_i)} - e^{-\omega(x_{j-1} - t_i)} \right] \mathbb{1}_E, \end{aligned}$$

where $\mathbb{1}_E$ is the indicator function for the event $E = \{t_i < x_j\}$ and $s(t_i)$ is the index of t_i in $\{x_0, \ldots, x_l\}$. Here P is equivalent to the definition of P in (5) since $e^{-\beta N(\alpha, t_i)}$ is a factor in the numerator and denominator of each fraction, thereby cancelling.

Our log-likelihood function now contains five unknowns (μ, k, ω, α, and β). For μ, k, ω, and β, we can find the respective partial derivatives of $\mathbb{E}[\mathscr{L}]$ and set them equal to 0 in order to maximize the log-likelihood. This gives the following formulas:

$$\mu = \frac{\sum_{i=1}^{n} P_{i,i}}{\sum_{i=1}^{l} e^{-\beta n_i}(x_i - x_{i-1})}, \tag{7}$$

$$k = \frac{-\sum_{j<i} P_{i,j}}{\sum_{i=1}^{n} \sum_{j=1}^{l} e^{-\beta n_j} \left[e^{-\omega(x_j - t_i)} - e^{-\omega(x_{j-1} - t_i)} \right] \mathbb{1}_{\{t_i < x_j\}}}, \tag{8}$$

$$\begin{aligned} 0 = {} & \frac{1}{\omega} \sum_{j<i} P_{i,j} - \sum_{j<i} P_{i,j}(t_i - t_j) \\ & + k \sum_{j=1}^{l} e^{-\beta n_j} \sum_{i=1}^{n} (t_i - x_j) e^{-\omega(x_j - t_i)} \mathbb{1}_E \\ & - k \sum_{j=1}^{l} e^{-\beta n_j} \sum_{i=1}^{n} (t_i - x_{j-1}) e^{-\omega(x_{j-1} - t_i)} \mathbb{1}_E, \end{aligned} \tag{9}$$

and

$$\begin{aligned} 0 = {} & -\sum_i P_{i,i} n_{s(t_i)} - \sum_{j<i} P_{i,j} n_{s(t_i)} \\ & + \mu \sum_{i=1}^{l} n_i e^{-\beta n_i}(x_i - x_{i-1}) \\ & - k \sum_{i=1}^{n} \sum_{j=1}^{l} n_j e^{-\beta n_j} \left[e^{-\omega(x_j - t_i)} - e^{-\omega(x_{j-1} - t_i)} \right] \mathbb{1}_E. \end{aligned} \tag{10}$$

However, as $N(\alpha, t)$ is not differentiable with respect to α, we must maximize over α using some other method. For now, assume α is given, in which case the remaining parameters μ, k, ω, and β could be found using the same basic E-M algorithm given in Fig. 2, with the maximization step done using (7)-(10). In practice, however, simultaneously solving (9) and (10) can be quite computationally demanding. Hence, in our implementations throughout the remainder of this paper, we have chosen instead to perform a parameter sweep over β values, numerically solving only (9) for each of the β values swept over, thereby also obtaining μ and k from (7) and (8), then simply choosing the parameter combination that resulted in the highest value for $\mathscr{L}$.

To estimate α, we also perform a parameter sweep, noting the maximal log-likelihood obtained for each test value of α and then simply selecting that α, and its accompanying μ, k, ω, and β, with the greatest overall log-likelihood.

We tested our method over the following sets of parameters for the preventative action:

$$\alpha \in \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5\},$$

$$\beta \in \{0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045\}.$$

We tested each value in each of these sets by choosing one parameter to vary while holding the other parameter constant at the fifth value in its set. So while we were testing the different values of α, β was fixed at 0.025. Likewise, while we were testing the different values of β, α was fixed at 2.5. For each test, μ, k, and ω were fixed at 0.65, 0.65, and 50, respectively.

For a particular set of values of α and β, 100 realizations of a Hawkes process on time interval $[0, 1000]$ were created with the given parameters. When simulating these processes, we employed the thinning algorithm. This was done so that each realization would yield two datasets: one a standard Hawkes process, which we will refer to as the hypothetical dataset, which could be interpreted as the set of crimes that might have occurred with no police intervention; the other, which we will refer to as the self-limiting dataset, a corresponding subset of the hypothetical dataset representing the full self-limiting process, which can be thought of as the set of crimes that might have actually occurred when police deterrence was in place.

For each realization, the parameters μ, k, and ω were then estimated using the standard Hawkes process E-M algorithm on both resulting datasets, and using our self-limiting E-M algorithm on the self-limiting dataset only. For this test, when using our self-limiting E-M algorithm on the self-limiting dataset, we used the true values of α and β to estimate the other parameters. This allows us to determine the extent to which the self-limiting aspect of the process affects parameter estimation. This is an important point to consider, given that current applications to crime data do not explicitly consider the effects of police activity, and therefore may have systematically biased estimates for parameter values.

To determine how well parameters have been estimated in each of these various scenarios, we consider relative error metrics for each of the estimated parameters. For example, if $\mu_e^{(i)}$ is the estimated value of μ for the i^{th} Hawkes process in one of the three scenarios, then our average relative error over the 100 realizations for that scenario is

$$\frac{1}{100}\sum_{i=1}^{100}\frac{|\mu_e^{(i)} - \mu|}{\mu};$$

corresponding values are computed for the other parameters.

Results are shown in Fig. 4. Each subfigure consists of nine different plots. Each row shows the relative errors in estimates of μ, k, and ω under one of the three scenarios. The top row shows the errors in estimation for the hypothetical datasets using the standard Hawkes E-M algorithm, and serves as a control. The middle row shows the errors when the standard Hawkes E-M algorithm is used on the self-limiting datasets. The bottom row shows the errors in estimation when using our self-limiting E-M algorithm on the self-limiting datasets. Across α and β values, we find that estimation error for each of the three parameters is roughly the same when comparing the hypothetical datasets estimated via standard Hawkes E-M (top rows) to the self-limiting datasets estimated via our self-limiting E-M (bottom rows). This shows that our algorithm is able to estimate the parameters of a self-limiting Hawkes process as well as can be done for a standard Hawkes process of the same parameters. The middle rows show how mis-specification of the model – using standard Hawkes as the model when the process is in reality self-limiting – can lead to significant errors in parameter estimation, in a systematic way. Recall from section III that α can be thought of as the memory of the police force and β can be thought of as the intensity of the police force. So as α and β increase, the number of events prevented within the hypothetical datasets increases. As more events are prevented, we should expect that the accuracy of the estimation via standard Hawkes E-M should decrease as we lose more information about the underlying Hawkes process, which is precisely what we find for parameters μ and k. However, the error in estimating parameter ω is not very sensitive to the precise value of α or β used.

Next, we tested how well the parameters are estimated when using the parameter sweep method for estimating α and β. To do this, we first chose two sets of self-limiting Hawkes parameters on which to test the estimation:

$$\{\mu = 0.65, k = 0.65, \omega = 50, \alpha = 2, \beta = 0.05\},$$

$$\{\mu = 0.65, k = 0.65, \omega = 50, \alpha = 5, \beta = 0.01\}.$$

We then generated 100 self-limiting Hawkes processes for each set of parameters. For each process, we used the self-limiting E-M algorithm along with the parameter sweep method to estimate the parameters. Tables I and II show the average values of each of the five parameters estimated using this method as well as the percent error between the average estimated and true values of the parameters for both sets of true parameters. As we can see, even though the estimation

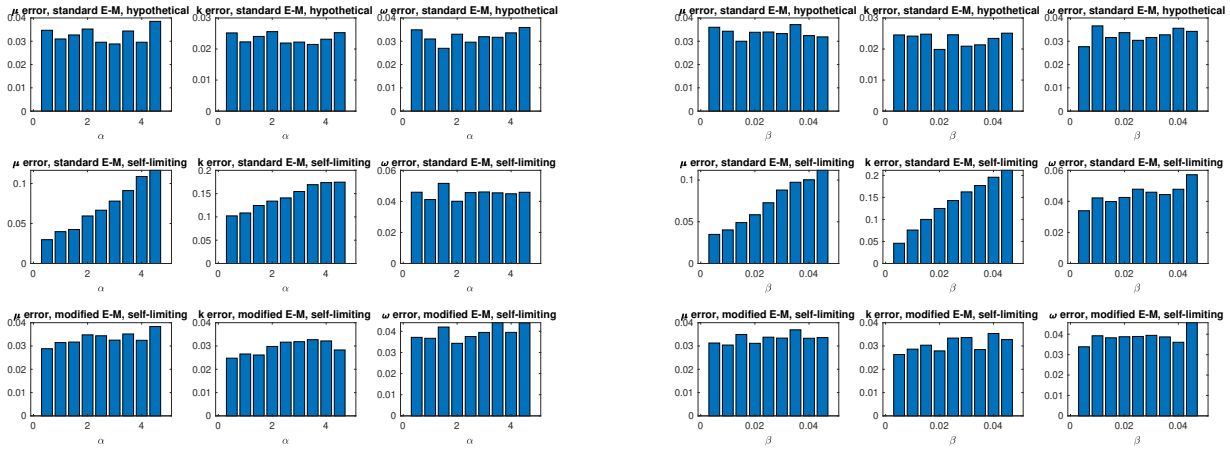

Fig. 4. Parameter estimation error when α is varied (left) and when β is varied (right), under various testing conditions described in the text.

TABLE I
THE AVERAGE ESTIMATED PARAMETERS COMPARED WITH THE TRUE PARAMETERS FOR THE FIRST SET OF TRUE PARAMETERS.

Parameter	True Values	Average Estimated Values	Percent Error (%)
μ	0.65	0.6637	2.11
k	0.65	0.6449	-0.78
ω	50	51.5862	3.17
α	2	3.15	57.50
β	0.05	0.0429	-14.20

TABLE II
THE AVERAGE ESTIMATED PARAMETERS COMPARED WITH THE TRUE PARAMETERS FOR THE SECOND SET OF TRUE PARAMETERS.

Parameter	True Values	Average Estimated Values	Percent Error (%)
μ	0.65	0.6955	7.00
k	0.65	0.7138	9.82
ω	50	50.6129	1.23
α	5	3.745	-25.10
β	0.01	0.0261	161.00

of α and β leads to much higher errors than is found with the other parameters, the values found are still reasonable and lead to accurate estimation of μ, k, and ω.

V. Results Using Real Crime Data

Here, we employ our self-limiting Hawkes process on crime data from Chicago, obtained via their open access portal [2]. We stress here that the main purpose of this analysis is not to establish that the self-limiting Hawkes process is superior the the standard Hawkes process in describing real world crime data; indeed, as we will show, neither model seems especially well fitting to the data analyzed here, for reasons we hypothesize below. Instead, this analysis is meant only to establish the plausibility of the model with respect to an often-used alternative (the standard Hawkes with exponentially decaying excited kernel), and show that one can readily fit the self-limiting Hawkes to real world datasets.

The data set considered contains the times and locations of all burglaries in Chicago from 2001 to 2020. We have chosen to focus on burglary here, as it is often used as an example of a crime type with strong self-excitation. Though location information is provided in the dataset, we are only considering purely temporal processes here. Prior work [8] has shown that parent-daughter crime pairs are often separated by relatively small distances. To allow our temporal processes to account for this, we have binned our data into squares with sides 1500 feet long, considering each such bin separately. In a square this size, any crime could conceivably be the daughter of any crime that occurred before it.

Having spatially binned our data, we only consider the ten squares with the highest total crime counts. For each such bin, we first divide the events into two sets, a training set (the first half of the events) and a testing set (the second half of the events). We then estimate the parameters of the training set using both the standard and self-limiting Hawkes models. As we mentioned in Sec. IV, when estimating the parameters of the self-limiting Hawkes model, we will be using the E-M algorithm to estimate μ, k, and ω, and these estimates will be done independently for each square. For the estimation of α and β, we use the same values for all squares; this seems plausible, as the response of police to crime numbers is likely more consistent across space than crime rates themselves. After performing the sweeps, we choose the values of α and β that maximize the number of squares where the self-limiting Hawkes process outperformed the standard Hawkes process on the square's training set. Here, we say that model A outperforms model B if the parameters estimated using model A result in a higher log-likelihood value and lower Akeike information criteron value (to be defined later) than the parameters estimated using model B. The values of the parameters found in this way are given in Tables III and IV.

To determine goodness of fit for each of these estimates, we compute the residuals $\{\tau_1, \ldots, \tau_n\}$ for each of the two models

TABLE III
THE VALUES OF THE PARAMETERS USING THE STANDARD HAWKES MODEL.

Square	μ (days^{-1})	k	ω (days^{-1})
1	0.0917	0.393	0.0890
2	0.0590	0.603	0.0204
3	0.0948	0.144	0.4550
4	0.0513	0.564	0.0287
5	0.1111	0.046	11.1671
6	0.0987	0.216	0.3034
7	0.874	0.311	0.1268
8	0.0734	0.447	0.0531
9	0.0737	0.443	0.0553
10	0.0720	0.461	0.0620

TABLE IV
THE VALUES OF THE PARAMETERS USING THE SELF-LIMITING HAWKES MODEL USING AVERAGE BEST FIT VALUES $\alpha = 1.124$ DAYS AND $\beta = 0.03$.

Square	μ (days^{-1})	k	ω (days^{-1})
1	0.0821	0.4703	0.0798
2	0.0581	0.6152	0.0275
3	0.0689	0.3901	0.0712
4	0.0585	0.4981	0.0463
5	0.1110	0.0485	10.8960
6	0.0599	0.5416	0.0460
7	0.0823	0.3621	0.1170
8	0.0692	0.4905	0.0592
9	0.0657	0.5177	0.0539
10	0.0710	0.4761	0.0724

in each testing set, where

$$\tau_i = \int_0^{t_i} \lambda(t)dt.$$

Here, $\{t_1, \ldots, t_n\}$ are the times of the crimes in the testing set that occurred in the current square and λ uses our best fit parameters from the training set for that square. Note that the times in the testing set are shifted so that the final event of the training set is time $t = 0$ for the testing set.

If a model correctly represents the data set, then the residuals should be distributed in a way consistent with a unit rate homogeneous Poisson process. Graphically, this means that when plotted as points $(i,\ \tau_i)$, the resulting curve should lie close to the line $y = x$. We measure the goodness of fit of each model using the Kolmogorov–Smirnov test statistic

$$\text{KS} = \max_{1 \le i \le n} |\tau_i - i|;$$

results are given in Table V. We found that the self-limiting Hawkes model has a smaller KS test statistic than the standard Hawkes model in seven of the ten squares tested, meaning that the self-limiting model is statistically significant at a higher confidence level than the standard Hawkes process for these seven squares. However, neither was model statistically significant at the 95% confidence level in any of the squares. One possible reason for this finding is that the excited kernel for real data is not exponentially decaying; this can be observed in [13], where a non-parametric method is used to estimate the kernel g of the Hawkes process, and the results are clearly not exponential decay.

TABLE V
THE KOLMOGOROV–SMIRNOV TEST STATISTICS OF THE RESIDUALS USING BOTH MODELS. SQUARE NUMBERS WRITTEN IN GREEN DESIGNATE SQUARES WHERE THE SELF-LIMITING MODEL OUTPERFORMED THE STANDARD MODEL WHILE NUMBERS IN RED DESIGNATE THE OPPOSITE.

Square	KS (Standard)	KS. (S-L)
1	96.1088	91.6981
2	69.5929	71.0726
3	71.6870	49.8222
4	41.1858	45.4628
5	65.4766	65.3632
6	81.0281	55.3470
7	80.7324	79.6442
8	88.6134	87.9864
9	99.2195	93.5650
10	101.9143	103.3003

TABLE VI
THE AIC VALUES FOR EACH SQUARE USING BOTH MODELS. SQUARE NUMBERS WRITTEN IN GREEN DESIGNATE SQUARES WHERE THE SELF-LIMITING MODEL OUTPERFORMED THE STANDARD MODEL WHILE NUMBERS IN RED DESIGNATE THE OPPOSITE.

Square	Standard Hawkes	Self-Limiting Hawkes	Relative Likelihood
1	13279.8524	13284.0058	0.1253
2	12619.7345	12621.4520	0.4237
3	11987.1279	11967.3027	4.9547×10^{-5}
4	11711.5026	11718.6872	0.0275
5	11518.8463	11522.4911	0.1616
6	11575.4314	11549.3261	2.1444×10^{-6}
7	11454.0408	11452.7481	0.5239
8	11417.2304	11421.1992	0.1375
9	11126.1679	11119.5940	0.0374
10	11025.7570	11031.2991	0.0626

Another way to measure the goodness of fit between models is to compare the Akeike information criterion (AIC) values for the two models. The AIC is defined as

$$\text{AIC} = 2(p - \mathscr{L}),$$

where p is the number of estimated parameters in the model and $\mathscr{L}$ is the log-likelihood value of the estimated parameters. The set of parameters that minimizes the AIC is the more likely model.

For this analysis, we divide the data set up into the same squares, training sets, and testing sets as we used for the residual analysis above. Once again, we estimate the parameters of the training sets using both models. We then compute the log-likelihood and AIC values of each testing set using the parameters estimated using the corresponding training sets. The AIC values are given in Table VI. The third column of Table VI lists the relative likelihoods of the models, defined as

$$\text{relative likelihood} = e^{\frac{\text{AIC}_{\min} - \text{AIC}_{\max}}{2}}.$$

This measure represents how probable the higher AIC model is to minimize the information loss relative to the lower AIC model.

As we can see in Table VI, the self-limiting Hawkes model resulted in lower AIC values in four of the ten squares. Thus, in these four squares, it is more likely that the data follows a

self-limiting Hawkes process rather than a standard Hawkes process. Moreover, in three of these squares, the relative likelihood values indicate that the probability of the standard Hawkes model resulting in a smaller information loss than the self-limiting Hawkes model is effectively zero. Hence, the self-limiting Hawkes process based on an exponentially decaying excited kernel is in some circumstances a better fitting model to our crime data than the standard Hawkes process with the same form of kernel.

VI. Conclusions

In this work, we introduced a self-limiting Hawkes process, a variant of the Hawkes process where the self-exciting component is counteracted by a self-limiting component. In the context of modelling crime data, the self-exiting component represents the likelihood that crime at a point in time will likely lead to more crime in the near future. The self-limiting component represents the efforts of a police force in preventing crime events from happening.

We provide methods for simulating the self-limiting Hawkes process, as well as a method for estimating the parameters of the self-limiting Hawkes process given a dataset of event times. Using maximum likelihood estimation, it has been shown that the parameters of a standard Hawkes process can be estimated with high accuracy [15]. Using a variation of this method that takes into account the preventative action of the self-limiting Hawkes process, we show that one can still estimate the parameters of the underlying Hawkes process with high accuracy. Further, using real crime data, we were able to show that the self-limiting Hawkes process is a plausible alternative to the standard Hawkes process, though neither of the two processes were very likely fits to the data.

Future work in this area could include the incorporation of a spatial component to the self-limiting Hawkes model so that spatial binning of the data for parameter estimation would be unnecessary. It has been shown that a spatial component can be incorporated into a the standard Hawkes model [13], so we believe that a similar approach could be used for the self-limiting Hawkes model, and may enhance its ability to fit real crime data. Another avenue of inquiry along these lines would be testing self-limiting Hawkes models with excited kernels g that are not decaying exponentials, which would also likely enhance the ability of the model to fit real-world data.

Acknowledgements

The authors were supported by NSF grant DMS-1737925.

References

[1] L Adamopoulos. “Cluster models for earthquakes: Regional comparisons”. In: *Journal of the International Association for Mathematical Geology* 8.4 (1976), pp. 463–475.

[2] City of Chicago. *Crimes - 2001 to Present*. 2001. URL: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2 (visited on 08/31/2020).

[3] Angelos Dassios and Hongbiao Zhao. “Exact simulation of Hawkes process with exponentially decaying intensity”. In: *Electronic Communications in Probability* 18.62 (2013).

[4] Eric W Fox et al. “Modeling e-mail networks and inferring leadership using self-exciting point processes”. In: *Journal of the American Statistical Association* 111.514 (2016), pp. 564–584.

[5] Stephen J Hardiman, Nicolas Bercot, and Jean-Philippe Bouchaud. “Critical reflexivity in financial markets: a Hawkes process analysis”. In: *The European Physical Journal B* 86.10 (2013), p. 442.

[6] Alan G Hawkes. “Spectra of some self-exciting and mutually exciting point processes”. In: *Biometrika* 58.1 (1971), pp. 83–90.

[7] Valerie Isham and Mark Westcott. “A self-correcting point process”. In: *Stochastic processes and their applications* 8.3 (1979), pp. 335–347.

[8] Shane D Johnson. “Repeat burglary victimisation: a tale of two theories”. In: *Journal of Experimental Criminology* 4.3 (2008), pp. 215–240.

[9] Shane D Johnson, Kate Bowers, and Alex Hirschfield. “New insights into the spatial and temporal distribution of repeat victimization”. In: *The British Journal of Criminology* 37.2 (1997), pp. 224–241.

[10] Shane D Johnson and Kate J Bowers. “The burglary as clue to the future: The beginnings of prospective hot-spotting”. In: *European Journal of Criminology* 1.2 (2004), pp. 237–255.

[11] PA W Lewis and Gerald S Shedler. “Simulation of nonhomogeneous Poisson processes by thinning”. In: *Naval research logistics quarterly* 26.3 (1979), pp. 403–413.

[12] George O Mohler et al. “Randomized controlled field trials of predictive policing”. In: *Journal of the American statistical association* 110.512 (2015), pp. 1399–1411.

[13] George O Mohler et al. “Self-exciting point process modeling of crime”. In: *Journal of the American Statistical Association* 106.493 (2011), pp. 100–108.

[14] Yosihiko Ogata. “On Lewis’ simulation method for point processes”. In: *IEEE transactions on information theory* 27.1 (1981), pp. 23–31.

[15] 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.