Prequel to Hawkes Processes: An Overview of Temporal and Spatio-Temporal Point Processes and Some Simulations

Frances Lin

June 2021

Major Professor: James Molyneux

Committee Members: Lisa Madsen & Charlotte Wickham

Abstract

1 Introduction

Real-world data are often spatial, temporal, or spatio-temporal in nature. Spatial data (e.g. soil properties, housing prices) often involves locations such as points and areas. Temporal data (e.g. sensor readings, stock prices) often involves times such as moments and intervals. Spatio-temporal data involves both locations and times. Examples of spatio-temoral data include chemical wastes, forest inventories, remotely sensed images, earthquake epicenters, disease cases, map services and travel times, to name a few.

There are various models suitable for modelling spatial, temporal, or spatio-temporal data, and which models to use depend on questions of interest. We focus primarily on point process models here. Point process models are useful for describing phenomena that occurs at random locations, times, or locations and times, and the questions of interests typically are: Does the rate of events vary with locations, times, or locations and times? Do events appear clustered? Do events trigger subsequent events?

Spatial data can be broadly categoried into three types: geostatistical (point process) data, areal data, and point pattern data. We are in the 3rd category in which questions about point pattern data typically are: Is there clustering of events? Can we define a point process that captures the events? Examples of such models include Cox and cluster processes.

On the other hand, when dealing with temporal data, (marked) point processes are sometimes used interchangeable with time series, and vice versa. One major distinction, though, is that in point processes, time intervals are treated as continuous, whereas in time series, they are treated as discrete. Examples of such models include (temporal) Poisson and Hawkes processes.

Hawkes process is also known as self-exciting point process. More specifically, the original Hawkes processes (e.g. ETAS model) are temporal, whereas the more recently developed self-exciting point processes have been extended from temporal Hawkess processes to account for both the spatial and temporal aspects of the data.

The defining characteristic of Hawkes processes is that it self-excites. In other words, the occurrence of an event increases the occurrence of future events. For example, in seismology, an event can be an earthquake occurrence that causes aftershocks. In criminology, an event can be a gang rivalry that triggers retaliations following the gang crime.

In addition to modelling earthquake epicenters in seismology and crime patterns in criminology, Hawkess processes have also been used in modelling events such as forest wildfires, insurance claims, financial transcations, social network events, neuron activities, and disease spread or transmission. Thus, it can be found

in a wide variety of fields such as emergency and disaster management, insurance, finance, social network, neuroscience, and epidemiology.

More recent work have been focused on modelling..... and leveraging the advancement of computing to process and model big data.

Given the flexibility and applicability of Hawkess processes, it is surprising to see that Hawkess processes have not gain enough attention from the machine learning communities, which would find their predicting capabilities beneficial. In addition, understanding Hawkess processes would benefit from knowing some of the relevant point processes (e.g. nonhomogeneous Poisson, Cox and cluster processes) which are often left out from graduate-level, introductory spatial statistics and stochastic processes courses.

The objective of this project is to give an overview of various types of point processes so that readers of interest have the necessary background knowledge to understand Hawkes and self-exciting processes. The outline of this project is as follows: In *Section 2*, we introduce, define and discuss properties of counting process, (homogeneous and nonhomogeneous) Poisson process, cluster process, Hawkes process, and spatio-temporal self-exciting process. In *Section 3*, we discuss in particular the thinning algorithms (acceptance-rejection method) and simulate selected processes in 1D and 2D. In *Section 4 and 5*, we discuss recent and future work of Hawkes and self-exciting processes.

2 Introductions, Definitions and Properties

2.1 Counting Process

Let us begin with counting process. Counting processes count the occurrences (or numbers) of events over time. For example, if we were to count the numbers of events $N(t)$ such as the numbers of customers arriving at a supermarket or the numbers of phone calls receiving at the help line up to some time t, we can use counting processes.

The processes require that the total number of events $N(t)$ has to be greater than zero, the total number of events is an integer,..... In addition, counting processes is independent, stationary, and homogeneous. In other words, the numbers of events $N(t)$ occur in disjoint interval t are independent, the distribution of the numbers of events $N(t)$ depends only on the length of the interval t, and the transition probability between any two states at two times depends only on the difference between the states.

Before formally define counting process, first we need to define stocastic process and point process. Stocastic process is a collection of random variables indexed by time t (or by space s or by space-time, but we restrict ourselves to time t for now). The differences between stocastic process and point process are not clear, but they are included to define counting process.

Definition 2.1.1 (Stochastic Process) A stochastic process is a family of random variables indexed by time t and is defined as

$$\{X_t\}_{t\in T}$$

Definition 2.1.2 (Point Process) Let $\{T_i\}_{i\in N}$ be a sequence of non-negative random variables such that $T_i < T_{i+1}\ \forall i \in N$, a point process on R^+ is defined as

$$\{T_i\}_{i\in N}$$

Next, we define a function $N(t)$ such that it counts the total number of events up to some time t. For example, the total number of event from $t = 0$ up to $t = 1$ is denoted as $N(0)$, the total number of events from $t = 1$ up to $t = 2$ is denoted as $N(1)$, and so on.

Definition 2.1.3 (Counting Process) Let $N(t)$ be the total number of events up to some time t such that the values are nonnegative, interger and nondecreasing, a stocastic process is said to be a counting process and is defined as

$$\{N(t), t \geq 0\}$$

Here, if we let the first occurrence of event occurs at $time = T_1$, the 2nd occurence of event occurs at $time = T_2$, and so on, then with counting process, we again add up the numbers incremently. Thus, $N(0)$ indicates the total number of events from $t = 0$ up to $t = 1$, which is 0 since there is no event occur in $(0, 1]$, $N(1)$ is the total number of events from $t = 1$ up to $t = 2$, which is 1 since there is a total of one event occur in $(1, 2]$, and so on.

A PICTURE HERE WOULD HELP

Definition 2.1.4 (Counting Process) Let $\{T_i\}_{i \in N}$ be a point process, a counting process associated with $\{T_i\}_{i \in N}$ is defined as

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

Corollary **2.1.1** A counting process satisfies that

1. $N(t) \geq 0$
2. $N(t)$ is an integer
3. If $s \leq t$, then $N(s) \leq N(t)$
4. If $s < t$, then $N(t) - N(s)$ is the number of events occur in the interval (s, t]

In other words, 1. the total number of events has to ≥ 0, 2. the total number of events is an integer,. . . . . .

Proposition 2.1.1 A counting process has the following properties

1. Independence
2. Stationarity
3. Homogeneity

In other words, 1. the numbers of events $N(t)$ occur in disjoint interval t are independent, 2. the distribution of the numbers of events $N(t)$ depends only on the length of the interval t, and 3. the transition probability between any two states at two times depends only on the difference between the states.

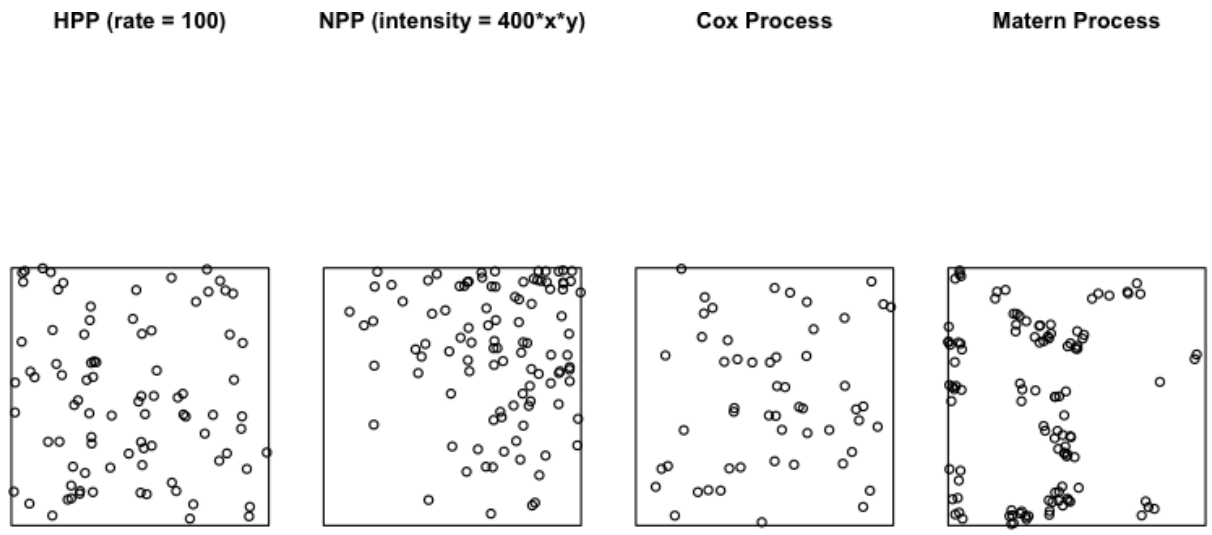

Figure 1: Left: HPP (rate = 100) Middle 1: NPP (intensity = $400xy$) Middle 2: Cox (intensity = exp(n = 1, rate = 1/100)) Right: Matern (kappa = 20, r = 0.05, mu = 5)

A Motivating Example

Before we delve into various point processes, let us first look at a motivating example. The figure on the left is a realization of HPP (homogeneous Poisson process) with constant $rate = 100$ in $\mathbb{R}^2$. The figure in the middle 1 is a realization of NPP (nonhomogeneous Poisson process) with $intensity\ function = 400xy$ in $\mathbb{R}^2$. The figure in the middle 2 is a realization of Cox process with $intensity\ function = exp(1, 1/100)$. The figure on the right is a realization of Matern cluster process with $kappa = 20, r = 0.05, mu = 5$ in $\mathbb{R}^2$.

We can see that while HPP points appear randomly spaced, NPP points seem to concentrate at the upper-right corner. Both Cox process and Matern cluster process points appear clustered, but the way the points cluster differ. Points in Cox process cluster accordingly to some specified distribution, whereas points in cluster process cluster in some defined area. We elaborate more and discuss in details each type of point process in the following sections.

2.2 Poisson Process

Homogeneous Poisson process (HPP) is one of the simplest yet most-widely used point processes. For example, if we were to model the numbers of events such as the numbers of bus arrivals at a bus stop, the numbers of car accidents at a site or the requests for documents on a web server, we can use Poisson processes.

Like counting process, HPP is also independent, stationary, and homogeneous. In addition, we assume that the numbers of events $N(t)$ follows a Poisson distribution with a constant rate λ and the interarrival times between events W are exponentially distributed.

Definition 2.2.1 (Poisson Process) If the following conditions hold, a counting process $\{N(t), t \geq 0\}$ is said to be a Poisson Process with constant rate (or intensity) $\lambda > 0$

1. $N(0) = 0$
2. $N(t)$ has independent increments
3. $P(N(t+h)) - N(t) = 1)) = \lambda h + o(h)$
4. $P(N(t+h)) - N(t) \geq 2)) = o(h)$

where $o(h)$ function of little o. . . . . .

In other words, 1. the process starts at $t = 0$, 2. the increments are independent, 3. λ is the rate (or intensity), and 4. no 2 or more events can occur at the same location.

Proposition 2.2.1 Homogeneous Poisson Process has similar properties as those of counting process; it is independent, stationary, and homogeneous. These properties follow nicely from the above definition.

A PICTURE HERE WOULD HELP

Again, we denote the total number of events as $N(t)$. We also denote the interarrival times between events as W. For example, while T_1 is the time of the first occurence of event and T_2 the time of the second occurence of event, the elapsed time between the none and first event is W_1 and the elapsed time between the first event and second event is W_2.

We assume that the numbers of events $N(t)$ follows a Poisson distribution with a constant rate λ and the interarrival times between events W are exponentially distributed.

Proposition 2.2.2 Poisson process has additional properties

1. The number of events in any interval t, $N(t)$, $\sim Pos(\lambda t)$. That is, for all $s, t \geq 0$

$$P(N(t+s) - N(s) = n) = P(N(t) - N(0)) = P(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

$n = 0, 1, ...$

2. The interarrival times, W, $\overset{iid}{\sim} exp(\frac{1}{\lambda})$. That is, for rate $\lambda > 0$, the interarrival time W_i $i = 1, 2, ...$

$$P(W_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$

This is because p(1^{st} arrivial arrives after time t) is the same as p(no arrivial in the interval [0,t]). Similarly, W_2 also $\sim exp(\frac{1}{\lambda})$ since

$$P(W_2 > t|W_1 = s) = P(N(t+s) - N(s)|N(s) - N(s^-) = 1) = P(N(t+s) - N(s)) = P(N(t) = 0) = e^{-\lambda t}.$$

2.3 Nonhomogeneous Poisson Process

Assuming that the rate is constant is often not realistic in practice. We may want a model that allows for more flexibility. Indeed, nonhomogeneous Poisson processes (NPP) is a generalization of homogeneous Poisson processes that allow the rate (or intensity) λ to vary with function of time t.

Previously, we assume that the intensity λ is constant. If we have reasons to believe that the intensity is not constant, we should model using nonhomogeneous Poisson processes. Back to the supermarket example, if we were to model the number of customers arriving at a supermarket and we have reasons to believe that the arrivial rate of customers is higher during lunch time as compared to say, 2pm, we could model using nonhomogeneous Poisson processes.

In contrast to HPP, NPP is independent but not stationary nor homogeneous. In addition, for NPP, we assume that $N(t)$ follows a Poisson distribution with an intensity function $\lambda(t)$. That is, the intensity now varies with the location of the time interval.

Definition 2.3.1 (Nonhomogeneous Poisson Process) If the following conditions hold, a counting process $\{N(t), t \geq 0\}$ is said to be a nonhomogeneous Poisson Process with intensity function of time $\lambda(t), t > 0$

1. $N(0) = 0$
2. $N(t)$ has independent increments
3. $P(N(t+h)) - N(t) = 1)) = \lambda(t)h + o(h)$
4. $P(N(t+h)) - N(t) \geq 2)) = o(h)$

Proposition 2.3.1 Nonhomogeneous Poisson process is independent but not stationary nor homogeneous. Homogeneous Poisson process has stationary increments since the distribution of the numbers of events $N(t)$ that occur in any interval of time t depends only on the length of the interval t but not the location of the interval t. In contrast, nonhomogeneous Poisson process does not have stationary increments since the distribution of $N(t)$ can change when shifted in t. Since stationarity implies homogeneity, nonhomogeneous Poisson process is nonhomogeneous.

Recall that for HPP, we assume that the numbers of events $N(t)$ follows a Poisson distribution with a constant intensity λ, here we assume that $N(t)$ follows a Poisson distribution too but with an intensity function $\lambda(t)$ such that the intensity vary with function of time.

Proposition 2.3.2 Nonhomogeneous Poisson process has additional properties

1. The number of events in any interval t, $N(t)$, $\sim Pos(\Lambda(t) = \int_0^t \lambda(s)ds)$. That is, for all $s, t \geq 0$

$$P(N(t) = n) = \frac{(\int_0^t \lambda(s)ds)^n e^{-\int_0^t \lambda(s)ds}}{n!}$$

$n = 0, 1, ...$

2. The law of occurance for the next point can be determined by

$$P(N(a, b] = 0) = e^{-\int_a^b \lambda(s)ds}.$$

2.4 Cox and Cluster Process

Even more flexible models than NPP are Cox and cluster processes that allow dependence between events. Previously, we assume independence between events. That is, whether events occur at a constant rate λ (e.g. HPP) or depend on an intensity function $\lambda(t)$ (e.g. NPP), they occur independently. Here, we discuss models that allow dependence between events.

Examples that can be modelled using Cox and cluster processes include seedlings and saplings of California redwood, locations of emergent plants, and locations of trees. In these examples, the patterns appear to be clustered.

We can think of Cox process as a hierarchical model with two levels and cluster process such as Neyman-Scott process a hierarchical model with three levels. **ELABORATE MORE**.

In Cox processes (or doubly stochastic Poisson process), the randomness arises from two parts. Not only the randomness occurs at different location of the time interval as in the case of a NPP, but the governing function $\Lambda(u)$ is also random, instead of governing by a determinist function $\lambda(t)$ also as in the case of a NPP. In other words, the intensity function $\Lambda(u)$ is also treated as random. **PICK ONE TO PLOT THEN TALK IN DETAILS**. Other examples of Cox processes include mixed Poisson process, log Gaussian Cox process, and shot noise Cox process.

In cluster processes, the randomness also arises from two steps: First, 'parent' points $\mathbf{Y}$ is generated. Then, each 'parent' point $y_i \in \mathbf{Y}$ gives to 'offspring' points z_{ij}. All the 'offspring' points Z_{ij} form a cluster process $\mathbf{X}$ and only $\mathbf{X}$ is observed.

Specific models of cluster processes depend on the choices of assumptions. Matern cluster process, for example, involves generating homogeneous Poisson parents and each parent gives rise to Poisson number of offspring uniformmly distributed in a disc of radius r centered around the parent. Other examples of cluster processes include Neyman-Scott process and Thomas cluster process.

In addition, the differences between Cox process and cluster process are such that Cox process conditions on (?) random field, whereas cluster process conditions on..... (?) in bounded region.

Definition 2.4.1 (Cox Process) Let $\Lambda = (\Lambda(u))_{u \in S \subseteq R^d}$ be a non-negative random field such that $\Lambda(u)$ is a locally integrable function. If $X \mid \Lambda \sim Pos(\Lambda)$, then X is said to be a Cox process driven by Λ with intensity function $\lambda(u) = E(\Lambda(u))$. That is,

$$P(N(u) = n) = \frac{(\lambda(u))^n e^{-\lambda(u)}}{n!} = \frac{(E(\Lambda(u)))^n e^{-E(\Lambda(u))}}{n!} =$$

(?)

$$\Lambda(u) \overset{a.s.}{=} \int_u \lambda(x)dx$$

(?) How to get here

$$= \int_0^\infty \frac{x^n e^{-x} F_u(dx)}{n!}$$

Note. Λ is a random field means that $\Lambda(u)$ is a random variable $\forall u \in S$.

Note. $\Lambda(u)$ is a locally integrable function means that $E(\Lambda(u))$ exists and is locally intergrable with probabiliy 1.

Proposition 2.4.1 Propertities of Cox process are as follow

1. Propertities of Cox process X follow immediately from the properties of Poisson process X | Λ. For example, if Λ is stationary, then X is stationary.
2. For bounded $B \subseteq S$, the void probabilities are given by

$$\nu(B) = E(P(N(B) = 0)|\Lambda) = E(exp(-\int_B Z(u)du)).$$

Note. The void (or avoidance) probability ν is defined as the probability that no points of a point process N existing in B where B is a subset of the underlying space R^d.

Definition 2.4.2 (Cluster Process) Let x be points in a point process N and replacing every x with a cluster of points N_x, then the union of all the clusters forms a cluster process N_c. That is,

$$N_c = \bigcup_{x \in N} N_x$$

Note. Each N_x is a finite point process 'centered' at x and it is assumed that each N_x is independent of one another.

Note. N_x can be thought of as 'offspring'??

In addition, condition on..., the cluster process has the intensity function...

Proposition 2.4.2 Cluster process has the following model assumptions

1. Poisson parents
2. Independent clusters
3. Identically distributed clusters
4. Offsprings independent within a cluster
5. Poisson number of offsprings
6. Isotropic clusters

In other words, 1. 'parent' points follow a Poisson distribution, 2. clusters are independent of each other, 3. clusters, when shifted, have the same distributions, 4. the locations of 'offspring' points of each parent point are independently and identically distributed, 5. the numbers of 'offspring' points of each parent point follow a Poisson distribution, and 6. the distribution of 'offspring' points for each parent point depends only on the distance between the 'parent' and the 'offspring'.

Under assumption 1 - 4, it is a Neyman-Scott process. Under assumption 1 - 5, the cluster process is a Cox process. Under assumption 1 - 6, it can be a Matern cluster process or a Thomas cluster process.

2.5 Hawkes Process

Hawkes process is also known as a self-exciting point process.

Like Cox and cluster process, the model also allows depdence between events. However, their depdence differ. In Hawkes process, the occurrence rate of the events depends not only on time t but also past events $\mathcal{H}_t^N$ up to some time t. Neither Cox nor cluster process captures the past history of events.

Because the intensity is now as a function of past history, a defining characteristic of Hawkes processes is that it self-excites. Examples that can be modelled using Hawkes processes include locations of earthquake epicenters, locations of crimes, and locations of patients with a disease. In these examples, the occurrence of an event increases the occurrence of subsequent events.

We can think of Hawkes process as a model that incorporates cluster process and conditional intensity function.

Next we want to define conditional intensity function since it is through which a point process is fully characterized.

Conditional intensity function can be thought as the instantaneous rate of events per unit time (, per unit space, or per unit space-time). For example, for HPP, $\lambda(t|\mathcal{H}_t) = \lambda$, for HPP, $\lambda(t|\mathcal{H}_t) = \lambda(t)$, for Cox process, $\lambda(t|\mathcal{H}_t)$ is a Papangelou conditional intensity function. For Hawkes process, $\lambda(t|\mathcal{H}_t)$ is a function of past history.

Definition 2.5.0 (Conditional Intensity Function) Let $N(t)$ be the numbers of events $N(t)$ that occur in any interval of time t , the conditional intensity function $\lambda(t)$ with respect to $\mathcal{H}_t$ is defined as

$$\lambda(t|\mathcal{H}_t) = \lim_{h\to 0^+} \frac{P(N(t,t+h] > 0|\mathcal{H}_t)}{h} = \lim_{h\to 0^+} \frac{E(N(t,t+h)|\mathcal{H}_t)}{h}$$

where $\mathcal{H}_t$ is the history prior to time t here in the case of a Hawkes process.

Definition 2.5.1 (Hawkes Process) A counting process $\{N(t), t \geq 0\}$ associated with past events $\{\mathcal{H}_t^N, t > 0\}$ is said to be a Hawkes process with conditional intensity function $\lambda(t|\mathcal{H}_t^N), t > 0$ and takes the form

$$\lambda(t|\mathcal{H}_t^N) = \lambda_0(t) + \sum_{i:T_i<t} \phi(t - T_i)$$

where

- $\lambda_0(t)$ is the base intensity function (or μ the constant background rate)
- $T_i < t$ are the events time occur before current time t
- $\phi(\cdot)$ is the kernel function (or $g(\cdot)$ the triggering function) through which intensity function depends on past events
- $\mathcal{H}_t^N$ is the natural filtration (or simply $\mathcal{H}_t$ the past history) which represents the internal history of N up to time t

Corollary **2.5.1** Hawkes process satisfies that

1. $N(t) = 0$
2. $\lambda(t|\mathcal{H}_t^N) = \lambda_0(t) + \int_{-\infty}^{t} \phi(t - T_i)dN(s) = \lambda_0(t) + \sum_{i:T_i<t} \phi(t - T_i)$
3. $P(N(t + h)) - N(t) = 1)|\mathcal{H}_t^N) = \lambda(t)h + o(h)$

4. $P(N(t+h)) - N(t) \geq 2)|\mathcal{H}_t^N) = o(h)$

2.5.1 Choices of $\phi(\cdot)$ include, for example, exponentially decaying function and power-law kernel, and they take the form of

$$\phi(x) = \alpha e^{-\beta x}$$

$$\phi(x) = \frac{\alpha}{(x+\beta)^{\eta+1}}$$

2.5.2 There are two ways to view Hawkes processes

1. Intensity-based Hawkes Process

Here, Hawkes process is defined through conditional intensity process.

In **Section 3.5**, we discuss algorithm for simulating intensity-based Hawkes process in details.

2. Cluster-based Hawkes Process

Alternatively, Hawkes process can also be defined through marked Poisson cluster process.

2.5 Spatio-Temporal Hawkes Process

Spatio-temporal Hawkes processes is an extention of temporal Hawkes processes. Recall that temporal Hawkes processes take the form of

$$\lambda(t|\mathcal{H}_t) = \mu + \sum_{i:T_i<t} g(t - t_i)$$

Spatio-temporal Hawkes processes take the form of

$$\lambda(t|\mathcal{H}_t) = \mu(s) + \sum_{i:T_i<t} g(s - s_i, t - t_i)$$

where

- $s_i, i = 1, 2, ...$ are the sequence of locations of events
- $t_i, i = 1, 2, ...$ are the times of events

Next, we simulate some of the aforementioned point processes in 1D ($\mathbb{R}^1$) through algorithms and the processes in 2D ($\mathbb{R}^2$) using `R` packages.

3 Algorithms and Simulations

3.2 Poisson Process

Algorithm 1

3.3 Nonhomogeneous Poisson Process

There are multiple ways to simulate nonhomogeneous Poisson process: 1) inversion, 2) order statistics, 3) thinning and 4) hybride (inversion + thinning).

In this example, we use the thinning algorithm (or acceptance-rejection method) to simulate nonhomogeneous Poisson process with the intensity function $\lambda(t) = \ldots$ since it is one of the most popular choices for both temporal and spatio-temporal cases.

Broadly put, thinning algorithm involves randomly deleting points from a point pattern. **ELABORATE MORE**.

Algorithm 2

3.5 Hawkes Process

Similarly, there are multiple ways to simulate Hawkes process. Here, we use thinning algorithm (or acceptance-rejection method).

Algorithm 3

3.6 Simulations in 2D using the spatstat package in `R`

All of the corresponding plots can be found in the Appendix section and all of the plots in this section are created using the **spatstat** package in `R`.

HPP

NPP

Cox process

Matern cluster process involves generating homogeneous Poisson parents and each parent gives rise to Poisson number of offspring uniformmly distributed in a disc of radius r centered around the parent. `kappa` controls the intensity of the cluster centers and allows us to specify the number of clusters. `r` specifies how far away cluster is from one another in radius, and `mu` gives the mean number of points per cluster.

The following functions use thinning algorithm. Simulations of Matern I and Matern II processes are generated using the `rMaternI` and `rMaternII` functions of the **spatstat** package.

4 Conclusions and Discussion

5 Recent Advancement

Acknowledgments

Reference

Chen, Y. (2016). Thinning algorithms for simulating point processes. Florida State University, Tallahassee, FL.

Obral, K. (2016). Simulation, estimation and applications of hawkes processes. (Master's thesis, University of Minnesota, Twin Cities, United States).

Rizoiu, M. A., Lee, Y., Mishra, S., & Xie, L. (2017). A tutorial on hawkes processes for events in social media. arXiv preprint arXiv:1708.06401.

Krishna, R. (2015). Simulation of Non-Homogeneous Poisson Processes.

Pasupathy, R. (2010). Generating homogeneous poisson processes. Wiley encyclopedia of operations research and management science.

Terminology

Stochastic Process

Counting Processes

Poisson Process

Nonhomogeneous Poisson Process

Cox Process

Appendix