

POINT PROCESS MODELLING OF A LIMIT ORDER BOOK

Oden Petersen

Supervisor: Dr. Tom Stindl

School of Mathematics and Statistics
UNSW Sydney

September 2024

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF
BACHELOR OF ADVANCED MATHEMATICS WITH HONOURS

Plagiarism statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration I am agreeing to the statements and conditions above.

Signed: _____

Date: _____

Acknowledgements

Abstract

This is the abstract

Contents

Chapter 1	Introduction	1
1.1	Limit Order Books	2
1.2	The Matching Algorithm	3
1.2.1	The Bid and Ask	3
1.2.2	Liquidity	3
1.2.3	Price-Time Priority	4
1.3	Queueing Models of the Order Book	4
1.4	Overview of Point Processes	4
1.4.1	Definition	5
1.4.2	The Counting Measure	5
1.4.3	The Expectation Measure	6
1.4.4	Intensity of a Point Process	6
1.4.5	Residuals	7
1.4.6	Marked and Multivariate Point Processes	7
1.5	Point Process Models of Book Updates	8
1.5.1	Poisson Processes	8
1.5.2	Inhomogeneous Poisson Process	9
1.5.3	Autoregressive Point Processes	9
1.5.4	State Dependence	10
1.5.5	Regression on marks	10
1.5.6	Adding structure to the event space	10
1.5.7	Hidden Events	10
1.6	the most general model	11
Chapter 2	Estimation	12
2.1	Empirical Counting Measure	12
2.2	Maximum Likelihood Estimation	12
2.3	EM	12
2.3.1	Information Theory and EM	12
2.3.2	EM for Hawkes Processes	13
2.3.3	Generalised Fixed Point Methods	13
2.3.4	Method of Scoring (Newton's Method)	13
2.4	Uncertainty quantification for MLE	13
2.5	Inference for the Mark Process	13
2.6	Computational Concerns	13
2.7	Bayesian Approach	13
2.8	Model Selection	13

Chapter 3	Generative Sampling	14
3.1	Simulation Methods	14
3.2	Simulation Study of Estimation Methods	14
3.3	Impulse Response Function	14
Chapter 4	Application to KOSPI/SPY Data	15
4.0.1	Clustering Ratio	16
Chapter 5	Conclusion	17
Chapter 6	Appendix: Foundations of Probabilistic Models	18
6.1	Measure Theory Fundamentals	18
6.2	Probability Spaces	19
6.2.1	Conditionalisation	20
6.3	Density of a Measure	20
6.4	Stochastic Process Fundamentals	21
References		22

CHAPTER 1

Introduction

On a single day in 2023, the US stock market saw an average of around \$500 billion dollars worth of shares traded on various exchanges and other venues, exceeding the annual GDP of a typical European country [7]. Modern securities markets facilitate the exchange of shares and other financial assets at extremely high frequency, as a result of aggressive investment in specialised networking hardware, custom-made computer architectures, and high-throughput machine learning systems. In facilitating capital flows for the global economy, financial markets have at the same time managed to claim an increasing fraction of resources and attention, with economic consequences that are not yet fully understood [1].

Despite many mysteries and open questions about the origins and dynamics of market phenomena, the financial sector itself has readily adapted to the increasing scale and complexity of the markets in which it operates. The practical design of exchange rules and trading systems has in large part been an empirical endeavour on the part of market participants, operators, and regulators. Many phenomena have been observed to emerge in an apparently decentralised fashion from the application of exploitative heuristics and predictive algorithms that interact with exchanges and aggregate to produce desirable outcomes. While users of trading strategies aim to maintain acceptable risk levels while generating profits over the long term, market operators and regulators are tasked with the design of incentive mechanisms that exploit this self-interested behaviour to improve market outcomes, including reduced transaction costs, fast and accurate incorporation of external information (such as economic news or earnings reports), and adherence to various concepts of fairness, propriety, and legality.

In this thesis, I describe and extend prior work from the empirical market microstructure literature, making extensive use of the state-dependent Hawkes process model for event arrivals. I begin with a conceptual overview of the trading mechanism, and formalise the mathematical tools that will be used to construct and describe variations on the basic Hawkes process model. I continue with a summary of existing literature on point processes as applied to market data, including both mathematical foundations and empirical findings. Next, I explore techniques to reduce the computational burden of parametric inference for point processes on large datasets. Finally, empirical applications and findings are discussed, including applications of generative modeling to a variety of open problems in market microstructure.

1.1 Limit Order Books

Intraday trading allows participants to respond to exogenous news and endogenous market events in a manner that maintains acceptable levels of risk and generates profits over the long term. This activity has a significant influence on the formation of market prices and plays a key role in reducing transaction costs while increasing the speed at which large institutions can control their exposure to various financial risks and opportunities.

Securities exchanges facilitate automated matching of buyers and sellers at prices favourable to both. Understanding the dynamics of this exchange process at a high degree of resolution can provide insights into the design of automated *matching engines* that produce desirable market behaviour, as well as insights into the design of *trading strategies* that exploit the dynamics of the exchange process to generate profits.

A matching engine is tasked with receiving and acting on various messages from market participants indicating their intent to buy or sell a particular security with particular conditions. As a result of this process, trades may be formed that match buyers and sellers at a mutually agreeable price and quantity. Trade reports are broadcast to relevant participants and may be used for the purposes of risk management and forecasting, as well as for the ultimate transfer of the assets that have been traded (which often occurs after trading hours).

A typical matching engine permits two kinds of incoming messages, known as *order insertion* and *order cancellation*, and maintains an internal state consisting of a single data structure, known as a *limit order book*.

An order insertion message indicates a participant's willingness to buy (or sell) some quantity of a particular asset at or below (respectively, above) a particular price, and results in the addition of an *order* to the limit order book \mathcal{L} .

Conversely, an order cancellation message results in the removal of a particular order from the limit order book, either in part (by reducing the remaining volume associated with the order) or in full (by removing the order entirely from the book). It might be good to include a statistic about roughly how large cancellation rates are, to highlight the importance of this message type.

Formally, a limit order book can be defined as a set of tuples ("orders") of the form

$$(\text{side}, \text{price}, \text{time}, \text{size}) = (s, p, t, q) \in \{-1, 1\} \times P \times T \times Q.$$

Each published order represents an intention to buy (or sell) some quantity of an asset at a maximum (respectively, minimum) price, as illustrated in figure 1.1.

$$\left(\underbrace{-1}_{\substack{+1 \text{ for a buy order} \\ -1 \text{ for a sell order}}}, \underbrace{84.1}_{\substack{\text{The least favourable price} \\ \text{(maximum for buy,} \\ \text{minimum for sell)} \\ \text{at which the order can trade}}}, \underbrace{9 : 52}_{\substack{\text{The time at which the} \\ \text{order was first published}}}, \underbrace{2}_{\substack{\text{The maximum quantity} \\ \text{of the product that will} \\ \text{be traded with this order}}} \right)$$

Figure 1.1: Components of an example limit order

1.2 The Matching Algorithm

1.2.1 The Bid and Ask

When an order insertion message is received, the exchange will attempt to form trades with the existing orders in the book such that as much volume as possible is matched. Any volume that cannot be matched will be added to the orders already in the book.

As a result of this matching, the total volume posted to the book may be depleted over time, even in the absence of cancellations. Any time a buy order has a price equal to or greater than that of a sell order, volume will be matched and a trade will occur. Consequently, the most competitive buy price (known as the *bid*) will always be less than the most competitive sell price (known as the *ask*), i.e.

$$\text{bid}_{\mathcal{L}} = \max_{\{p:(1,p,t,q) \in \mathcal{L}\}} p \leq \min_{\{p:(-1,p,t,q) \in \mathcal{L}\}} p = \text{ask}_{\mathcal{L}}.$$

An excessively high bid price will be depleted by traders seeking to sell the product at a premium to its true value. Conversely, an excessively low ask price will be pushed up by participants hoping to buy at a discount. It is therefore common to regard the bid and ask as lower and upper bounds respectively on the consensus fair price of the product.

Motivated by this, the *midprice* is a naive point estimate for the consensus fair price, defined by

$$\text{mid}_{\mathcal{L}} = \frac{1}{2} (\text{bid}_{\mathcal{L}} + \text{ask}_{\mathcal{L}}).$$

Many other proxies for the consensus fair price exist, and it is common to use these as prediction targets in the construction of trading signals. **I might discuss them later**

1.2.2 Liquidity

At any point in time, the contents of the limit order book represent trading opportunities presented to all market participants. The abundance of these opportunities, also known as *liquidity*, represents a positive externality insofar as it allows

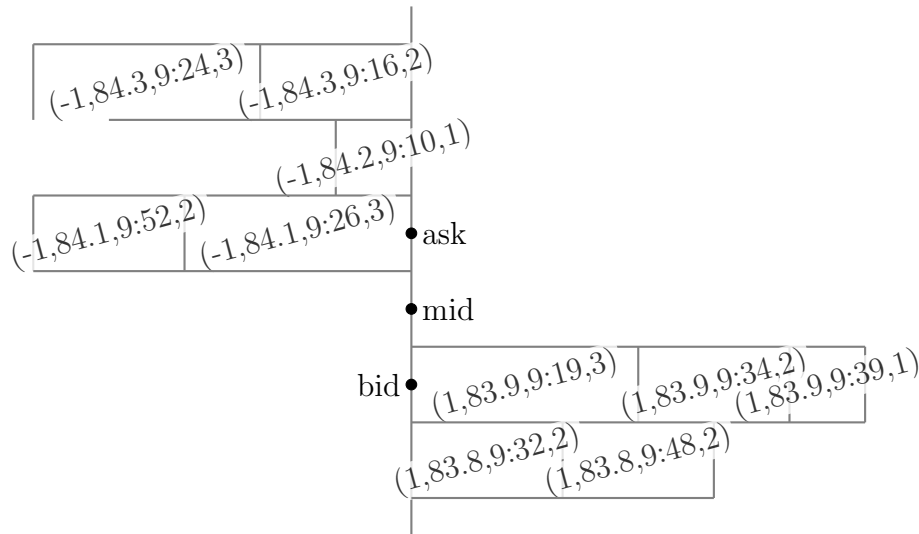


Figure 1.2: An example order book, arranged by order price and time

impatient traders (*liquidity takers*) to buy or sell products precisely under those circumstances where it is favourable to them. Conversely, order publishers (*liquidity providers*) must adhere to the terms of trades formed against a posted order, regardless of whether it is in their interests.

One common measure for market liquidity is the *bid-ask spread*, defined as

$$\text{spread}_{\mathcal{L}} = \text{ask}_{\mathcal{L}} - \text{bid}_{\mathcal{L}}.$$

Notably, if $\text{mid}_{\mathcal{L}}$ is taken to represent the fair value of the product, the *half-spread*, i.e. $\frac{1}{2}\text{spread}_{\mathcal{L}}$, represents the premium paid to liquidity providers by participants trading against the posted bid or ask. There are many other components and measures of liquidity **which I may discuss later**.

1.2.3 Price-Time Priority

Explain

1.3 Queueing Models of the Order Book

Queue depletions are related to price changes [3]

Evolution of the limit order book over time \rightarrow motivate the emphasis on modeling arrival times

Price evolution as a jump process (what is a jump process)

Importance of arrival time modeling

<https://www.amazon.com/Point-Processes-Queues-Martingale-Statistics/dp/0387905>

1.4 Overview of Point Processes

With the goal of modelling the arrival time process, I will now provide an introductory overview of point processes. In the next section, I will highlight some key variants described in existing literature that are relevant to order book modeling.

This section will make extensive use of theoretical concepts described in the appendix **Make sure to update this if there ends up being more than one appendix.** For an overview of measure theory, probability spaces, and stochastic processes, please refer to the relevant sections.

1.4.1 Definition

Given a probability space $(\Omega, \Sigma, \mathbb{P})$, and a measurable space (T, Σ_T) representing times, a point process is any increasing sequence of random times

$$\mathcal{T} : \Omega \rightarrow T^{\mathbb{N}},$$

meaning that $\mathcal{T}(\omega)_n$ is increasing in n for any $\omega \in \Omega$.

Concretely, T may be chosen to be $\mathbb{R}_{\geq 0}$, and equipped with the Borel σ -algebra $B(\mathbb{R}_{\geq 0})$.

The times in the point process are often referred to as *event times*, with the implication that the sequence represents the times at which some event of interest occurs (for instance, the arrival of messages sent to a matching engine).

For any point process, there exists a corresponding càdlàg stochastic process $N_{\mathcal{T}} : T \times \Omega \rightarrow \mathbb{N}$ known as the *counting process*, that gives the number of events having occurred before or at a given time. This is defined as

$$N_{\mathcal{T}}(t, \omega) = |\{i \in \mathbb{N} | \mathcal{T}(\omega)_i \leq t\}|.$$

Overlay barcode plot for trade times with corresponding counting process

All **(or perhaps many)** of the point processes considered in this thesis will additionally be adapted with respect to some filtration \mathcal{F} indexed by T . **In what sense?**

Furthermore, I will only consider *nonexplosive point processes*, defined as those point processes \mathcal{T} for which

$$\lim_{n \rightarrow \infty} \mathcal{T}(\omega)_n = \infty, \mathbb{P}\text{-a.s.}$$

For any bounded interval of time (t_{\min}, t_{\max}) , a nonexplosive point process will almost surely contain only a finite set of times in that interval, i.e.

$$(t_{\min}, t_{\max}) \cap \mathcal{T}(\omega) \text{ finite, } \mathbb{P}\text{-a.s.}$$

1.4.2 The Counting Measure

For any $\omega \in \Omega$, we can define the *counting measure* Λ_{ω} of the point process as

$$\Lambda_{\omega} : \Sigma_T \rightarrow \mathbb{N}$$

$$\Lambda_{\omega}(S) := |S \cap \mathcal{T}(\omega)|.$$

For any finite or countable collection of disjoint sets $A_n \in \Sigma$, we have

$$\Lambda_\omega \left(\bigcup_n A_n \right) = \left| \bigcup_n (A_n \cap \mathcal{T}(\omega)) \right|.$$

Is this a measure on (T, Σ_T) ? Since subsets of disjoint sets are also disjoint, the terms in the union on the right-hand side will be disjoint, and so

$$\Lambda_\omega \left(\bigcup_n A_n \right) = \sum_n |A_n \cap \mathcal{T}(\omega)| = \sum_n \Lambda_\omega(A_n).$$

So this is indeed a measure.

1.4.3 The Expectation Measure

Taking the expectation of the counting measure with respect to some measure \mathbb{P} on a measurable space (Ω, Σ) gives the *expectation measure*,

$$\Lambda_{\mathbb{P}}(S) := \mathbb{E}_{\mathbb{P}} [\Lambda_\omega(S)] = \int_{\Omega} \Lambda_\omega(S) d\mathbb{P}(\omega).$$

Is this a measure on (T, Σ_T) ? Since $\Lambda_\omega(S)$ is a non-negative function of ω , it follows that

$$\Lambda_{\mathbb{P}} : \Sigma_T \rightarrow \mathbb{R}_{\geq 0} \cup \{\infty\}.$$

Furthermore, for any finite or countable sequence of $A_n \in B(\mathbb{R})$, we have

$$\Lambda_{\mathbb{P}} \left(\bigcup_n A_n \right) = \int_{\Omega} \Lambda_\omega \left(\bigcup_n A_n \right) d\mathbb{P}(\omega) = \int_{\Omega} \sum_n \Lambda_\omega(A_n) d\mathbb{P}(\omega),$$

by the countable additivity of Λ_ω .

To handle the sum inside the integral, we can write each term as a sum of indicator functions:

$$\Lambda_\omega(A_n) = |A_n \cap \mathcal{T}(\omega)| = \sum_i 1_{A_n}(\mathcal{T}(\omega)_i).$$

Then, since A_n are measurable sets, the indicator functions are each measurable functions of ω . Because finite sums and pointwise limits of measurable functions are measurable, we have that $\Lambda_\omega(A_n)$ is measurable. Finally, since each term in the sum over n is a nonnegative measurable function, the integral commutes with the sum, and hence $\Lambda_{\mathbb{P}}$ is countably additive. Therefore the expectation measure is also a valid measure.

1.4.4 Intensity of a Point Process

For a point process \mathcal{T} , the counting measure of a half-open interval $(t, t + \epsilon]$ can be written in terms of the counting process $N_{\mathcal{T}}$ in the form

$$\Lambda_\omega((t, t + \epsilon]) = N_{\mathcal{T}}(t + \epsilon, \omega) - N_{\mathcal{T}}(t, \omega).$$

Integrating out ω with respect to \mathbb{P} then gives us the identity

$$\Lambda_{\mathbb{P}}((t, t + \epsilon]) = \mathbb{E}_{\mathbb{P}}[N_{\mathcal{T}}(t + \epsilon)] - \mathbb{E}_{\mathbb{P}}[N_{\mathcal{T}}(t)],$$

allowing us to write the expectation measure in terms of a finite difference of the first moment of $N_{\mathcal{T}}$.

If $\mathbb{E}_{\mathbb{P}}[N_{\mathcal{T}}(t)]$ is right-differentiable at t , we can further say that

$$\lim_{\epsilon \rightarrow 0} \frac{\Lambda_{\mathbb{P}}((t, t + \epsilon])}{\epsilon} = \lim_{\epsilon \rightarrow 0} \frac{\mathbb{E}_{\mathbb{P}}[N_{\mathcal{T}}(t + \epsilon)] - \mathbb{E}_{\mathbb{P}}[N_{\mathcal{T}}(t)]}{\epsilon} = \frac{d\mathbb{E}_{\mathbb{P}}[N_{\mathcal{T}}(t)]}{dt}$$

for sufficiently small ϵ , where the last term is the right-derivative. For brevity, I will write

$$\lambda_{\mathbb{P}}(t) := \frac{d\mathbb{E}_{\mathbb{P}}[N_{\mathcal{T}}(t)]}{dt}.$$

Assuming further that $\mathbb{E}_{\mathbb{P}}[N_{\mathcal{T}}(t)]$ is differentiable on some open interval $(t, t + \epsilon')$,

Prove this is Radon-Nikodym derivative of $\Lambda_{\mathbb{P}}$ in the open interval

It then follows from the fundamental theorem of calculus that

$$\int_t^{t+\epsilon} \lambda_{\mathbb{P}}(s) ds = \lambda_{\mathbb{P}}(t) \cdot \epsilon + o(\epsilon), \quad \epsilon \rightarrow 0.$$

So **since $\lambda_{\mathbb{P}}$ is a radon-nikodym derivative** we can write **Why?**

$$\Lambda_{\mathbb{P}}((t, t + \epsilon]) = \lambda_{\mathbb{P}}(t) \cdot \epsilon + o(\epsilon), \quad \epsilon \rightarrow 0$$

Therefore the expected number of events occurring in a small interval $(t, t + \epsilon]$ is approximately proportional to its length.

It is common to refer to $\lambda_{\mathbb{P}}(t)$ as the *arrival rate* of the point process at t , since it represents the expected number of events occurring per unit time.

If we instead take the expectations above conditioned on \mathcal{F}_t , we can define a stochastic process

$$\lambda(t, \omega) := \lambda_{\mathbb{P}|\mathcal{F}_t}(t, \omega),$$

known as the *intensity* of the point process. This represents our best estimate of the number of events that will arrive in the next ϵ units of time, for very small ϵ , based on all the information contained in the filtration up to t .

Existence and uniqueness shown in <https://projecteuclid.org/journals/annals-of-probability/volume-24/issue-3/Stability-of-nonlinear-Hawkes-processes/10.1214/aop/10657251> under restrictions on the kernel

1.4.5 Residuals

Define residuals, explain why they're exponentially distributed

1.4.6 Marked and Multivariate Point Processes

A point process T may be *marked*, in which case it is associated with one or more random sequences of marks

$$\mathcal{X} : \Omega \rightarrow X^{\mathbb{N}},$$

drawn from a set X , that represent additional information about each event. These may be adapted to the filtration \mathcal{F} **In what sense?**

A common special case of this is when X is partitioned into a finite number of *event types*, in which case we may refer to \mathcal{T} as a *multivariate point process*.

Each part E of the partition then has an associated point process

$$\mathcal{T}_E = \mathcal{T}_{\{n \in \mathbb{N} : \mathcal{X}_n \in E\}},$$

formed by the subsequence of times where the corresponding element of \mathcal{X} is in E . These point processes will then have their own counting functions, intensities, and other characteristics. It is common to arrange these in vector form, with one entry for each event type.

On the other hand, given a collection of point processes adapted with respect to a common filtration, we can form a single marked point process by interleaving the sequences of event times. The counting process, counting measure, expectation measure, and intensity of the combined process will be the sum of those for the individual point processes.

Barcode plot and counting process for buys and sells, and the combined process

1.5 Point Process Models of Book Updates

For each extension: - show some example realisations for various parameter choices: barcode and intensity plot

1.5.1 Poisson Processes

Probably the simplest modeling assumption we can make is to assume that the intensity of the process is a constant $\nu \in \mathbb{R}$.

A Poisson process is a point process for which the following two properties hold.

1. For any finite collection of disjoint measurable sets $A_n \in \Sigma_T$, their counting measures $\Lambda_\omega(A_n)$ are independent random variables.
2. For any measurable set $A \in \Sigma_T$, we have

$$\Lambda_{\mathbb{P}}(A) = \nu \mu_{\text{Lebesgue}}(A) \propto \mu_{\text{Lebesgue}}(A).$$

Why is this the same as saying that the intensity is constant?

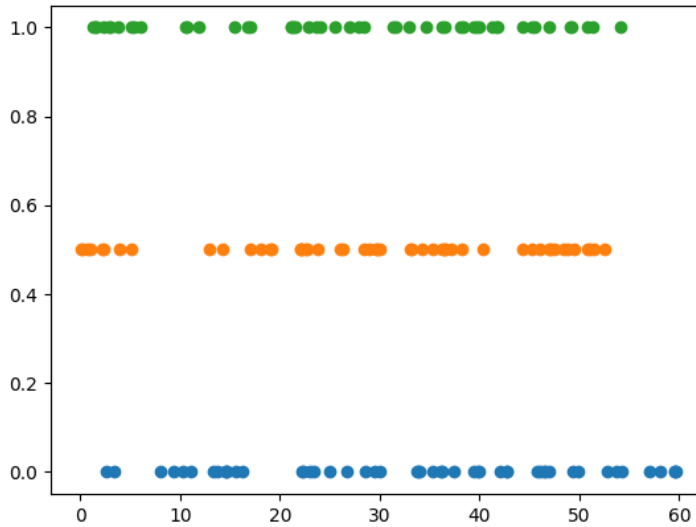
In the multivariate case with n event types, we have a vector $\nu \in \mathbb{R}_{>0}^n$ containing the constant intensities of each individual point process.

This definition might require a reference.

Is there a proof of existence and nonexplosiveness?

Explain why the times between events are exponentially distributed. Can use the linearity of residuals to prove this succinctly, since residuals are exponentially distributed.

There is a link here to the sante fe order book model, described in the Bouchaud source, which is an example of the application of poisson arrivals.



I think I should change this to be separate plots. Can show the intensity function for each; e.g. for inhomogenous it's the same for every realisation, but for hawkes process the intensity function is itself random.

1.5.2 Inhomogeneous Poisson Process

An inhomogeneous Poisson process is a generalisation of the Poisson process for which the second condition is replaced by the weaker requirement that every measurable set $A \in \Sigma_T$ with finite Lebesgue measure $\mu_{\text{Lebesgue}}(A)$ has a finite expectation measure. Is this weaker version even needed, or is it sufficient to just drop the condition entirely?

As a result, although the intensity $\lambda(t, \omega)$ is constant in ω (i.e. deterministic) how do we know it is?, it can vary as a function of t . This is useful for modeling deterministic seasonality in the point process, such as having more events near the start/end of each trading session or near known events such as news releases.

Show U-shape from empirical data Show U-shape from quadratic IPP

1.5.3 Autoregressive Point Processes

i.e. Hawkes Processes Intuitively, self-excitation or self-inhibition; mutually exciting/inhibiting

Original hawkes process paper <https://academic.oup.com/biomet/article-abstract/58/1/83/224>

immigration/birth interpretation branching matrix & complete data likelihood

Long-run mean

$$\bar{\lambda} = \lim_{t \rightarrow \infty} \mathbb{E}[\lambda(t)]$$

$$\bar{\lambda} = \nu + \bar{\lambda} \int_0^\infty k(t) dt$$

Stability requires $\int_0^\infty k(t) dt < 1$. Not sure if sufficient condition.

Multidimensional:

$$\bar{\lambda} = \nu + K\bar{\lambda}$$

$$\bar{\lambda} = (I - K)^{-1}\nu$$

Requires $|\det K| < 1$ (why? and what about if there are directions orthogonal to ν ?) **Define K**

According to <https://ieeexplore.ieee.org/document/7416001> <https://arxiv.org/pdf/> Hawkes processes are fully determined by the first two moments of the intensity function. Proof in the paper.

Special case of $\beta = 0$ may be worth including in the final kernel. (Integrated point process?)

“Swishchuk and Huffman (2020) construct a compound Hawkes process” modeling price changes with a jump process where jump sizes are a markov chain. Jump process is controlled by one-dimensional point process.

1.5.4 State Dependence

Vinovskaya proposes regime-switching independent of the point process Moriaru-Patrichi and Pakkanen generalise this by coupling state transitions with Moriaru-Patrichi and Pakkanen find that conditioning on the state of the order book improves the fit of the model. (How?)

They consider conditioning on only the state of the triggering event (their 2.3), as well as conditioning on both the state of the triggering event and the current state (their 4.2).

The first of these:

$$\tilde{\lambda}_{ex}(t) = \phi_e(X(t), x) \left(\nu_e + \sum_{e' \in \mathcal{E}, x' \in \mathcal{X}} \int_{[0, t)} k_{e', e}(t - s, x') d\tilde{N}_{e'x'}(s) \right), t \geq 0, e \in \mathcal{E}, x \in \mathcal{X}.$$

The second:

$$\tilde{\lambda}_{ex}(t) = \nu_{ex} + \sum_{e' \in \mathcal{E}, x' \in \mathcal{X}} \int_{[0, t)} k_{e'x'ex}(t - s) d\tilde{N}_{e'x'}(s), t \geq 0, e \in \mathcal{E}, x \in \mathcal{X}.$$

1. markov switching 2. linear state space model / kalman filter (we'll see if i get around to figuring out how to work this)

1.5.5 Regression on marks

Multiply the kernel by $\exp(X \cdot \beta)$ where X is a vector of mark variables and β are coefficients. **transfer functions should be separable, see [6.25 Definition] in <https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/151886/eth-1112-02.pdf>**

1.5.6 Adding structure to the event space

For a finite number of event types (e.g. 10 book levels) this just corresponds to constraining the coefficient matrices.

For a continuous event space it is a bit more complicated

Another way to look at this is just having marks but the distribution of the marks is not constant - this works too. This approach is nice because it admits Kalman filtering. (Is it possible to mix Kalman filtering in event time, volume time, and wall time? Maybe better to just use wall time though.)

1.5.7 Hidden Events

Seems relevant <https://www.cambridge.org/core/journals/advances-in-applied-probability/article/elementary-derivation-of-moments-of-hawkes-processes/79A7355542F08087C8AE828C664A304>

Hawkes Process Inference with Missing Data <https://aaai.org/papers/12116-hawkes-process-inference-with-missing-data/> makes use of monte carlo EM algorithm <https://www.jstor.org/stable/2346222>

Hawkes processes with hidden marks <https://www.tandfonline.com/doi/full/10.1080/1351847X.2017.1351847>

Meta-orders. Kernel is product of an ordinary kernel and an EMA on the meta-order hidden events. See model in [3] 10.4.3.

Could also be used to correct for the quasi-EM approximation.

1.6 the most general model

Here for ideas/reference but I dont think this section will be in the final report
State-dependent multi-kernel multivariate hawkes process with increasing kernel components and possible inhibition.

Nonparametric estimate of nu, sim study for u-shaped true nu. Probably splines, maybe kde if its fast enough. Compare to quadratic regression

The marks evolve according to state-space model involving various model variables, but have no causal impact on the point process (though obviously there is state-dependence, and the state may include the marks).

The state might be generalised to a kalman-filter-ish model [15]

$$IRF_{e,e',k}(s, t) = \alpha_{e,e',k} e^{X(s) \cdot c_{e,e',k}} \exp(-\beta_{e,e',k}(t - s))$$

$$\lambda_{e,e',k}(t, \omega) = \int_0^t IRF_{e,e',k}(s, t) d\Lambda_{e',\omega}(s)$$

$$\lambda_e(t, \omega) = \sum_{e'} \sum_k \lambda_{e,e',k}(t, \omega) + \text{spline intensity}(t)$$

CHAPTER 2

Estimation

To enable analysis of empirical order book data, I will now explain the principles involved in parameter estimation for point process models.

2.1 Empirical Counting Measure

Define $\lambda_{e,(T,E)}(t)$ and $\Lambda_{e,(T,E)}(t)$ for fully observed point processes for (T, E)

2.2 Maximum Likelihood Estimation

Given a parametric family of marked point processes,

$$\{(\mathcal{T}_\theta, \mathcal{X}) : \theta \in \Theta\}$$

Derive the likelihood function

For a particular observation (T, E) , we have

$$\ell((T, E); \theta) = \sum_e \left(\int_0^T \log(\lambda_{e,\theta,(T,E)}(t)) d\Lambda_{e,(T,E)}(t) - \int_0^T \lambda_{e,\theta,(T,E)}(t) dt \right)$$

What is a likelihood function

Clearly λ_e must be positive at the event times.

At stationary points, loglikelihood has zero derivatives.

$$\frac{\partial \ell((T, E); \theta)}{\partial \theta} = \sum_e \left(\int_0^T \frac{\partial \lambda_{e,\theta,(T,E)}(t)}{\partial \theta} \cdot \frac{1}{\lambda_{e,\theta,(T,E)}(t)} d\Lambda_{e,(T,E)}(t) - \int_0^T \frac{\partial \lambda_{e,\theta,(T,E)}(t)}{\partial \theta} dt \right)$$

Explain why no explicit solution

Observe that if intensity is a sum of multiple kernels it separates into the EM form

Introduce the branching matrix

Motivate iterative methods: find an iteration whose fixed points are stationary points. Is there a way to show that this converges?

2.3 EM

2.3.1 Information Theory and EM

Can cite original dempster paper? As well as the one that corrects the errors in that paper

2.3.2 EM for Hawkes Processes

Branching matrix vs exo/endo classification. Usefulness for multikernel and other limitations of gradient based MLE.

$$B_{e,e',k}(s,t) = \frac{IRF_{e,e',k}(s,t)}{\lambda_e(t)}$$

$$\sum_{e'} \sum_k \int_0^t B_{e,e',k}(s,t) d\Lambda_{e',\omega}(s) = 1$$

Branching matrix:

$$B_{i,j,k} = \frac{IRF_{e_j,e_i,k}(t_i,t_j)}{\lambda_{e_j}(t_j)}$$

- Visualisation of branching matrix. How does it evolve throughout the fitting procedure?
- Quasi EM approximation - Closed form constant time M step - Negative probabilities. Conditions for kernel nonnegativity? (Probably not tractable)
- Complex probabilities? For periodic/wavelet kernels

2.3.3 Generalised Fixed Point Methods

Is it a contraction mapping?

2.3.4 Method of Scoring (Newton's Method)

2.4 Uncertainty quantification for MLE

Asymptotic normality

Parametric Bootstrap

2.5 Inference for the Mark Process

Kalman filter? Ways to couple this with the point process? (Monte carlo EM?)

2.6 Computational Concerns

Sensor fusion for parallelisation

Momentum (analyse autocorrelation of parameter changes throughout the learning process)

Exploiting sparsity [14]

2.7 Bayesian Approach

- Uncertainty quantification - Less likely to get stuck in local optima - Jeffreys prior: $\sqrt{\det \mathcal{I}}$

2.8 Model Selection

Performance of information criteria for selection of Hawkes process models of financial data <https://www.tandfonline.com/doi/full/10.1080/14697688.2017.1403140>

CHAPTER 3

Generative Sampling

3.1 Simulation Methods

Immigration-birth interpretation [13] (aka Watson-Galton models)

Pseudocode

Ogata thinning

Pseudocode

Do these have different time complexity? Memory complexity?

3.2 Simulation Study of Estimation Methods

Convergence analysis for simple models (eg univariate)

3.3 Impulse Response Function

Causal analysis, price impact of orders.

What does inserting a single exogenous event do to the order book, price, etc? Are there analytical formulas for this?

[19] [20] sources that use Hawkes process models to simulate an order book for analysis purposes

CHAPTER 4

Application to KOSPI/SPY Data

Hawkes processes and their applications to finance: a review <https://www.tandfonline.com/doi/https://www.tandfonline.com/journals/rej20/collections/Hawkes-Processes-in-Finance>

- Basic poisson model
- Univariate hawkes
- Multivariate hawkes - what are the event types?
- State-dependent hawkes - which state variables are relevant?
- Full model
- Replicate findings from [13]
- Hidden events (& events on different exchanges) - either poisson distributed or more complex
- Modeling changes in the entire order book
- Market impact (are there any datasets on this? square-root law, other common findings. power law impact for hawkes processes is explicitly studied here <https://arxiv.org/pdf/1805.07134>)
- Optimal execution - VWAP, TWAP, Almgren-Chriss
- Midprice change prediction/explanation - explicit formula or simulation?
- Realised volatility prediction
- Correlated products (with low beta, preferably - or see what is done in literature studies of correlated products)
- Options (if I can get data) - would give lots of (nonlinearly) correlated products. Can estimate the correlation between products at any point in time using factor loadings & historical factor correlations. Here is one source: https://www.nber.org/system/files/working_papers/w29369/w29369.pdf. *Options pricing and execution in a Hawkes process*
<https://arxiv.org/abs/1602.03043> *Closed form for optimal execution of signals (including impact)*
- This paper arxiv.org/pdf/2401.11495 shows a functional limit theorem for hawkes processes behaving as integrated CIR processes. These are a popular volatility model so this makes sense!

<https://quant.stackexchange.com/questions/59593/what-are-some-currently-open-problems-in-the-theory-of-hawkes-processes>

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4844711 *Option Pricing Using Hawkes Processes*

A slightly depressing jump model: intraday volatility pattern simulation

<https://www.tandfonline.com/doi/full/10.1080/14697688.2017.1403139> Implementation and evaluation of the Heston-Queue-Hawkes option pricing model <https://uu.diva-portal.org/showFileFullText?id=11111111>

4.0.1 *Clustering Ratio*

Variance of interevent time divided by expectation. Bouchaud p 165. Should this section be elsewhere? Does this mean something for residuals too?

Point process control https://pages.stern.nyu.edu/~rcaldent/courses/B60.4308_files/Process.pdf

Two choices for handling inhomogeneity: - Reparameterise to 'business time' in preprocessing, as recommended by [3] in 9.3.1 - Splines/polynomials etc - Could also combine both

Databento info: CME: ES MES OPRA: SPX - Index options SPY - Index ETF options <https://databento.medium.com/getting-futures-tick-sizes-and-notional-t>

CHAPTER 5

Conclusion

This is the conclusion

CHAPTER 6

Appendix: Foundations of Probabilistic Models

In order to describe precisely the various models explored in this thesis, it is necessary to introduce some key mathematical concepts that form the basis for model specifications and related derivations. In this appendix, I cover the basics of measure, probability, and stochastic processes.

6.1 Measure Theory Fundamentals

In order to formalise the concept of a stochastic process, it is necessary to introduce the concept of a measure space.

For any set X , we can construct the power set

$$\mathcal{P}(X) = \{S : S \subseteq X\}$$

which is a new set that contains as its elements every subset S of X . We say that a subset Σ of $\mathcal{P}(X)$ is a σ -*algebra* on X if and only if it satisfies the following three properties:

1. Containment of the full space, i.e.

$$X \in \Sigma.$$

2. Closure under complements, i.e.

$$\forall S \in \Sigma, X \setminus S \in \Sigma.$$

3. Closure under countable union, i.e.

$$\forall \{A_n\}_{n=0}^{\infty} \in \Sigma^{\mathbb{N}}, \quad \bigcup_{n=0}^{\infty} A_n \in \Sigma.$$

Elements of Σ are known as *measurable sets*. A common example of a σ -algebra is the Borel σ -algebra $B(X)$ of a topological space X (e.g. \mathbb{R}), defined as the smallest σ -algebra such that every open set is measurable.

We then say that a *measure* on (X, Σ) is any function $\mu : \Sigma \rightarrow \mathbb{R}_{\geq 0} \cup \{\infty\}$ that is *countably additive*, meaning that for any finite or countable sequence of disjoint sets $A_n \in \Sigma$, we have

$$\mu \left(\bigcup_n A_n \right) = \sum_n \mu(A_n).$$

Assuming that at least one set $S \in \Sigma$ has finite measure, we then have

$$\mu(S) = \mu(S \cap \emptyset) = \mu(S) + \mu(\emptyset) \Rightarrow \mu(\emptyset) = 0.$$

We refer to the combined triple (X, Σ, μ) as a *measure space*.

Informally, a measure formalises intuitions about the size, mass, or significance of a set of points. For instance, a set $S \subseteq X$ with measure zero is known as a *null set*, and a property that holds only for points in a null set is said to be true *almost nowhere* in X . Conversely, a property that holds for every point except those in a null set is said to be true *almost everywhere* (a.e.) in X . In this way, a measure quantifies how significant or negligible the exceptions to a heuristic principle may be.

Similarly, familiar concepts of *length*, *area* and *volume* are all formalised by a family of translation-invariant measures on \mathbb{R}^n , known as the *n-dimensional Lebesgue measures*.

The concept of a measure is foundational to the definition of the Lebesgue integral. Integration of a function f over a measurable set S with respect to a measure μ is written as

$$\int_{t \in S} f(t) d\mu(t),$$

while integration with respect to the Lebesgue measure will often be written simply as

$$\int_S f(t) dt.$$

I will not cover the definition of the Lebesgue measure or Lebesgue integral here, but they can be found in most textbooks on measure theory.

6.2 Probability Spaces

In the special case where $\mu(X) = 1$, we refer to μ as a *probability measure*, and to X as a *sample space*. Correspondingly, the term *almost everywhere* is replaced with *almost surely* (a.s.), indicating a property that holds on a set of points with probability measure one.

A function from one σ -algebra to another is called *measurable* if and only if the preimage of any measurable set in the codomain is a measurable set in the domain. A measurable function whose domain is a sample space Ω equipped with a probability measure \mathbb{P} is known as a *random variable*.

Measurable sets in a sample space are often referred to as *events*, and the measure of such a set is called the probability of the event. For instance, the set S of points $\omega \in \Omega$ for which a random variable $X : \Omega \rightarrow \mathbb{R}$ satisfies a particular property $P : \mathbb{R} \rightarrow \{\text{True}, \text{False}\}$ will have measure equal to the probability of that property being true.¹

¹The predicate P must be measurable.

A more general concept is *expectation*. For a real-valued random variable $X : \Omega \rightarrow \mathbb{R}$, we define the expectation of X with respect to \mathbb{P} to be the linear functional

$$\mathbb{E}_{\mathbb{P}}[X(\omega)] := \int_{\omega \in \Omega} X(\omega) d\mathbb{P}(\omega).$$

For binary-valued random variables $X : \Omega \rightarrow \{0, 1\}$, we have the identity

$$\mathbb{E}_{\mathbb{P}}[X(\omega)] = \mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\}).$$

6.2.1 Conditionalisation

By defining ‘events’ as Σ -measurable subsets of Ω , we allow knowledge of an event to inform us about which possible values of the hidden ω could have produced the result we actually observe. In this way, the knowledge that ω has produced one observed event allows us to draw probabilistic conclusions about the occurrence of a different, related event. For instance, learning that a die has rolled an even number tells us that it cannot possibly have rolled a 3, and makes the proposition that a 2 has been rolled a more reasonable guess than otherwise.

In order to capture the relationships between events, and to make systematic inferences about hidden events from observed ones, it is necessary to describe mathematical rules for *conditionalisation*.

Given two σ -algebras Σ_1, Σ_2 on ω , we say that Σ_1 is a *sub- σ -algebra* of Σ_2 if and only if $\Sigma_1 \subseteq \Sigma_2$. In the case where $\Sigma_1 \neq \Sigma_2$, Σ_1 is said to be *coarser* than Σ_2 , in the sense that Σ_2 contains events that cannot be expressed as Σ_1 -measurable sets. Knowing which Σ_2 -events the hidden ω falls under can therefore give us more information about the exact value of ω .

One common example of a sub- σ -algebra arises by considering the set

$$\{X^{-1}(A) : A \in \mathbb{R}\} \subseteq \mathcal{P}(\Omega),$$

consisting of all the preimages of Borel-measurable sets under a random variable $X : \Omega \rightarrow \mathbb{R}$. This is known as the σ -algebra generated by X .

expectation conditional on sigma algebra <https://math.stackexchange.com/questions/1111111/expectation-conditional-on-sigma-algebra>
 expectation conditional on an event

probability is expectation of indicator function (notation for indicator function is 1_A)

random measure

What is independence independence of RVs if and only if independence of generated sigma algebras

Conditional probability

6.3 Density of a Measure

Absolute continuity: $\mu_1 \ll \mu_2$ (μ_1 is dominated by μ_2) if and only if $\mu_2(A) = 0 \Rightarrow \mu_1(A) = 0$ for every measurable set A . Can also say that one measure is absolutely continuous wrt another on a subset of the measure space Radon-Nikodym Theorem

6.4 Stochastic Process Fundamentals

What is a filtration

What is a stochastic process

What is a realisation of a stochastic process

What is a cadlag The order book is a cadlag over time. Also the counting process is a cadlag.

What is a martingale - do I need this?

References

- [1] Palley, T. I., *Financialization: What It Is and Why It Matters*, Working Paper No. 525, The Levy Economics Institute, December 2007. Paper presented at the conference on "Finance-led Capitalism? Macroeconomic Effects of Changes in the Financial Sector," sponsored by the Hans Boeckler Foundation, Berlin, Germany, October 26{27, 2007. Available at: https://www.levyinstitute.org/pubs/wp_525.pdf
- [2] Balakrishnan, S., Wainwright, M. J., and Yu, B., *Statistical guarantees for the EM algorithm: From population to sample-based analysis*, *The Annals of Statistics*, 45(1) (2017), 77--120. doi = 10.1214/16-AOS1435 various facts about EM algorithm in general, useful in particular because it shows the general principle that likelihood gradient = EM gradient. balakrishnan, wainwright, yu page 82 as cited at <https://stats.stackexchange.com/questions/45652/what-is-the-difference-b>
- [3] Bouchaud, J.-P., Bonart, J., Donier, J., and Gould, M., *Trades, Quotes and Prices: Financial Markets Under the Microscope*, Cambridge University Press, 2018. ISBN = 9781107156050
- [4] Chen, F. and Stindl, T., *Direct Likelihood Evaluation for the Renewal Hawkes Process*, *Journal of Computational and Graphical Statistics*, 27(1) (2018), 119--131. doi = 10.1080/10618600.2017.1341324
- [5] Daley, D. J. and Vere-Jones, D., *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, 2nd edition, Springer, 2003. doi = 10.1007/b97277
- [6] Embrechts, P., Liniger, T., and Lin, L., *Multivariate Hawkes processes: An application to financial data*, *Journal of Applied Probability*, 48(A) (2011), 367--378. doi = 10.1239/jap/1318940477
- [7] Financial Industry Regulatory Authority, *2024 Industry Snapshot: Market Data*, FINRA, 2024. Available at: <https://www.finra.org/media-center/reports-studies/2024-industry-snapsho> Accessed: 2 September 2024.
- [8] Jamshidian, M. and Jennrich, R. I., *Acceleration of the EM Algorithm by Using Quasi-Newton Methods*, *Journal of the Royal*

- Statistical Society. Series B (Methodological)*, 59(3) (1997), 569--587. Available at: <http://www.jstor.org/stable/2346010>
- [9] Jamshidian, M. and Jennrich, R. I., *Standard errors for EM estimation*, *Biometrika*, 89(1) (2002), 63--75. doi = 10.1111/1467-9868.00230
 - [10] Jiang, A. Z. and Rodriguez, A., *Improvements on Scalable Stochastic Bayesian Inference Methods for Multivariate Hawkes Processes*, *Statistics and Computing*, 34, article 85 (2024). doi = 10.1007/s11222-024-10392-x
 - [11] Laub, P. J., Lee, Y., and Taimre, T., *The Elements of Hawkes Processes*, Springer, 2021. doi = 10.1007/978-3-030-84639-8
 - [12] Lewis, E. and Mohler, G., *A nonparametric EM algorithm for multiscale Hawkes processes*, *Journal of Nonparametric Statistics*, 1(1) (2011), 1--20. one example of EM algorithm applied to point processes. may or may not deserve a direct citation, but can see if anything cited here is useful
 - [13] Morariu-Patrichi, M. and Pakkanen, M. S., *State-Dependent Hawkes Processes and Their Application to Limit Order Book Modelling*, *Quantitative Finance*, 22(3) (2022), 563--583 doi = 10.1080/14697688.2021.1983199
 - [14] Nickel, M. and Le, M., *Learning Multivariate Hawkes Processes at Scale*, arXiv preprint arXiv:2002.12501 [cs.LG], 2020. doi = 10.48550/arXiv.2002.12501
 - [15] Smith, A. C. and Brown, E. N., *Estimating a state-space model from point process observations*, *Neural Computation*, 15(5) (2003), 965--991. doi = 10.1162/089976603765202622
 - [16] Veen, A. and Schoenberg, F. P., *Estimation of Space-Time Branching Process Models in Seismology Using an EM-Type Algorithm*, *Journal of the American Statistical Association*, 103(June) (2008), 614--624. doi = 10.1198/016214508000000148
 - [17] Zhou, F., Li, Z., Fan, X., Wang, Y., Sowmya, A., and Chen, F., *Efficient inference for nonparametric Hawkes processes using auxiliary latent variables*, *The Journal of Machine Learning Research*, 21(1) (2020), 9745--9775.
- Unformatted:
- [18] On Lewis' simulation method for point processes
<https://ieeexplore.ieee.org/document/1056305>
 - [19] LONG TIME BEHAVIOUR OF A HAWKES PROCESS-BASED LIMIT ORDER BOOK
<https://hal.science/hal-01121711v5/document>
 - [20] https://link.springer.com/chapter/10.1007/978-88-470-1766-5_4
<https://www.idescat.cat/sort/sort461/46.1.1.Worrall-et al.pdf>
<https://www.tandfonline.com/doi/full/10.1080/10618600.2022.2050247>
<https://www.santafe.edu/research/results/working-papers/studies-of-the-limit-o>
<https://arxiv.org/pdf/1903.03223>

https://link.springer.com/chapter/10.1007/978-88-470-1766-5_4
<https://www.jstor.org/stable/2334319>
<https://www.tandfonline.com/doi/full/10.1080/14697688.2021.1983199>
<https://www.tandfonline.com/doi/full/10.1080/10618600.2017.1341324>
https://www.researchgate.net/publication/4742983_Estimation_of_Space-Time_Branching_Processes
<https://pubmed.ncbi.nlm.nih.gov/12803953/>
http://paleo.sscnet.ucla.edu/Lewis-Molher-EM_preprint.pdf
<https://opus.lib.uts.edu.au/bitstream/10453/145676/2/19-930.pdf>
<https://projecteuclid.org/journals/annals-of-statistics/volume-45/issue-1/Statistical-Inference-for-Branching-Processes-with-Applications-to-Neural-Networks>
<https://www.tandfonline.com/doi/full/10.1080/1351847X.2021.1917441>
<https://projecteuclid.org/journals/annals-of-probability/volume-24/issue-3/Statistical-Inference-for-Branching-Processes-with-Applications-to-Neural-Networks>
 Data Sketching for Large-Scale Kalman Filtering
<https://arxiv.org/pdf/1606.08136>