

HAWKES PROCESSES IN FINANCE: A REVIEW WITH
SIMULATIONS

by

GRAHAM SIMON

A THESIS

Presented to the Department of Mathematics
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2016

An Abstract of the Thesis of

Graham Simon for the degree of Bachelor of Science
in the Department of Mathematics to be taken June 2016

Title: Hawkes Processes in Finance: A Review with Simulations

Approved: _____

Christopher D. Sinclair

Abstract

Hawkes processes are flexible robust models for simulating many of the numerous self-exciting point processes seen in empirical data. Using a Hawkes process creates clusters in modeled data that are frequently seen in different natural environments. Some frequent areas of use for Hawkes processes include the study of earthquakes, neural networks, social media sharing, and financial trading data. This work builds an accessible framework for the undergraduate study of Hawkes processes through building step-by-step from point processes to Poisson processes and eventually Hawkes models. A literature review of current research and utilizations for Hawkes processes is then done to demonstrate some of the dramatic growth seen in this field of research. Point clusters, kernel estimation, parameter estimation, and algorithms for implementation are also discussed with simple simulations performed in Excel.

Acknowledgements

First, I would like to thank Professor Chris Sinclair for all his help and guidance during the thesis process. Not only is Chris a wonderful guide in the labyrinth of mathematics, but he also provided the perspective, wisdom, and encouragement to make this research process a truly multifaceted learning experience. Second, I would like to thank Hayden Harker, Eric Merchant, and Aaron Montgomery for their wonderful instruction and advice during my time at the University of Oregon. Finally, I would like to thank my grandparents, parents, and wife for their unwavering love and support in this and all my pursuits.

Table of Contents

Introduction	1
Undergraduate Research of Hawkes Processes	4
Theoretical Background	7
Point Processes	7
Point Process Models	8
Overview of Hawkes Processes	14
Hawkes Process	14
Initial Kernel Development & Relationship with Autoregressive Models	16
Different Kernels	18
Beginning Analysis of Point Clusters	20
Data Simulation	22
Estimation of the Kernel	23
Literature Review	27
Other Interesting Areas of Self-Exciting Point Process Research	30
Algorithms for Simulation and Estimation	32
Poisson Processes	32
Univariate Hawkes Process with Exponentially Decaying Intensity	36
Multivariate Hawkes Process with Exponentially Decaying Intensity	42
Estimation of a Power Kernel	44
Implemented Processes	46
In an Excel workbook each of these algorithms, except the multivariate Hawkes process were simulated. Some of the simulated results are:	46
Future Work	49
Improve Simulation Implementations	49
Bibliography	50

List of Figures

Figure 1: Graph of Exponential Distribution with $\lambda = 5$	9
Figure 2: Graph of Exponential Density Function with $\lambda = 5$	9
Figure 3: Graph of Poisson distribution with $\lambda = 5$	10
Figure 4: Graph of Poisson Density Function with $\lambda = 5$	11
Figure 5: Kernel estimate showing individual kernel. Window width of 0.4 (Silverman, 1986).	26
Figure 6: Graph of $-\frac{1}{\lambda} \ln U$ for $\lambda = 5$	33
Figure 7: Example of a Generic Exponential Intensity Decay	38
Figure 8: Homogenous Poisson Process with $\lambda = 4$ and $T = 50$	46
Figure 9: Non-homogenous Poisson Process: $\lambda = 7.5$, $T = 7.5$, & $\lambda(t) = \frac{n}{\tau} \exp\left(-\frac{t}{\tau}\right)$	46
Figure 10: Intensity of Univariate Hawkes Process: $a = 0.45$, $\delta = 0.25$, $\lambda_0 = 0.9$, & Intensity Jumps of Size 0.40	47
Figure 11: Histogram of Number of Events per Unit Time for Univariate Hawkes Process: $a = 0.45$, $\delta = 0.25$, $\lambda_0 = 0.9$, & Intensity Jumps of Size 0.40	47

Introduction

Properly functioning capital markets (e.g. stock, bond, and real estate markets) not only provide efficient capital allocation, but also the necessary gains for the funding of pension funds, retirement accounts, and university endowments. Ideally these important institutions undertake their fiduciary duty free from speculation and undertake an *investment* operation “which upon thorough analysis, promises safety of principal and a satisfactory return.” By undertaking this process investment managers seek not only objectively sound returns, but also risk adjusted returns above the average market participant. However, the existence of such a process has been one of the most debated financial subjects of the last century. Most academics have now concluded sustained outperformance is impossible and profess belief in the *efficient market hypothesis*. Under the efficient market hypothesis all public and nonpublic information is continuously priced into financial assets so any performance above the average must be short term luck. Thus after a period of outperformance a fund will likely enter a period of underperformance pushing the individual average down to, or below (after fees), the larger industry average.

Against this academic backdrop, speculators have spent centuries attempting to profit from short term fluctuations in the value of stocks, bonds, commodities, and other investment vehicles. George Soros offered an explanation for this phenomenon in his book *The Alchemy of Finance* through the idea of *market reflexivity*. Market reflexivity presents the idea that financial markets combine an interplay of fundamental exogenous events (breaking news, new financial results being release, and other economic fundamentals) with market created endogenous events (price changes from previous

trades or other purely market based forces). Reflexivity carries especially high weight today since different historical bubbles have spawned legions of day traders and other speculators, but modern trading occurs primarily through computers. Computerized trading allows for much larger blocks of stocks or bonds to be rapidly moved for minute gains unachievable by slower human traders. This means rigorous analysis may now examine trading data in time intervals as small as a microsecond, and focus upon the system itself as much as the economic fundamentals affecting the financial asset.

Unlike stocks and bonds, which are tied to the future cash flow generations of companies or governments, commodities (e.g. oil, natural gas, corn, and cocoa) are exchange traded assets with no value beyond their eventual use. In commodity financial markets future contracts are bought and sold that give the owner the right to buy or sell a certain commodity at a predetermined price sometime in the future. For example one of the most common oil trades is the price currently agreed to be paid upon delivery of a barrel of West Texas Intermediate (WTI) oil in Cushing, Oklahoma one month in the future. Companies use commodity contracts to help set and plan for the prices received and paid for their good and inputs in the future. However, companies do not act alone in this arena as speculators thrive, and die, guessing upon potential changes in commodity prices. Overall, based on the uniform character of commodities these assets tend to experience greater endogenous effects than stocks and bonds and trade in ways heavily influenced by the market itself.

When building an initial framework for understanding short term trading in this area, prior econometricians realized they needed a model with a “time series of irregularly spaced points that show a clustering behavior” (Fonseca and Zaatour, 2016).

During the financial trading day the clustering of orders appears most obviously at the beginning and ending of the day, but orders between these heavy times also tend to clump together. Alan Hawkes first popularized a quickly understood process fulfilling this criteria in 1971. Now known as a *Hawkes process* this model created a self-exciting process (i.e. one event increases the probability another event will follow shortly) with exponential (rapid) time decay mimicking the clustering of neuron firing, earthquakes, and financial trading data. As high frequency trading began representing a larger and larger portion of total trading volume during the 1990's and 2000's the study of Hawkes processes also became a more active area of research.

This paper will first review the mathematical creation and improvement of Hawkes models before exploring its financial capabilities and end with simulations of different point processes. This thesis is designed to serve as an accessible work for undergraduate students seeking to learn about Hawkes processes as the model operates in a flexible manner with application in many different fields. Unfortunately while this model seems well taught and utilized at the graduate level its recent development means little writing exists at an undergraduate level. With time this will most likely change due to the ever expanding amount of available electronic data in modern society and the need for rapid and precise responses to high intensity events. Finally, utilizing a Hawkes model upon commodity ETF data will both serve to give another data point in the study of commodities and be one of the first to view the effects of clustered trading upon ETFs which are far more accessible to retail (Ma & Pa) investors.

Undergraduate Research of Hawkes Processes

This thesis is designed to be accessible to undergraduate students who have previously taken courses in probability or statistics and wish to learn about the unique properties of Hawkes processes. However, a couple notes should be made about studying these processes and the approach undertaken within this thesis. First, attempting to implement Hawkes processes, like most advanced models, requires a fair amount of coding. In order to fully research these processes it is therefore necessary to spend a lot of time gaining familiarity with how to implement probability code in addition to learning the rigorous underpinnings of the Hawkes process. Possessing both these skills then allows researcher the opportunity to quickly adapt or tune a model to best fit empirical data. Given the range of skills necessary for this it is therefore no surprise that a couple quantitative finance research departments produce an outsized proportion of current research on this topic.

This thesis focuses on the knowledge necessary to understand and discuss Hawkes processes, but utilizes more simplistic models for simulations. Additionally, while there are certainly many interesting avenues for applied research with Hawkes processes, not enough progression was made to construct models from empirical data. Not reaching this point was certainly a disappointment, but it also serves as a highly instructive learning moment on the difficulties of research. Unfortunately progression does not occur linearly, and gauging the full scope of these projects cannot be seen until they are well underway. Eventually research for this thesis converged upon understanding Hawkes process, their properties and attributes, and a couple algorithmic ways of simulating point processes.

One of the biggest areas for expansion of this thesis, if research was to be continued, would be to develop a better understanding of maximum likelihood estimation and other forms of parameter estimation. Since methods like maximum likelihood estimation provide the numerical values for empirical modeling a more robust understanding of these techniques would be paramount to progressing further in the subject matter. Greater time spent coding and building a firm understanding of statistical software would also be critical to adding new results. Assuming this knowledge was acquired, accumulating the necessary financial data to create empirical models would be the next difficult step. There are a couple different financial services firms that maintain data detailed enough for the full benefit of modeling with Hawkes processes, but access to these data sets costs in excess of \$20,000 and cannot be replicated from free sources. Methods, like a Brownian bridge, exist to distribute accumulated discrete points across a range in order to create smaller breaks in the data set, but these then alter the underlying structure of the data. High costs associated with financial data therefore offer another reason why authors from the same departments and centers possess the best data available for modeling Hawkes processes.

Finally, research in this area seems to have done a good job of better describing some attributes of financial markets, but even the best models fail to present a compelling model for profits. While it is quite possible that researchers with profitable ideas quietly implemented them, and reaped profits themselves, the available literature focuses more on the complexity of modern financial markets. In particular researchers demonstrate that the rapidity of change within financial markets and how little short term financial trading resembles the trading world of only a decade ago. Thus,

progression from new research requires simultaneous progress in both the applied and theoretical aspects. New theoretical ideas must be created, implemented, tested against empirical data, and then revised in order to make meaningful improvement. Attacking these problems is therefore quite difficult at an undergraduate level, but the Hawkes process is flexible enough that benefits from using it may eventually be found in a multitude of yet unknown areas.

Theoretical Background

Point Processes

In order to analyze the effects of many scientific and financial processes it is necessary to begin counting the frequency of events over time. Mathematically, keeping track of when random events occur during a known time window is a *point process*. Each point then represents a “time and/or location of an event, such as a lightning strike,” earthquake, or stock trade (Schoenberg, 2016). In general a point process, N , is defined as a random increasing step function on a “metric space S taking values in the non-negative integers” (Schoenberg, 2016). This simply means that N is a function which represents an integer count between 0 and infinity (inclusively) for the number of points filling in a subset A of S . Even more simply N is just a function counting the number of events during any time window. While an infinite number of points may appear in a subset, most point processes building upon real world data remain focused upon areas where N may contain “only finitely many points on any bounded subset of S ” (Schoenberg, 2016). Focusing on situations with a finite number of events then allows for applied analysis to reach meaningful conclusions.

Restricting this general definition to the needs of this paper we can consider a *temporal point process* (all events occur between times 0 and T). For a temporal point process N is then simply an ordered list $\{t_1, t_2, \dots, t_n\}$ of event times. Alternatively the list may be thought of as inter-event times $u_i = t_i - t_{i-1}$ and provide the list of gaps between events $\{u_1, u_2, \dots, u_n\}$, taking $t_0 = 0$. In order to get the specific count of events at any positive time $t < T$ we may then use the notation $N(t)$ to reference the number of

points occurring at or before time t . The process $N(t)$ will then be non-decreasing (since an event cannot unhappen), right-continuous, take only non-negative integer values (there cannot be negative events), and have left jump discontinuities at each event time t_j (when a new event pushes the function value up one). Thus, a temporal point process could alternatively be defined “as any non-decreasing, right-continuous \mathbf{Z}^+ -valued process” (Schoenberg, 2016). Defining point processes this way then fulfills all the criteria necessary and immediately denotes the flexibility of allowing any such function. A point process is called *simple* if, with probability one, all its points t_i occur at distinct times and *orderly* if for any time t , $\lim_{\Delta t \rightarrow 0} \frac{P\{N(t, t + \Delta t] > 1\}}{\Delta t} = 0$.

Beyond simply time values a point process can also contain additional variables to make it a *marked point process* (i.e. a function with multiple input variables). A financial example of a marked point process would be a set which not only contains the times of different market trades, but also the size of the trade, if it moved the quoted price, or who made the trade. Marked point processes are also then very similar to time series data. In principle the different realizations of a marked point process could be viewed as the dataset of a time series and vice versa. However, a time series dataset and a marked point process differ by allowing an event to take any time in a continuum, whereas as time intervals are deterministic for time series data.

Point Process Models

Before defining the most common point process models recall that a random variable T is said to have an exponential distribution with rate $\lambda > 0$, or $T = \text{exponential}(\lambda)$, if:

$$P(T \leq t) = 1 - e^{-\lambda t} \quad \text{for all } t \geq 0.$$

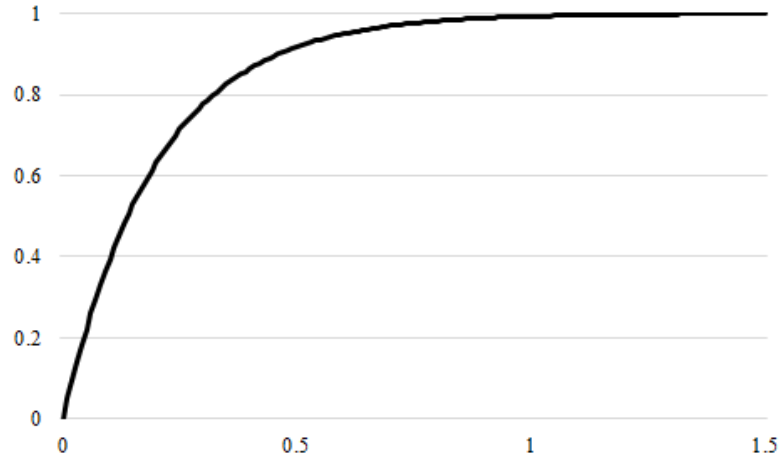


Figure 1: Graph of Exponential Distribution with $\lambda = 5$

This will then produce a density function $f_T(t)$ equal to:

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0. \end{cases}$$

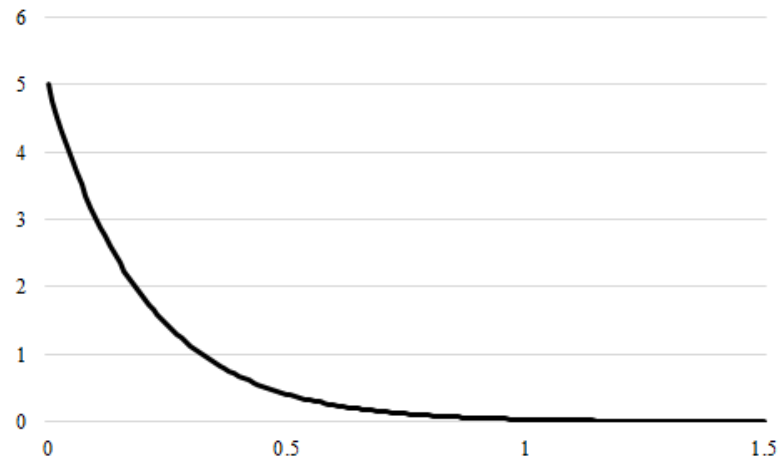


Figure 2: Graph of Exponential Density Function with $\lambda = 5$

Using these definitions we may then define one of the most utilized point process models, and the most important one for this paper, the *Poisson process* (Durrett, 1999).

Definition: Let τ_1, τ_2, \dots be independent exponential(λ) random variables. Let $T_n = \tau_1 + \tau_2 + \dots + \tau_n$ for $n \geq 1$, $T_0 = 0$, and define $N(s) = \max\{n : T_n \leq s\}$.

Homogenous Poisson processes (i.e. those with a constant value for λ) $N(s)$ are then distributed with mean λs and variance λs .

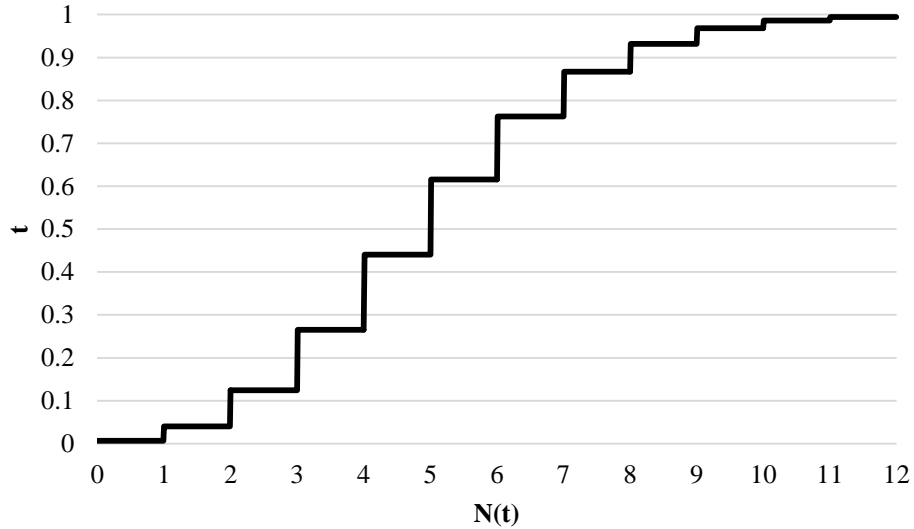


Figure 3: Graph of Poisson distribution with $\lambda = 5$

One of the nice attributes of Poisson processes are that they have *independent increments*. This means that a Poisson process $N(t)$ is such that if $t_0 < t_1 < \dots < t_n$, then $N(t_1) - N(t_0)$, $N(t_2) - N(t_1)$, ..., $N(t_n) - N(t_{n-1})$ are independent. Using this language a Poisson process may also be defined as (Durrett, 1999):

Theorem: If $\{N(s), s \geq 0\}$ is a Poisson process then:

- (i) $N(0) = 0$
 - (ii) $N(t + s) - N(s) = \text{Poisson}(\lambda t)$
 - (iii) $N(t)$ has independent increments
- Conversely, if (i), (ii), and (iii) hold, then $\{N(s), s \geq 0\}$ is a Poisson process.

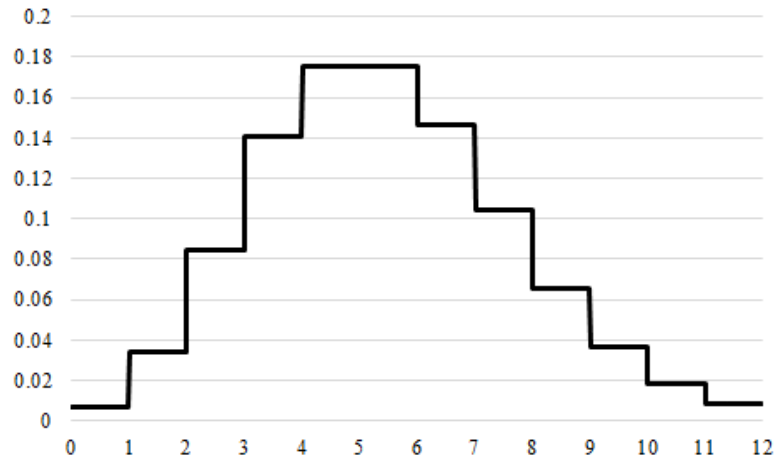


Figure 4: Graph of Poisson Density Function with $\lambda = 5$

However, these theoretical models have a flaw in that most financial events do not occur with uniform probability. Rather, the occurrence of an event tends to either increase or decrease the probability of another event happening in response to the initial one. Therefore a modified Poisson process with a varying rate for λ can be much more accurate, and is called a *Non-homogenous Poisson process*. Such a process is defined as:

- $\{N(s), s \geq 0\}$ is a Poisson process with rate $\lambda(r)$ if:
- (i) $N(0) = 0$
 - (ii) $N(t)$ has independent increments
 - (iii) $N(t) - N(s)$ is Poisson with mean $\int_s^t \lambda(r) dr$.

The function $\lambda(t)$, represents the infinitesimal rate at which events are expected to occur around a particular time t . One way of thinking of this would be to consider it a hurdle rate for a new randomly generated value to have to clear in order for a point there to be included (this will actually be part of the method of modeling non-homogenous processes later). This rate is known as the *conditional intensity* of the non-homogenous Poisson process and is based on the prior history of the point process.

Estimating $\lambda(t)$ can be done both parametrically and nonparametrically (i.e. with external constraints or without external constraints), to allow $\lambda(t)$ to vary properly with t . For a temporal point process originating at time 0 the *compensator*, $A(t)$, may be defined as the integral of the conditional intensity from time 0 to time t . An equivalent definition would be the compensator is the unique non-negative non-decreasing predictable process $A(t)$ such that $N[0,t] - A(t)$ is a martingale (Note: $N[s,t] = N(t) - N(s)$).

A *renewal process* is a point process where the inter-event times $\{u_1, u_2, \dots, u_n\}$ are independent random variables. Density functions governing each inter-event time are thus known as *renewal density* functions. Such models describe situations in which the probability of an event occurring depends only on the time since the most recent event (e.g. in fire hazard analysis such a model is consistent with wood fuel loading followed by complete fuel depletion in the event of a fire). Overall these characteristics provide the ability to generate a point process with varying degrees of intensity based on the history of events up to any time t . However, these processes can still be improved further by adding factors that create the clustering behavior seen in many market situations.

Setting the stage for Hawkes processes comes the idea of *self-exciting* and *self-correcting* point processes. A point process N is self-exciting if $cov\{N(s,t), N(t, u)\} > 0$ for $s < t < u$ and is self-correcting if the covariance is negative. More specifically this means “the occurrence of points in a self-exciting point process causes other points to be more likely to occur, whereas in a self-correcting process, the points have an inhibitory effect” (most simply self-exciting processes clump together in bunches

because one event increases the chances of another happening while self-correction processes spread out event times through the opposite process) (Schoenberg, 2016). Connecting this to the previous discussion of homogenous and non-homogenous Poisson processes this means self-exciting and self-correcting processes are noticeably affected by previous events in the same way a non-homogenous Poisson process has varying intensity with time. While a Poisson process is neither self-exciting nor self-correcting by definition, $\lambda(t)$ may be modified to produce results similar to a self-exciting or self-correcting process. Utilization of self-exciting processes emerged from clustering in the point process enabled by these models.

Overview of Hawkes Processes

Hawkes Process

A Hawkes process is a point process with a response function (or kernel) $\phi(t - t_i)$ which reflects the influence of past events on the conditional intensity. The beauty of this model is that it is more general than the Poisson process and has greater potential to explain some of the phenomena seen in financial markets. Specifically, Daley and Vere-Jones claim the Hawkes process, “comes closest to fulfilling, for point processes, the kind of role that the autoregressive model plays for conventional time series” (Daley, 2003). Given a counting process $N(t) = \max\{i : t_i \leq t\}$ and filtration $F_t = \{t_1, \dots, t_i : \forall i < N(t)\}$, representing the information about the process up to time t , a linear continuous Hawkes process may be defined as a point process $\{t_i\}_{i \in \mathbb{Z}^+}$ with conditional intensity given by:

$$\lambda(t | F_{t-}) = \mu(t) + \int_{-\infty}^t \phi(t - s) dN(s).$$

The conditional intensity is defined as:

$$\lambda(t | F_{t-}) = \lim_{h \rightarrow 0} \frac{E[N(t+h) - N(t) | F_{t-}]}{h}.$$

Intuitively “the conditional intensity is an infinitesimal expected rate at which the events occur around time t given the history of the process N before t ” (Morzywolek, 2015).

Within this model the two most important moving pieces are then $\mu(t)$ and the kernel function ϕ . Generally $\mu(t)$ is seen as the background intensity responsible for accounting for the arrival of exogenous (external) events while the kernel function ϕ ,

satisfying causality condition $\phi(t) = 0$ for $t < 0$, determines the correlation properties of the process. (Note: Throughout the rest of this paper exogenous events will frequently be referred to as immigrant events since these events were caused by external factors, and events generated by ϕ will be referred to as descendants or children since these are system created events “descended” from the initial immigrant.) Then the *branching ratio*, n , for the process may be defined as

$$n = \int_0^{\infty} \phi(t) dt.$$

The differential of the counting process may then be rewritten as a sum of delta functions

$$dN(t) = \sum_{t_i < t} \delta(t - t_i).$$

To form the discrete time form of the conditional intensity

$$\lambda(t | F_{t-}) = \mu(t) + n \sum_{t_i < t} h(t - t_i)$$

where $h(t) = \phi(t)/n$ is called a *bare kernel*. Note that the bare kernel is a probability density function and the Hawkes process is *stationary* for $n < 1$. Stationarity implies if X_1 and X_2 are independent copies of a random variable with $a, b > 0$ being constants then $aX_1 + bX_2$ has the same distribution as $cX + d$ for some positive constants c, d .

Assuming stationarity and taking $\mu(t)$ to be constant, which will simply be denoted μ from now on, the average total intensity Λ may be calculated as

$$\Lambda = E[\lambda(t | F_{t-})] = E[\mu(t) + \int_{-\infty}^t \phi(t - s) dN(s)] = \mu + \Lambda \int_0^{\infty} \phi(\tau) d(\tau)$$

which implies

$$\Lambda = \frac{\mu}{1 - n}.$$

Note that if $n > 1$ this formula implies that $\Lambda \rightarrow \infty$ exponentially quickly, and hence the counting process $N(t)$ eventually explodes. Hence understanding the branching ratio is critical to properly analyzing a Hawkes process. Essentially the branching ratio represents the average number of first-generation daughters (market created events) of a single mother (actual company or economic event). If $n = 0$ then the model collapses back down to a non-homogenous Poisson process since there will be no further events triggered by the initial immigrants. Therefore, the Hawkes process may be viewed as a generalization of the Poisson process that depends on both the time and history of a process. Further, the critical case $n = 1$ separates the model into subcritical ($n < 1$) and supercritical ($n > 1$) states. If $n > 1$ for a sustained amount of time the modeled intensity may explode beyond applied analysis so most study focuses upon subcritical cases. Finally, “whenever the intensity μ is a constant and the process is in the subcritical ($n < 1$) or in the critical ($n = 1$) regime the branching ratio can be used as a measure of the proportion of events that are generated inside the model (by the presence of the exponential kernel, i.e. endogenously generated events) to all events” (Lorenzen, 2012).

Initial Kernel Development & Relationship with Autoregressive Models

Early development of the Hawkes process focused on the exponential kernel $\phi(t - t_i) = \alpha e^{-\beta(t-t_i)}$ which leads to the conditional intensity

$$\lambda_t(t) = \mu(t) + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}.$$

To begin seeing how this process resembles an autoregressive model consider the intensity at some past specified time t_i . Then the intensity will be

$$\lambda(t_i) - \mu(t_i) = \sum_{t_k < t_i} \alpha e^{-\beta(t_i - t_k)}$$

where t_k represents all events that occurred before t_i . Next if we multiply both sides of the previous equation by $\exp[-\beta(t - t_i)]$ we have

$$[\lambda(t_i) - \mu(t_i)]e^{-\beta(t - t_i)} = \sum_{t_k < t_i} \alpha e^{-\beta(t - t_k)}.$$

Now the response function can be decomposed into

$$\sum_{t_i < t} \alpha e^{-\beta(t - t_i)} = \sum_{t_k < t_i} \alpha e^{-\beta(t - t_k)} + \sum_{t_k > t_i < t} \alpha e^{-\beta(t - t_k)}.$$

Combining the last two equations we can write $\lambda(t) - \mu(t)$ as

$$\lambda(t) - \mu(t) = [\lambda(t_i) - \mu(t_i)]e^{-\beta(t - t_i)} + \sum_{t_k > t_i} \alpha e^{-\beta(t - t_k)}.$$

This then deeply resembles the continuous time form of an autoregressive model

$$X_t - \mu = e^{-\beta(t - s)} (X_s - \mu) + \text{sum of innovations}$$

“where the term $[\lambda(t_i) - \mu(t_i)]e^{-\beta(t - t_i)}$ is the autoregressive term and the term

$\sum_{t_k > t_i} \alpha e^{-\beta(t - t_k)}$ represents the sum of the innovations in the AR process” (Lorenzen, 2012).

For the exponential kernel the unconditional intensity Λ for a trading day may be calculated as

$$E(\lambda) = E(\mu) + E \left[\int_{-\infty}^t \alpha e^{-\beta(t-s)} dN(s) \right].$$

Then assuming stationarity as above we get the expected intensity as

$$E(\lambda) = \frac{\mu}{1 - \frac{\alpha}{\beta}}$$

Hence for the exponential kernel the characteristics and behavior of the point process are determined by the ratio of $\frac{\alpha}{\beta}$ since

$$n = \int_0^{\infty} \alpha e^{-\beta t} dt = \frac{\alpha}{\beta}.$$

Different Kernels

In many applied settings simply observing the distribution of events provides enough evidence to decide what “the kernel should look like or what properties it should have” (Morzywolek, 2015). However, there are still a couple different options for the kernel in a Hawkes process. First, the power law kernel is defined as

$$\phi_{pow}(t) = \frac{n\theta c^\theta}{(t+c)^{1+\theta}} \chi(t)$$

with $\chi(t)$ representing the unit step function (i.e $\chi(t)$ is 0 if $t < 0$ and 1 if $t \geq 0$). The unit step function guarantees the causality and is often used in geophysical applications.

Second, the exponential kernel can be written in a slightly modified form from $\phi(t - t_i) = \alpha e^{-\beta(t-t_i)}$ and be defined as

$$\phi_{exp}(t) = \frac{n}{\tau} \exp\left(-\frac{t}{\tau}\right) \chi(t).$$

Exponential kernels provide a much faster decay in the probability distribution than power kernels and are generally used in short-memory processes. Finally, a couple modified kernels such as the cut-off kernel (Kagan and Knopoff, 1981)

$$\phi_{cut}(t) = \frac{n\epsilon\tau_0}{t^{1+\epsilon}} \chi(t - \tau_0)$$

and the double exponential kernel (Rambaldi, Pennesi, and Lillo, 2014).

$$\phi_{de}(t) = \left(\alpha_1 \exp\left(-\frac{t}{\tau_1}\right) + \alpha_2 \exp\left(-\frac{t}{\tau_2}\right) \right) \chi(t)$$

have been suggested to help bridge the gap between the two most common kernels.

Specifically for financial processes there continues to be debate about which kernel best represents the dependencies of price changes in financial assets, but the exponential kernel is more frequently utilized. Despite this frequency the power kernel and its longer term price correlation are still advocated by many (Hardiman, Bercot, and Bouchaud, 2013) (Bacry, Dayri, and Muzy, 2012). This contrasts with the general consensus which maintains previous price movements have limited impact and therefore exponential kernels make more sense (Filimonov and Sornette, 2012) (Filimonov and Sornette, 2015). Arguments for the exponential kernel also include the Markov property frequently assumed for financial assets. Financial models often assume the Markov property, that previous trading data does not affect the future probabilities for an asset, which lends credence to the short-memory process and kernel. Finally, the nature of the financial data being studied also affects the selected kernel. Data sets with frequent sampling, such as high frequency trading, tend to use exponential kernels more often than longer time horizon studies that use power kernels.

This is because things like stock market trades depend far less on long ago trades than corporate bond default rate trends that can take months or years to play out.

Beginning Analysis of Point Clusters

As alluded to above, the beauty of a Hawkes point process model is the point clusters generated by these models. Clusters emerge because an initial immigrant event, from an external news or economic event, causes the probability of a subsequent event occurring to increase. This increase in the probability of another event occurring makes the Hawkes process self-exciting and capable of modeling the clustering of data often seen in empirical data. While this might not initially seem like a groundbreaking discovery it allows for the creation of models much more in tune with reality than many deterministic models. In everything from the dynamics of views of YouTube views (MacKinlay, 2015) and Twitter retweets (Zhao, Erdogdu, He, Rajaraman, and Leskovec, 2015) to analyzing earthquake data a wide variety of point processes occur in clusters. Financial trading data also clusters around big news events, certain times of day, and other catalysts.

Looking beyond the surface level clustering of a point process one of the next questions is how overlapping or separated individual clusters are. In order to view the separation, or lack thereof, in clusters a couple quick definitions are necessary. First, a *renormalized kernel* describes the responses of points within a cluster to the initial immigrant point. This is in contrast with the previously discussed bare kernels which “is nothing else than the probability of a mother event triggering a daughter event” (Morzywolek, 2015). Hence using a renormalized kernel it is possible to talk about the overlapping nature of clusters within the point process.

Bare kernels, $h(t)$, and renormalized kernels, $R(t)$, are related by the equation

$$h(t) = R(t) - n \int_0^t h(t-s) R(s) ds.$$

For the exponential kernel the renormalized kernel is

$$R(t) = \frac{1}{\tau} \exp\left(-\frac{t(1-n)}{\tau}\right).$$

This example shows that the renormalized kernel is not a probability density function (PDF), but in general

$$\int_0^{\infty} R(t) dt = \frac{1}{1-n}.$$

Using this information it is then possible to calculate the average distance between immigrants and the average length of a cluster. Assuming as above that immigrants are generated from a homogenous Poisson process with intensity μ the average distance between points is then $\frac{1}{\mu}$, but the occurrence of descendants is defined by a renormalized kernel like $\frac{1}{\tau} \exp\left(-\frac{t(1-n)}{\tau}\right)$.

This means the average length of the cluster is

$$\int_0^{\infty} t(1-n) R(t) dt = \int_0^{\infty} \frac{(1-n)}{\tau} t \exp\left(-\frac{(1-n)}{\tau} t\right) dt = \frac{\tau}{1-n}$$

where the $(1-n)$ factor accounts for the renormalized kernel not being a PDF. With this information κ will be defined as the ratio of the average length of the cluster and the average distance between immigrants. If $\kappa < 1$ it follows that average cluster lengths are shorter than the average distance between immigrants and therefore the point clusters are well-separated. However, if $\kappa \geq 1$ then cluster lengths are much longer than the average distance between immigrants and most clusters overlap. Finishing the example

of a renormalized exponential kernel the value of κ for this renormalized kernel is simply

$$\kappa = \frac{\tau/(1-n)}{1/\mu} = \frac{\mu\tau}{1-n}.$$

Data Simulation

Currently there are two common ways to generate a simulated Hawkes process. The first, the *thinning method*, was initially proposed by Lewis and Shedler in 1978 and was used to simulate inhomogeneous Poisson processes. Ogata then applied the thinning method to the Hawkes process in 1981 and instituted it as one of the most common ways of simulating a Hawkes process. A thinning process works by first generating data points t_1, \dots, t_N from a homogeneous Poisson process with intensity λ_{maj} being majorant to the conditional intensity $\lambda(t)$ of the Hawkes process being generated, i.e. $\lambda_{maj} \geq \lambda(t), \forall t$. The majorant intensity then acts as a filter for the acceptance-rejection method. Using randomly generated values for a given point t_i points are then accepted with probability p_i given by

$$p_i = \frac{\lambda(t_i)}{\lambda_{maj}}$$

and otherwise removed from the sample.

Unfortunately there are two drawbacks to using the thinning method of Hawkes process generation. First, this process runs in $O(N^2)$ time which makes it rather inefficient when dealing with large near-critical Hawkes processes. Second, using this method also ignores the branching structure of the process and generates events created by both exogenous and endogenous factors simultaneously. Simulating events in this

manner makes it impossible to tell whether event is an exogenous immigrant or an endogenous child of a previous immigrant event. In order to deal with these shortcomings, more modern studies (Møller and Rasmussen, 2005) (Møller and Rasmussen, 2006) have suggested simulating all clusters in parallel generation by generation.

Parallel generation begins by simulating all the immigrant events from a homogeneous Poisson process with intensity $\lambda(t)$ equal to the background intensity of the Hawkes process μ . Next the first generation of descendants, which begins the cluster, is modeled for each immigrant event t_i . This is done by utilizing a nonhomogeneous Poisson process with the intensity $\lambda_i(t) = \phi(t - t_i)$. Similarly once the k^{th} generation has been constructed the points of the $(k+1)^{\text{th}}$ generation are produced by a nonhomogeneous Poisson process with intensity $\lambda_{k,i}(t) = \phi(t - t_{k,i})$ (where $t_{k,i}$ is the i^{th} point in the k^{th} generation of a cluster). Finally, this process is then repeated in parallel until there are no more offspring. Utilizing this process then allows the numerical complexity to decline to $O(\mu TK)$ where T is the size of the simulation window and K is the number of generations modeled. Additionally, this process allows for the reconstruction of the entire branching structure in order to determine the attributes of point clusters.

Estimation of the Kernel

Estimation of the kernel for a Hawkes process is of critical importance due to the large impact it has upon the shape and characteristics of a model. Since different kernels have their intensity decay at different rates, and different parameters will enhance or diminish these differences, it's important to estimate an accurate kernel.

Once a kernel has been estimated it is then used as the determining factor for the intensity of the Hawkes process at different times and therefore drives the attributes of the model. The values for a kernel also determine whether modeled point clusters are disjoint or overlapping. When estimating point clusters that are disjoint it might be possible to simply eye ball a model with reasonable accuracy, but if knowing the generations of an initial immigrant or the independence of clusters is important to a research they must estimate a kernel that properly reflects the underlying system. Proper parameter estimation also ensures that useful measurements like the average total intensity are branching ratio are calculated accurately. Therefore, without the proper estimation of a kernel it is difficult to model characteristics that are in real conformity with the studied point process.

When seeking to numerically estimate the parameters for a kernel the most popular method is Maximum Likelihood Estimation (Ogata, 1998). For the Hawkes process the log-likelihood function associated with it is

$$\log L(t_1, \dots, t_N) = \sum_{i=1}^N \log \lambda(t_i | F_{t_i-}) - \int_0^T \lambda(t | F_{t-}) dt$$

where $t_1, \dots, t_N \in (0, T]$ are the observed events. Using this function the parameters of the Hawkes model can then be calculated numerically by maximizing the log-likelihood function in the case of both the exponential and power law kernel. One of the more easily understood methods of estimating these values comes from *Kernel Density Estimation*. This process attempts to use time sections of varying size (described by their *window width*) to construct a smooth approximation of a kernel of a Hawkes process.

First, a potential kernel K must satisfy the condition that

$$\int_{-\infty}^{\infty} K(x) dx = 1.$$

Most of the time K will be a symmetric probability density function, but this does not always hold. Two of the most widely used kernels are the boxcar kernel

$$K(x) = \frac{1}{2} 1_{[-1,1]}(x)$$

and the Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Next, the *kernel estimator*, \hat{f} , formed from l real observations X_1, \dots, X_l with kernel K will be

$$\hat{f}(x) = \frac{1}{hl} \sum_{i=1}^l K\left(\frac{x - X_i}{h}\right)$$

where h is the window width, length of a sub-sections, within the total time span. The kernel function K then determines the shape of event “bumps” while the window width h determines their width. Figure 1 demonstrates an example of this where individual bumps, $l^{-1}h^{-1}K\{(x-X_i)/h\}$, are shown as well as the estimate of \hat{f} given by adding them up.

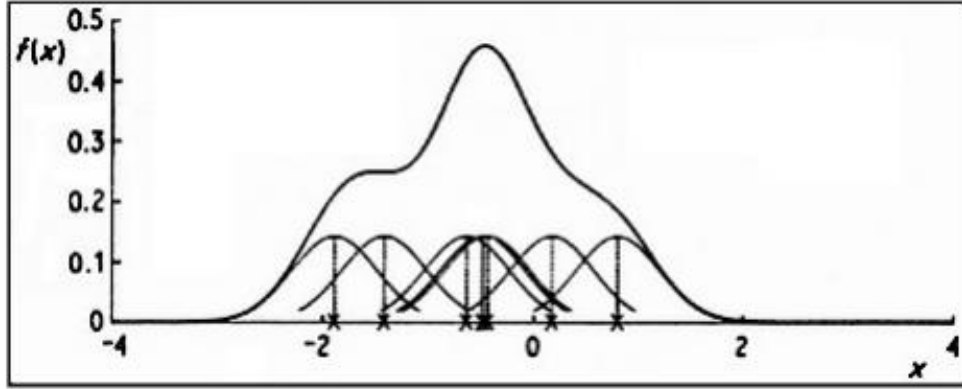


Figure 5: Kernel estimate showing individual kernel. Window width of 0.4 (Silverman, 1986).

Finally, the approximation of the kernel of the Hawkes process is of the form

$$\hat{\phi} = \frac{1}{2\delta T} \sum_{i=1}^N \sum_{j=1}^N K\left(\frac{t - |t_i - t_j|}{h}\right).$$

This method of kernel approximation provides for a couple nice properties.

First, if the kernel is everywhere non-negative and satisfies (1) then the kernel is a probability density function and therefore \hat{f} will also be a probability density.

Additionally, \hat{f} inherits all the continuity and differentiability properties of the underlying kernel. One example of this would be the fact that the Gaussian kernel immediately implies \hat{f} will be a smooth curve with derivatives of all orders. However, one drawback to this method can be seen when estimating kernels with long-tailed distributions. Since the window width is fixed across the entire sample, spurious noise in the tails can cause spikes within \hat{f} . Rectifying this flaw then requires the estimate being smoothed further, but this then begins to dilute the accuracy of the main bulge of the distribution, i.e. the main spike will be decreased at the expense of minimizing the spike in the tail of the distribution.

Literature Review

Beyond the works referenced in the earlier introduction and theoretical overview a few more deserve especial attention. First, while the Hawkes model was initially proposed in 1970's it took almost 30 years for this point process to gain traction in the financial community. Bowsher (2007) and Hewlett (2006) were two of the first to utilize Hawkes models when they studied mid-quote (price between the bid and ask prices) changes and order flow in the FX (foreign currency exchange) market respectively. Generalized Hawkes processes, described in terms of vector conditional intensities, are utilized by Bowsher to demonstrate a two-way interaction between trades and changes in mid-prices within General Motors Corporation (GM) stock during a 40 day study window from 2000. The first direction of influence is that the occurrence of a trade increases the intensity of mid-price changes, and the logical second influence is that mid-price changes do in fact increase trade intensity (i.e. trading volume and mid-price movements are positively correlated).

At the time available databases used timestamps with only one second of precision. Combined with the budding growth of high frequency trading this meant several orders were often combined into one second despite happening at materially different times. In order to combat this issue Bowsher set a precedent by adding a uniform random component that distinguished between equal timestamps (most current published studies now have access to data sets with precision greater than one second). For the GM data set studied only 0.26% of all trades shared a timestamp, but by 2010 in one trading day in February Yahoo's stock had ~30% of all trades share a second with at least one other trade. Such a rapid increase in the frequency of executed trades then

helps the stage for later study of the power high frequency trading has upon financial asset prices. Very close to the publishing of Bowsher's work Hewlett used a bivariate Hawkes process to predict future FX trading intensity conditional on recent trades. Hewlett utilized a bivariate Hawkes process because in that market liquidity takers (brokerage firm or other market maker) that need to fill a large order are faced with a dilemma. Either the market maker submits one large order which may perturb the current price or the large order is split into smaller orders while running the risk that others could front run the order if they see the pattern of buy and sell orders. Thus, Hewlett's model sought to combine order flow with the needs of brokers moving the vast sums of capital tied up in FX markets.

One of the most interesting / pertinent works for this paper then came in 2012 from Filimonov and Sornette and their study of market endogeneity (whether price changes are driven more by exogenous news events related to a firm or economy, or endogenously by market movements caused by positive feedback mechanisms that introduce correlation into the price changes). In particular this is when the branching ratio of a Hawkes process becomes a major concern as Filimonov and Sornette use that ratio as an appropriate proxy for market endogeneity. Bringing this framework to the E-mini S&P 500 future market the two authors then expanded the power of high frequency trading in setting prices. First, when considering E-mini data from 1998 to 2010 the dynamic branching ratio, estimated via Maximum Likelihood, Filimonov and Sornette calculated increased from the low value of ~ 0.3 to as high as 0.9 and consistently above 0.6 by 2004. However, while market endogeneity increased overall during this time it spikes did not occur during times of market stress that were

accompanied by fundamental exogeneous news like the debt downgrades of Greece and Portugal in 2010.

Contrasting with this fundamentally driven trading volume the authors do find a hard to explain increase in the branching ratio during the Flash Crash of May 6th, 2010. During a very short span of time (approximately 36 minutes total) the U.S. stock market declined nearly 9% during seemingly mundane afternoon trading before rapidly rebounding to close to market open prices. All of this occurred without any meaningful exogeneous news and therefore quickly demonstrated a far above average branching ratio for that trading day. Additionally, Filimonov and Sornette note that while no evidence has been discovered to directly link high frequency trading to the start of the Flash Crash it was associated with automated trading systems which might have exacerbated the extreme market movements experienced that day. Finally, their findings did reconfirm the increase of the model branching ratio alongside the rise of bigger and more active high frequency trading shops.

Multivariate Hawkes processes are also of great interest and Fauth and Tudor are one of the better examples of work in this arena. In particular the two researchers confirm the intuitive expectation that price fluctuations also vary based on trading on volume. Thus, when modeling stock trading activity with a *marked*, multivariate, Hawkes process greater accuracy is found than when using price changes alone. Strong evidence supports the conclusion that small quantities of shares do not affect the market in the same way huge quantities do and that quantities and price changes are closely linked. Additionally, Fauth and Tudor discuss the creation of a compound counting

process that allows for a growth process to change the underlying attributes of the process over time.

Other Interesting Areas of Self-Exciting Point Process Research

One of the most interesting applications of self-exciting processes outside of finance is the SEISMIC (Self-Exciting Model of Information Cascades) model. SEISMIC is a self-exciting point process that Zhao, Erdogdu, He, Rajaraman, and Leskovec from Stanford University created to model the final number of shares a social post will receive based on its resharing history thus far. Not only does this paper create a model very similar to the Hawkes process, but it also utilizes an infectiousness parameter in order to determine when the underlying behavior is *supercritical* or *subcritical*. Similar to a branching ratio the critical state determines whether SEISMIC model can or cannot predict the eventual number of total shares. Thus, instead of seeing changes in infectiousness as something to be measured like many finance paper the SEISMIC authors use it as a barometer to determine the potential effectiveness of a projection made at that time. Additionally, most tweets quickly fall to a subcritical state allowing for SEISMIC to make a prediction for 98.20% of all tweets after observing them for only 15 minutes.

SEISMIC therefore has many interesting properties in comparison to Hawkes model, but like the following algorithms it also remains easy to implement. Three of SEISMIC's most notable attributes are that it is a generative model, has scalable computation, and easy to interoperate. In order to implement SEISMIC only the time history of reshares and the degrees of the resharing nodes are needed without any parametric assumptions. Further, SEISMIC runs in linear computation time based on

the number of observed reshares and can be easily parallelized. Since the model synthesizes all its past history into a single infectiousness parameter this parameter conveys clear information about the information cascade and can be used as an input to other applications.

Algorithms for Simulation and Estimation

Throughout this section effort will be made to focus upon the implementation and intuition of the described processes. For a more rigorous treatment of the subject matter references have been made to the necessary articles, and some results assumed in order to maintain a more readable style. Overall, this style was deemed most consistent with the tone of the thesis and in greatest conformity with its goal of reaching undergraduate readers.

Poisson Processes

The first process simulation is a simple homogeneous Poisson process with rate λ . Recall that a Poisson process with rate λ has a distribution of $T(x)$, the interarrival time function for the process, is $T(x) = P(X \leq x) = 1 - e^{-\lambda x}$, $x \geq 0$; $E(X) = 1/\lambda$. Since this distribution function can be easily inverted with the natural log an algorithm for simulating a Poisson process with rate λ up to time T is:

Homogeneous Poisson Process Algorithm (Sigman, 2016):

1. $t = 0, N = 0$
2. Generate U (U is a random variable in $[0,1]$).
3. $t = t + [-(1/\lambda) \ln(U)]$. If $t > T$, then stop.
4. Set $N = N + 1$ and set $t_N = t$.
5. Go back to 2.

Reviewing the algorithm it can be seen that $-\frac{1}{\lambda} \ln(U)$ is replicating the exponential decay of the Poisson process. This is done by taking random values for U from $[0,1]$, making

the natural log this number positive and scaling by the inverse of λ . An examples of the values this takes on is:

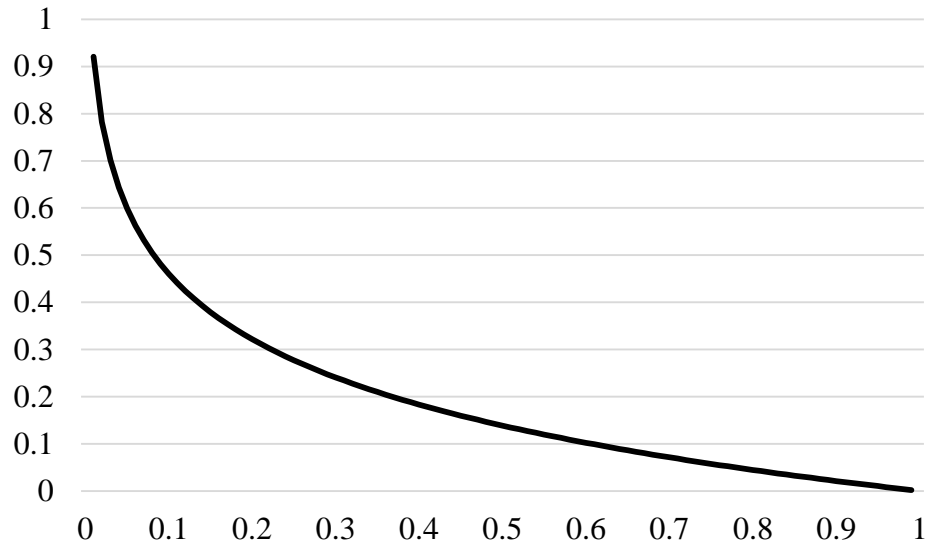


Figure 6: Graph of $\frac{-1}{\lambda} \ln(U)$ for $\lambda = 5$

Random values from this range are then added to the previous time in order to create the Poisson process.

With only a few minor modifications this algorithm can then be used to simulate two independent Poisson processes with rates λ_1, λ_2 up to time T :

Algorithm to Simulate Two Independent Poisson Processes¹:

1. $t = 0, t_1 = 0, t_2 = 0, N_1 = 0, N_2 = 0$, set $\lambda = \lambda_1 + \lambda_2$, set $p = \lambda_1/\lambda$.
2. Generate U .
3. $t = t + [-(1/\lambda) \ln(U)]$. If $t > T$, then stop.
4. Generate U . If $U \leq p$, then set $N_1 = N_1 + 1$ and set $t_{N_1} = t$; otherwise ($U > p$) set $N_2 = N_2 + 1$ and set $t_{N_2} = t$
5. Go back to 2.

Notice all that changed in this algorithm is that after the same first variable was created a second was used to determine which Poisson process would be allocated the event. Hence, the relative values of λ_1 and λ_2 became the determining factor in which process would advance more rapidly than the other. Additionally this algorithm easily generalizes to handle $k \geq 2$ independent Poisson processes. This is because all that would need to happen is a further bracketing of the relative sizes of each λ_i so that larger λ values may receive more points than smaller λ values.

However, as described in the point cluster section, the simulation of point clusters requires a non-homogeneous Poisson process. In order to simulate a non-homogenous Poisson process a thinning method may be used to slightly modify the structure of the above algorithms. The algorithm to generate a non-homogenous Poisson process with intensity $\lambda(t)$ bounded by λ^* up to time T with $N(T)$ arrival times $t_1, \dots, t_{N(T)}$ is then:

Non-homogenous Poisson Process With Intensity $\lambda(t)$ That is Bounded by λ^* Algorithm

(Sigman, 2016):

1. $t = 0, N = 0$
2. Generate U_1
3. $t = t + [-(1/\lambda^*) \ln (U_1)]$. If $t > T$, then stop.
4. Generate a U_2
5. If $U_2 \leq \lambda(t)/\lambda^*$, then set $N = N + 1$ and set $t_N = t$.
6. Go back to 2.

Notice that this process is extraordinarily similar to the independent Poisson process algorithm and simply checks to see if a point should be included or excluded from the

generated list. Now in order to prove this result we will begin with a nice fact about partitioning Poisson processes.

Theorem (Partitioning a Poisson process): If X comes from a Poisson process with rate α and if each object of X is, independently, type 1 or type 2 with probability p and $q = 1 - p$, then X_1 comes from a Poisson process with rate $p\alpha$, and X_2 comes from a Poisson process with rate $q\alpha$ and they are independent.

Proof²: We must show that

$$P(X_1 = k, X_2 = m) = e^{-p\alpha} \frac{(p\alpha)^k}{k!} e^{-q\alpha} \frac{(q\alpha)^m}{m!}. \quad (1)$$

By the properties of Poisson processes we know that

$$\begin{aligned} P(X_1 = k, X_2 = m) &= P(X_1 = k, X_2 = k + m) \\ &= P(X_1 = k \mid X = k + m) \times P(X = k + m). \end{aligned}$$

Given $X = k + m$, it then follows that $X_1 \sim \text{Binomial}(k + m, p)$ and

$$\begin{aligned} P(X_1 = k \mid X = k + m) \times P(X = k + m) &= \frac{(k + m)!}{k! m!} p^k q^m e^{-\alpha} \frac{\alpha^{k+m}}{(k + m)!} \\ &= e^{-\alpha} \frac{(p\alpha)^k}{k!} \frac{(q\alpha)^m}{m!}. \end{aligned}$$

Since $1 = p + (1 - p) = p + q$ we then know that $e^\alpha = e^{p\alpha} e^{q\alpha}$. This equality then shows the above equation reduces to (1) and proves the result. \square

This result then helps prove the initial result we sought:

Proof¹: [Thinning works] Let $\{M(t)\}$ be the counting process of the λ^* rate Poisson Process, and $\{N(t)\}$ be the counting process of the thinned process. In order to prove the result it must be shown that $\{N(t)\}$ has independent increments and that the increments are Poisson distributed with the correct mean, $m(t)$.

First, it is true that $\{N(t)\}$ has independent increments because $\{M(t)\}$ has independent increments and the thinning is done independently of $\{M(t)\}$. Thus, $\{N(t)\}$ inherits independence of increments from $\{M(t)\}$. So what is left to prove is that for each $t > 0$, $N(t)$ constructed from the thinning method has a Poisson distribution with mean $m(t) = \int_0^t \lambda(s) ds$. Since $M(t)$ is a Poisson distribution with mean $\lambda^* t$ it is possible to partition $M(t)$ into two types for each $t > 0$. If $N(t)$ represents the accepted points and $R(t)$ the rejected ones, $M(t)$ will be fully described by these two functions. Hence since $M(t)$ can be partitioned, it follows from the above Theorem that $N(t)$ has the desired Poisson distribution. Specifically the conditional on $M(t) = n$, it is possible to take n unordered arrival times as i.i.d. uniform $(0, t)$ random variables. Thus an arrival, denoted by $V \sim \text{Uniform}(0, t)$ will be accepted with conditional probability $\lambda(v)/\lambda^*$, conditional on $V = v$. Therefore the unconditional probability of acceptance is:

$$p = p(t) = E \left[\frac{\lambda(V)}{\lambda^*} \right] = \frac{1}{\lambda^* t} \int_0^t \lambda(s) ds$$

and we conclude from the partitioning theorem that $N(t)$ has a Poisson distribution with mean $\lambda^* t p = m(t)$. \square

Univariate Hawkes Process with Exponentially Decaying Intensity

For this first Hawkes process method we will consider a Hawkes process with exponentially decaying intensity. While this might be a special case of the more general Hawkes process it is numerically efficient to calculate and the “most widely implemented in practice” (Dassios, Angelos, and Zhao, 2013). The algorithm can be implemented quite easily because the random interarrival-times between events in the process are simulated by decomposing each event into two independent random

variables without inverting the underlying cumulative distribution function. Later this will be discussed as one of the frequent numerical limiting factors to implementing Hawkes processes, and therefore avoiding the inversion of the cumulative distribution function makes this implementation of Hawkes processes simpler and faster.

Additionally, this algorithm does not require stationarity. A few necessary definitions are:

- $a \geq 0$ is the constant reversion level (i.e. the background intensity)
- $\lambda_0 > 0$ is the initial intensity at time $t = 0$
- $\delta > 0$ is the constant rate of exponential decay
- $\{Y_k\}_{k=1,2,\dots}$ are sizes of self-excited intensity jumps, a sequence of i.i.d positive random variables with distribution function $G(y)$, $y > 0$

Written in our previous form this would therefore be the Hawkes process intensity

$$\lambda_t(t) = a + (\lambda_0 - a)e^{-\delta t} + \sum_{t_i < t} Y_i e^{-\delta(t - t_i)}.$$

Univariate Hawkes Algorithm⁴: *The simulation algorithm for one sample path of one-dimensional Hawkes process with exponentially decaying intensity $\{(N_t, \lambda_t)\}_{t \geq 0}$ conditional on λ_0 and $N_0 = 0$, with intensity jump-size distribution $Y \sim G$ and K event-times $\{T_1, T_2, \dots, T_K\}$:*

1. Set the initial conditions: $T_0 = 0$, $\lambda_{T_0^\pm} = \lambda_0 > a$, $N_0 = 0$ and $k \in \{0, 1, \dots, K - 1\}$
2. Simulate the $(k + 1)^{\text{th}}$ interarrival-time S_{k+1} by

$$S_{k+1} = \begin{cases} S_{k+1}^{(1)} \wedge S_{k+1}^{(2)} & D_{k+1} > 0 \\ S_{k+1}^{(2)} & D_{k+1} < 0 \end{cases}$$

where

$$D_{k+1} = 1 + \frac{\delta \ln U_1}{\lambda_{T^+k} - a}, \quad U_1 \sim U[0,1]$$

and

$$S^{(1)}_{k+1} = -\frac{1}{\delta} \ln D_{k+1},$$

$$S^{(2)}_{k+1} = -\frac{1}{a} \ln U_2, \quad U_2 \sim U[0,1].$$

Commentary: This then means that:

$$\begin{aligned} S^{(1)}_{k+1} \wedge S^{(2)}_{k+1} &= -\frac{1}{\delta} \ln \left(1 + \frac{\delta \ln U_1}{\lambda_{T^+k} - a} \right) \times -\frac{1}{a} \ln U_2 \\ &= \frac{1}{\delta a} \ln \left(1 + \frac{\delta \ln U_1}{\lambda_{T^+k} - a} \right) \times \ln U_2 \end{aligned}$$

where δ is the constant rate of exponential decay and $\lambda_{T^+k} - a$ is the current gap between the constant background intensity and the self-exciting endogenous excitement.

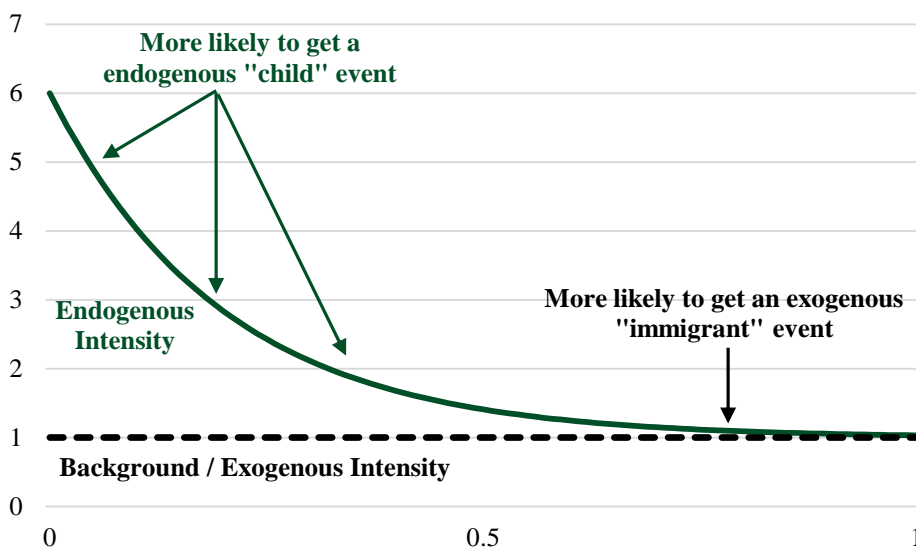


Figure 7: Example of a Generic Exponential Intensity Decay

Figure 7 helps demonstrate the changes in intensity by showing the large amount of endogenous intensity on the left, and the more dominant exogenous intensity on the right. Hence when the quantity $1 + \frac{\delta \ln U_1}{\lambda_{T^+k} - a}$ is large we are more likely to see a new child (endogenous event), and when this value is small a new child (exogenous event) is more likely. If $1 + \frac{\delta \ln U_1}{\lambda_{T^+k} - a}$ is positive then $D_{k+1} > 0$, we can solve the first natural log and multiply it with second logarithm to find the random value of the exogenous and endogenous points together. The random value of these two processes are then scaled by the exponential decay factor and the background intensity through $\frac{1}{\delta a}$. However, if this is negative then an exogenous point should be generated as the ratio was quite small (i.e. the endogenous intensity was small in relation to the constant exogenous intensity).

3. Record the $(k + 1)^{\text{th}}$ event-time T_{k+1} in the intensity process λ_t by

$$T_{k+1} = T_k + S_{k+1}$$

4. Record the change at the event-time T_{k+1} in the intensity process λ_t by

$$\lambda_{T^+_{k+1}} = \lambda_{T^-_{k+1}} + Y_{k+1}, \quad Y_{k+1} \sim G$$

where

$$\lambda_{T^-_{k+1}} = (\lambda_{T^+_k} - a)e^{-\delta(T_{k+1} - T_k)} + a.$$

Commentary: Recall that $\{Y_k\}_{k=1,2,\dots}$ are sizes of self-excited intensity jumps, a sequence of i.i.d. positive random variables with distribution function $G(y)$. So Y_{k+1} is the value by which the endogenous intensity of the process is increasing. However, this is an exponentially time decaying process with faster decay happening the larger the intensity currently is. This is why the calculation of

$\lambda_{T^-_{k+1}}$ not only has an exponential decay value, but also multiplies this by the size of the exogenous intensity minus the background exogenous intensity.

5. Record the change at the event-time T_{k+1} in the point process N_t by

$$N_{T^+_{k+1}} = N_{T^-_{k+1}} + 1.$$

For the proof of this result in its entirety the reader is encouraged to consult Dassios and Zhao (2013). Here we will give a rough overview of the proof, but will simply assume that the inverse of the cumulative distribution function can be replaced with two independent variables $S^{(1)}_{k+1}$ and $S^{(2)}_{k+1}$. While avoiding this proof loses some of the rigor of the result it makes the proof far more readable and focuses attention on the benefits this algorithm brings to implementation. The importance of this result for implementation that it allows for *exact simulation*,¹ which avoids “introducing discretization bias for associated estimators,” without numerical evaluation of the inverse of analytic distribution functions. Inverting analytic distribution functions generally requires using Brent’s method and involves intensive computations. Algorithmically solving this problem therefore keeps the benefits of the most precise methods of Hawkes process simulation while remaining easily computable.

Proof: Given a k^{th} event-time T_k , the point process then has the intensity process

$\{\lambda_t\}_{T_k \leq t \leq T_k + S_{k+1}}$ following the ODE

$$\frac{d\lambda_t}{dt} = -\delta(\lambda_t - a).$$

With the initial condition $\lambda_t|_{t=T_k} = \lambda_{T_k}$. The above ODE has a unique solution given by

¹ Here ‘exact’ simulation means a method of drawing an unbiased associated estimator through the entire simulation process.

$$\lambda_t = (\lambda_{T_k^+} - a) e^{-\delta(t-T_k)} + a, \quad T_k \leq t \leq T_k + S_{k+1}$$

and the cumulative distribution function of the $(k+1)^{\text{th}}$ interarrival-time S_{k+1} is given by

$$\begin{aligned} F_{S_{k+1}}(s) &= P\{S_{k+1} \leq s\} \\ &= 1 - P\{S_{k+1} > s\} \\ &= 1 - P\{N_{T_k+s} - N_{T_k} = 0\} \\ &= 1 - \exp\left(-\int_{T_k}^{T_k+s} \lambda_u du\right) \\ &= 1 - \exp\left(-\int_0^s \lambda_{T_k+v} dv\right) \\ &= 1 - \exp\left(-(\lambda_{T_k^+} - a) \frac{1-e^{-\delta s}}{\delta} - as\right). \end{aligned}$$

By the inverse transformation method it follows that

$$S_{k+1} \stackrel{D}{=} F_{S_{k+1}}^{-1}(U), \quad U \sim [0,1].$$

However, inverting the function $F_{S_{k+1}}(\cdot)$ can be avoided by decomposing S_{k+1} into two simpler and independent random variables $S_{k+1}^{(1)}$ and $S_{k+1}^{(2)}$ via

$$S_{k+1} \stackrel{D}{=} S_{k+1}^{(1)} \wedge S_{k+1}^{(2)}.$$

Assuming the ability to do this then means that the next time interval will be determined by the two random variables $S_{k+1}^{(1)}$ and $S_{k+1}^{(2)}$. These together then represent $F_{S_{k+1}}^{-1}(U)$ and give the next interarrival length. Therefore the $(k+1)^{\text{th}}$ event-time T_{k+1} in the Hawkes process will be given by

$$T_{k+1} = T_k + S_{k+1}$$

and the change in λ_t and N_t at time T_{k+1} then can be derived as $\lambda_{T_{k+1}^+} = \lambda_{T_{k+1}^-} + Y_{k+1}$,

$Y_{k+1} \sim G$ and $N_{T_{k+1}^+} = N_{T_{k+1}^-} + 1$ from steps 4 and 5 of the algorithm respectively. \square

Multivariate Hawkes Process with Exponentially Decaying Intensity

Making only a couple modifications to the univariate Hawkes process algorithm it is possible to extend the algorithm to multi-dimensional cases. With K joint event-times $\{T_1, T_2, \dots, T_K\}$ before time t a D -dimensional point process $\{N^{[j]}_t\}_{j=1,2,\dots,D}$ where $N^{[j]}_t = \{T^{[j]}_k\}_{k=1,2,\dots}$ can be defined by the following underlying intensity process:

$$\lambda^{[j]}_t = a^{[j]} + (\lambda^{[j]}_0 - a^{[j]}) e^{-\delta^{[j]}t} + \sum_{l=1}^D \sum_{0 \leq T^{[l]}_k < t} Y^{[j,l]}_k e^{-\delta^{[j]}(t-T^{[l]}_k)}$$

with $j \in \{1, 2, \dots, D\}$.

Where $\{Y^{[j,l]}_k\}_{j=l}$ are the sizes of self-excited intensity jumps and $\{Y^{[j,l]}_k\}_{j \neq l}$ are sizes of *cross-excited jumps*, and they are measurements of the impacts of *self-contagion* and *cross-contagion* respectively. Upon the arrival of an event in point process $N^{[l]}_t$, note that each marginal intensity process $\{\lambda^{[j]}_t\}_{j=1,2,\dots,D}$ experiences a simultaneous intensity jump of positive random size, and these intensity jumps can be either dependent or independent.

Multivariate Hawkes Process Algorithm (Dassios, Angelos, and Zhao, 2013): *The simulation algorithm for one sample path of a D -dimensional Hawkes process with exponentially decaying intensity $\{(N^{[j]}_t, \lambda^{[j]}_t)\}_{t \geq 0}$ for $j \in \{1, 2, \dots, D\}$ conditional on $\lambda^{[j]}_0$ and $N^{[j]}_0 = 0$, with K joint event-times $\{T_1, T_2, \dots, T_K\}$ in intensity processes:*

1. Set the initial conditions $T_0 = 0$, $\lambda^{[j]}_{T_0 \pm 0} = \lambda^{[j]}_0 > a^{[j]}$, $N^{[j]}_0$, $j \in \{1, 2, \dots, D\}$ and $k \in \{0, 1, 2, \dots, K-1\}$

2. Simulate the $(k+1)^{\text{th}}$ interarrival-time W_{k+1} by

$$W_{k+1} = \min\{S^{[1]}_{k+1}, S^{[2]}_{k+1}, \dots, S^{[D]}_{k+1}\}$$

where

$$W_{k+1} = S^{ll}_{k+1}$$

and each S^{lj}_{k+1} can be simulated in the same way as S_{k+1} as given by Step 2 of the univariate algorithm

Commentary: Notice that all that has changed in this situation is that interarrival time happens much faster because the algorithm merely picks the process that takes the least amount of time to cause an event. For actual implementations this means that the decay rate for time will generally be faster here than in the univariate case. This will then correctly allocate effects to other variables, and ensure the process continues at a reasonable rate.

3. Record the $(k+1)^{\text{th}}$ event-time T_{k+1} in the intensity process $\lambda^{[j]}_t$ by

$$T_{k+1} = T_k + W_{k+1}$$

4. Record the change at the event-time T_{k+1} in the intensity process $N^{[j]}_t$ by

$$N^{[j]}_{T^+_{k+1}} = \begin{cases} N^{[j]}_{T^-_{k+1}} + 1 & j = l \\ N^{[j]}_{T^-_{k+1}} & j \neq l \end{cases} \quad j \in \{1, 2, \dots, D\}$$

Commentary: This is simply a fancy way of updating a list of values to ensure that that we allocate the event to the right sub-process. Once the match has been made then the algorithm just pulls the right values in order to increase the right intensity.

5. Record the change at the event-time T_{k+1} in the intensity process $\lambda^{[j]}_t$ by

$$\lambda^{[j]}_{T^+_{k+1}} = \lambda^{[j]}_{T^-_{k+1}} + Y^{[j,l]}_{k+1}, \quad j \in \{1, 2, \dots, D\}$$

where

$$\lambda^{[j]}_{T^-_{k+1}} = (\lambda^{[j]}_{T^+_k} - a^{[j]})e^{-\delta^{[j]}(T_{k+1}-T_k)} + a^{[j]}.$$

Estimation of a Power Kernel

When seeking to estimate a point cluster with a non-homogenous Poisson process it is important to be able to find an estimate for $\lambda(t)$ in order to estimate the limiting factor for point acceptance appropriately. Previous work has been done with the exponential kernel, but in this situation the limiting factor can be estimated more easily with a power kernel. In particular the following method utilizes easy to follow steps that can be used on any list formatted data. Estimating $\lambda(t)$ through this method is done by using the power law function

$$M(t) = at^b$$

which has the intensity function

$$\lambda(t) = \frac{dM(t)}{dt} = abt^{b-1}.$$

Note that if $0 < b < 1$, the intensity function is decreasing and if $b > 1$ the rate is increasing. Then the modified *maximum likelihood estimation*² (MLE) coefficients, conditioned on time t_n , for the power model are (Tobias and Trindade, 2012)

$$\hat{b} = \frac{n-2}{\sum_{i=1}^{n-1} \ln \frac{t_n}{t_i}}, \quad \hat{a} = \frac{n}{t_n^{\hat{b}}}.$$

Similarly considering a fixed time T , so that the number of events N by time T is random, we can condition on N to get the MLEs as⁵

$$\hat{b} = \frac{N-1}{\sum_{i=1}^N \ln \frac{T}{t_i}}, \quad \hat{a} = \frac{N}{T^{\hat{b}}}.$$

These estimations can then be improved by estimating the kernel on k different time intervals, with similar conditions, to improve the estimates of the model

² The values that maximize the probability of obtaining a particular set of data, given the chosen probability distribution model.

parameters. When performing these tests it's possible to limit both by time and number of events to occur. Let T_q denote the length of time considered for the q th time interval, $q = 1, 2, \dots, k$ and let n_q denote the total number of events in the q th sample by the time T_q . Let t_{iq} be the i th event time within the q th sample. Now we will introduce a new variable N_q , which equal n_q if the data from the q th sample are time limited or equals $n_q - 1$ if the data on the q th system are event limited. Crow (1974) demonstrates that conditioning on either the number of events in each window or the length of time of each the unbiased MLE for b can be expressed in closed form as

$$\hat{b} = \frac{N_s - 1}{\sum_{q=1}^k \sum_{i=1}^{N_q} \ln \frac{T_q}{t_{iq}}}$$

where

$$N_s = \sum_{q=1}^k N_q.$$

The modified MLE for a is then

$$\hat{a} = \frac{\sum_{q=1}^k n_q}{\sum_{q=1}^k T_q^{\hat{b}}}.$$

Hence this process provides a way of using given data to estimate values for a and b that may be used to generate $\lambda(t)$ for a non-homogeneous Poisson process that produces children for point clusters.

Implemented Processes

In an Excel workbook³ each of these algorithms, except the multivariate Hawkes process were simulated. Some of the simulated results are:

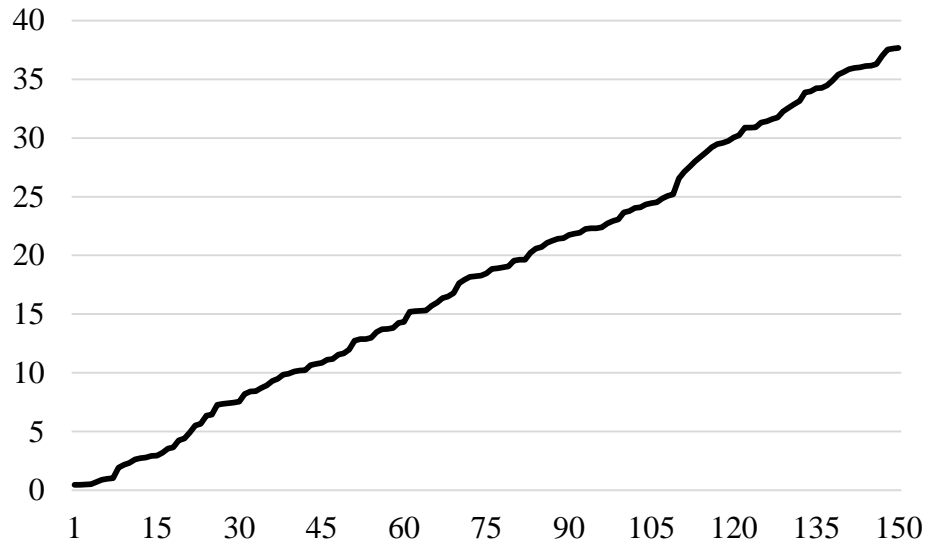


Figure 8: Homogenous Poisson Process with $\lambda = 4$ and $T = 50$

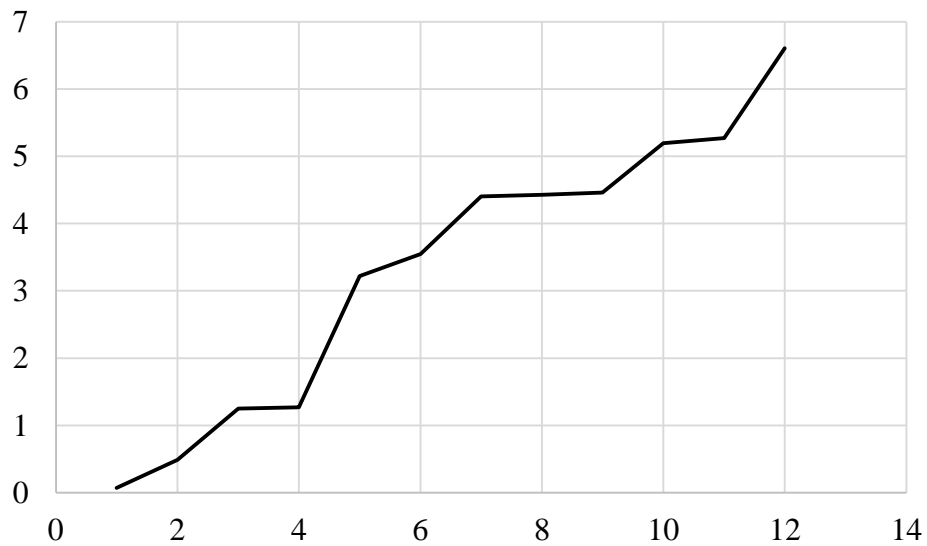


Figure 9: Non-homogenous Poisson Process: $\lambda = 7.5$, $T = 7.5$, & $\lambda(t) = \frac{n}{\tau} \exp\left(-\frac{t}{\tau}\right)$

³ <https://drive.google.com/open?id=0BxM3BpLkURvVVERBeHdQUnotYk0>

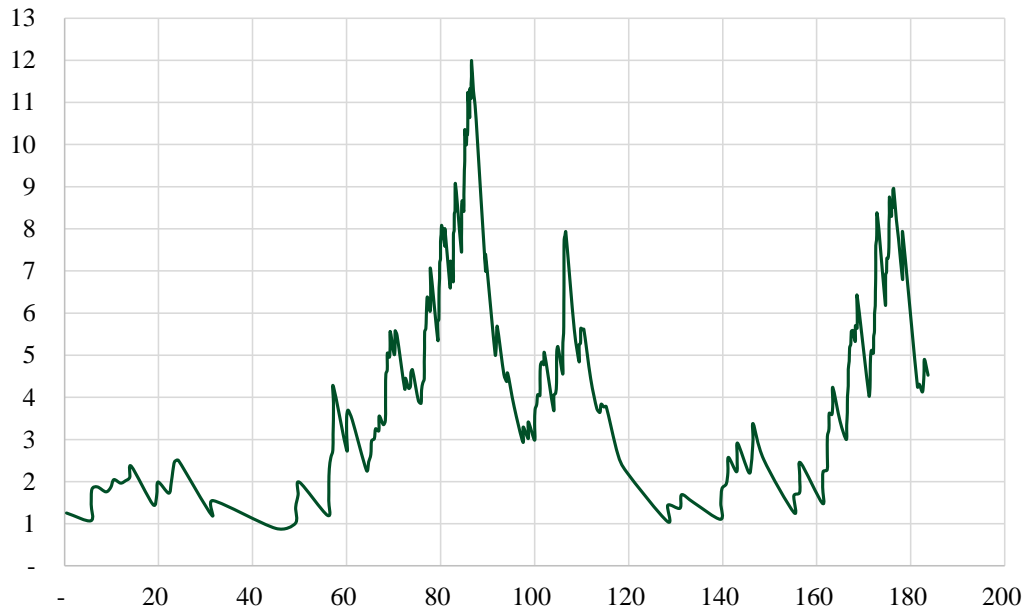


Figure 10: Intensity of Univariate Hawkes Process: $a = 0.45$, $\delta = 0.25$, $\lambda_0 = 0.9$, & Intensity Jumps of Size 0.40

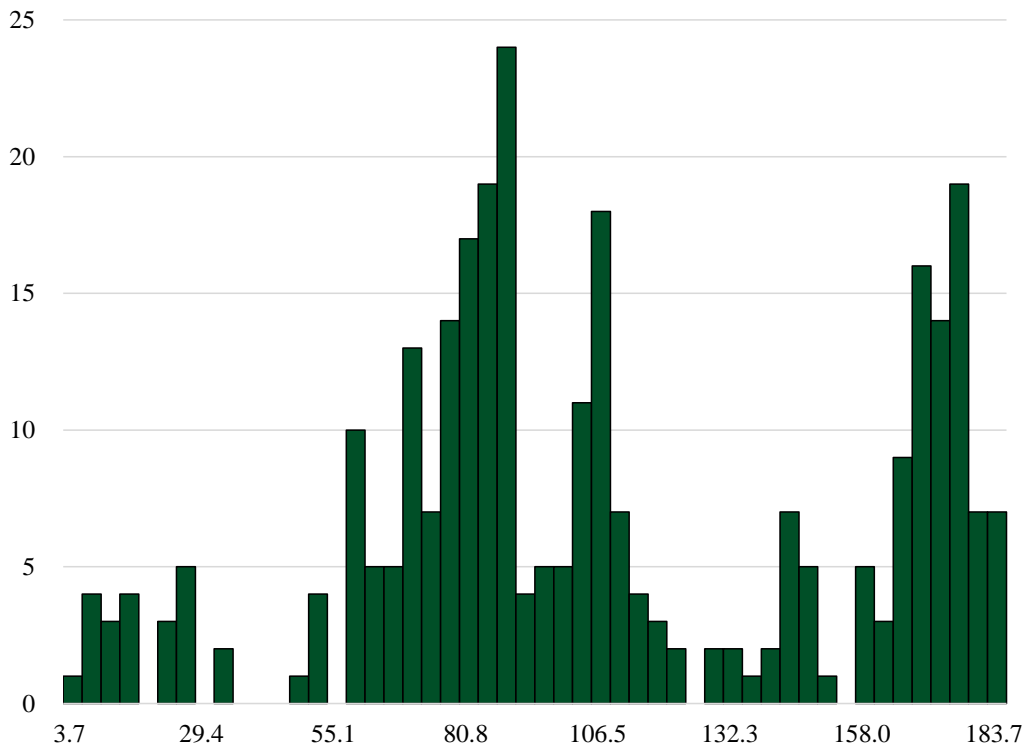


Figure 11: Histogram of Number of Events per Unit Time for Univariate Hawkes Process: $a = 0.45$, $\delta = 0.25$, $\lambda_0 = 0.9$, & Intensity Jumps of Size 0.40

Future steps for this work would then be to use financial data to better estimate parameters and kernels for these models. In some other fields it might be possible that

there are generally accepted values for how things are modeled, but financial models are almost always built around empirical data. This is also the case in many other fields where the relevant data is available to help tune and test a model before using for simulations or forecasts. Better tuning then ensures researchers find the best relationship between intensity (Figure 9) and the actual number of events occurring (Figure 10). While lacking such data does not immediately preclude a Hawkes process from being used to model a point process it does make the results of the model far less certain. Overall this reinforces that Hawkes processes are very robust and flexible models, but processes that are frequently most useful with adequate empirical data.

Future Work

Improve Simulation Implementations

Creating results in R, Matlab, or with a coding language would be one of the first necessary tasks. Accumulating the knowledge to do this would take a great deal of time, but would be necessary to continue research in an applied setting. Next, estimation methods would be researched to better estimate the parameters necessary for modeling empirical data. Then within one particular financial market a great deal of testing would need to be done to determine what time horizon to study, what kernel to use, how to best tune the model, what additional attributes would be meaningful for a multivariate Hawkes process, and how accurate models are for that arena. Having completed this survey theoretical modifications to the kernel would most likely be the most fruitful area of study, but finding numerically simple solutions for currently complex ones would also be an interesting pursuit.

Research such as this would then provide even greater clarity on the characteristics of some of the most important financial markets in the world. Most importantly work in this arena can explore the often dramatic reactions of financial markets to external events. Additionally, further research should help give a better image of the impact high frequency trading is having on global financial markets. While enormous development have occurred in high frequency trading some feel laws and regulators have failed to keep pace with these progressions. Studying the effects of highly computerized trading through Hawkes processes might therefore be quite helpful in modernizing outsiders understanding of this arena.

Bibliography

- Bacry, E.; Delattre, S.; Hoffmann, M.; Muzy, J. F. (2011). Modelling microstructure noise with mutually exciting point processes, Proceedings of the ICASSP.
- Bogle, John C. The Little Book of Common Sense Investing: The Only Way to Guarantee Your Fair Share of Market Returns. Hoboken, NJ: John Wiley & Sons, 2007. Print.
- Bowsher, C. (2007) Modelling security market events in continuous time: intensity based, multivariate point process models, Journal of Econometrics, 141(2).
- Buffett, Warren E. "The Superinvestors of Graham-and-Doddsville." Fiftieth Anniversary of Security Analysis. Columbia University, New York. 13 Mar. 2016. Speech.
- Crow, L. H. 1974. Reliability analysis for complex repairable systems. In Reliability and Biometry, ed. F. Proschan and R. J. Serfling, 379–410. Philadelphia: SIAM.
- Daley, D. J.; Vere-Jones, D. (2003) An introduction to the theory of point processes. Volume I. Springer: Heidelberg
- Dassios, Angelos, and Hongbiao Zhao. "Exact Simulation of Hawkes Process with Exponentially Decaying Intensity." Electronic Communications in Probability Electron. Commun. Probab. 18.0 (2013): n. pag. Web.
- Durrett, Richard. Essentials of Stochastic Processes. New York: Springer, 1999. Print.
- E. Bacry, K. Dayri, and J.-F. Muzy, "Non-parametric kernel estimation for symmetric Hawkes processes. application to high frequency financial data," European Physics Journal B 85 no. 5, (2012) 157.
- Fama, Eugene F. "Efficient Capital Markets: A Review of Theory and Empirical Work." The Journal of Finance 25.2 (1970): 383. Web. 13 Mar. 2016.
- Fauth, Alexis, and Ciprian Tudor. "Modeling First Line Of An Order Book With Multivariate Marked Point Processes." (2012): n. pag. 20 Nov. 2012. Web. 26 Apr. 2016.
- Filimonov, V. and D. Sornette (2012). Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. Physical Review E 85 (5), 056108.
- Filimonov, V. and D. Sornette (2015). Apparent criticality and calibration issues in the hawkes self-excited point process model: application to high-frequency financial data. Quantitative Finance (ahead-of-print), 1–22.

- Fonseca, José Da, and Riadh Zaatour. "Hawkes Process: Fast Calibration, Application to Trade Clustering and Diffusive Limit." SSRN Electronic Journal SSRN Journal (n.d.): n. pag. Web. 18 Feb. 2016.
- Graham, Benjamin, and David Dodd. *Security Analysis*. New York, NY.: McGraw-Hill Book, 1962. Print.
- Hardiman, S. J., N. Bercot, and J.-P. Bouchaud (2013). Critical reflexivity in financial markets: a hawkes process analysis. *The European Physical Journal B* 86 (10), 1–9.
- Hewlett, P. (2006) Clustering of order arrivals, price impact and trade path optimization, Workshop on Financial Modelling with Jump Processes, Ecole Polytechnique.
- Kagan, Y. Y. and L. Knopoff (1981). Stochastic synthesis of earthquake catalogs. *Journal of Geophysical Research: Solid Earth* (1978–2012) 86 (B4), 2853–2862.
- Large, J. (2007) Measuring the resiliency of an electronic limit order book, *Journal of Financial Markets*, 10.
- Lewis, P. A. and G. S. Shedler (1978). Simulation of nonhomogeneous Poisson processes by thinning. Technical report, DTIC Document.
- Lorenzen, F. "Analysis of Order Clustering Using High Frequency Data: A Point Process Approach." Thesis. Tilburg School of Economics and Management Finance Department, 2012. Print.
- MacKinlay, Daniel. Estimating Self-excitation Effects for Social Media Using the Hawkes Process. Thesis. Swiss Federal Institute of Technology Zurich, 2015. N.p.: n.p., n.d. Print.
- Malkiel, Burton Gordon. *A Random Walk down Wall Street: The Time-tested Strategy for Successful Investing*. New York: W.W. Norton, 2003. Print.
- Malkiel, Burton G. "The Efficient Market Hypothesis and Its Critics." *Journal of Economic Perspectives* 17.1 (2003): 59-82. Web. 13 Mar. 2016.
- Møller, J. and J. G. Rasmussen (2005). Perfect simulation of Hawkes processes. *Advances in applied probability*, 629–646.
- Møller, J. and J. G. Rasmussen (2006). Approximate simulation of Hawkes processes. *Methodology and Computing in Applied Probability* 8 (1), 53–64.
- Morzywolek, Pawel. "Non-parametric Methods for Estimation of Hawkes Process for High-frequency Financial Data." Thesis. Swiss Federal Institute of Technology Zurich, 2015. Print.

- Ogata, Y. (1981). On Lewis' simulation method for point processes. *Information Theory, IEEE Transactions on* 27 (1), 23–31.
- Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics* 50 (2), 379–402.
- Qingyuan Zhao, Murat A. Erdogdu, Hera Y. He, Anand Rajaraman, and Jure Leskovec. 2015. SEISMIC: A Self-Exciting Point Process Model for Predicting Tweet Popularity. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*. ACM, New York, NY, USA, 1513-1522. DOI=<http://dx.doi.org/10.1145/2783258.2783401>
- Rambaldi, M., P. Pennesi, and F. Lillo (2014). Modeling FX market activity around macroeconomic news: a Hawkes process approach. *arXiv preprint arXiv:1405.6047*.
- Schoenberg, Frederic Paik. "Introduction to Point Processes." UCLA, n.d. Web. 13 Mar. 2016.
- Sigman, Karl. 1 Poisson Processes, and Compound (batch) Poisson Processes (n.d.): n. pag. Columbia University. Web. 12 Apr. 2016. <<http://www.columbia.edu/~ks20/4703-Sigman/4703-07-Notes-PP-NSPP.pdf>>.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Volume 26. CRC press.
- S. J. Hardiman, N. Bercot, and J.-P. Bouchaud, "Critical reflexivity in financial markets: a Hawkes process analysis," *The European Physical Journal B* 86 no. 10, (2013) 1–9.
- Soros, George. *The Alchemy of Finance: Reading the Mind of the Market*. New York: Simon and Schuster, 1987. Print.
- Tobias, Paul A., and David C. Trindade. *Applied Reliability*. 3rd ed. Boca Raton: CRC, 2012. Print.
- Utsu, T. and Y. Ogata (1995). The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth* 43 (1), 1–33.