POINT PROCESS MODELLING OF A LIMIT ORDER BOOK

Oden Petersen

Supervisor: Dr. Tom Stindl

School of Mathematics and Statistics
UNSW Sydney

June 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

Modern financial markets facilitate the exchange of shares and other financial assets at extremely high frequency. The ability to trade throughout the day allows market participants to respond to exogenous news and endogenous market events in a manner that minimises their risk and generates profits over the long term. Intraday trading activity has a significant influence on the formation of market pricesm 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 play a key role in this ecosystem, facilitating the 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.

1.1 Limit Order Books

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.

Conversely, an order cancellation message results in the removal of a particular order from the limit order 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

$$(\text{side}, \text{price}, \text{time}, \text{size}) \in \{-1, 1\} \times P \times T \times Q,$$

where

- The order expresses an intention to buy (side$=1$) or sell (side$=-1$) the asset
- The price of the order is a real number. Typically prices are restricted to some discrete set of values, and the difference between successive prices is known as the *tick size* (usually constant).

I plan to also highlight in this section some key features of limit order book structures, such as the definition of bid/ask levels, the mechanics of the matching process and queue priority, and the evolution of the limit order book over time, motivating the emphasis on modeling arrival times. I also plan to highlight the fact that queue depletions are intimately related to price changes; this is discussed in the Bouchaud source and seems like important motivation.

1.2 Point Process Models of Book Updates

With the goal of modelling the arrival time process, we now turn to an introductory overview of point processes, and highlight some key variants described in existing literature that are relevant to this application.

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

1.2.1 Point Processes

Given a sample space Ω with probability measure $\mathbb{P}$, a point process is any increasing random sequence of times

$$\mathcal{T} : \Omega \to {\mathbb{R}_{\geq 0}}^{\mathbb{N}}.$$

These times 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 nondecreasing random function $N_{\mathcal{T}} : \mathbb{R}_{\geq 0} \times \Omega \to \mathbb{N}$ known as the *counting process*, that gives the number of times having occurred before or at a given time. This is defined as

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

All (or perhaps many) of the point processes considered in this thesis will additionally be *adapted point processes*, in the sense that $N_{\mathcal{T}}$ is a stochastic process adapted with respect to some filtration $\mathcal{F}$.

Perhaps I should define here what a filtration is?

A point process $\mathcal{T}$ may also be *marked*, in which case it is associated with one or more random sequences of marks $\mathcal{E} : \Omega \to E^{\mathbb{N}}$ drawn from a set E, typically representing additional information about each event. A common special case of this is when E is finite, in which case we may refer to the elements of E as *event* *types*, and refer to $\mathcal{T}$ as a *multivariate point process* with a *multivariate counting process*

$$\tilde{N}_{\mathcal{T},\mathcal{E}} : \mathbb{R}_{\geq 0} \times \omega \to \mathbb{N}^{|E|}$$
$$N_{\mathcal{T},\mathcal{E}}(t, \omega)_e = |\{i \in \mathbb{N} | \mathcal{T}(\omega)_i \leq t, \mathcal{E}(\omega)_i = e\}|.$$

1.2.2 The Expectation Measure

Let $B(\mathbb{R})$ be the Borel σ-algebra on $\mathbb{R}$. Does the σ-algebra need to be this restricted, or can it just be $\mathcal{P}(\mathbb{R})$?

For any point process $\mathcal{T}$ and for any $\omega \in \Omega$, we can define the *counting measure* Λ_ω as

$$\Lambda_\omega : B(\mathbb{R}) \to \mathbb{N}$$
$$\Lambda_\omega(S) = |S \cap \mathcal{T}(\omega)|.$$

Taking the expectation of the counting measure with respect to $\mathbb{P}$ (optionally conditional on some sub-σ-algebra of Ω) gives the *expectation measure* $\Lambda(S)$.

For the sake of intuition, it should probably be shown here that $d\Lambda = d(\mathbb{E}N)$

It is also important to define the expectation measure with respect to a sigma-algebra, and in particular one from the filtration.

1.2.3 Intensity Functions

These can be formulated using the Radon-Nikodym derivative of Λ with respect to the lebesgue measure on time, i.e. $d\Lambda = \lambda dt$. For intuition, this is the arrival rate.

1.2.4 Poisson Processes

Probably the simplest modeling assumption we can make is to assume that the arrival rate of the process is a constant conditional on any sigma-algebra in the filtration to which the process is adapted.

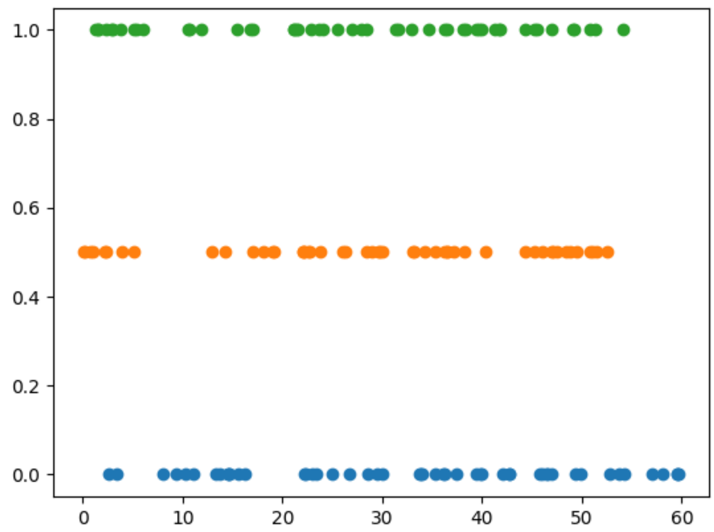
A Poisson process is a point process for which the following two properties hold.

1. For any collection of disjoint Borel sets $B \subset B(\mathbb{R})$, their couting measures $\{\Lambda_w(b) : b \in B\}$ are independent random variables.
2. For any Borel set b with finite Lebesgue measure $\mu(b)$, we have $\Lambda(b) = \lambda\mu(b) \propto \mu(b)$ for some fixed $\lambda \in \mathbb{R}$. Put differently, we have $d\Lambda = \lambda dt$.

This definition might require a reference.

Multivariate case

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.

1.2.5 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 Borel set b with finite Lebesgue measure $\mu(b)$ has a finite expectation measure. Is this weaker version even needed, or is it sufficient to just drop the condition entirely?

1.2.6 Autoregressive Intensity & Hawkes Processes

Intuitively, self-excitation or self-inhibition

immigration/birth interpretation

1.2.7 State-dependent Hawkes Processes

Moriaru-Patrichi and Pakkanen find that conditioning on the state of the order book improves the fit of the model.

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:

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

Chapter 2

Estimation

2.1 Introduction to estimation algorithms

2.1.1 MLE

2.1.2 EM

2.1.3 Bayesian Approach

- Uncertainty quantification - Less likely to get stuck in local optima

2.2 Prior work on estimation

Chapter 3

Application to NYSE Data

Chapter 4

Conclusion

References

Other: The elements of hawkes processes An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods, Second Edition Multivariate hawkes processes: an application to financial data Embrechts, liniger, lin direct likelihood evaluation for the renewal hawkes process state dependent hawkes processes and their applicvation to limit order book modeling morariu-patrichi & pakkanen estimation of space time branching process models in seismology using an em type algorithm learning multivariate hawkes processes at scale nickel, le improvements on scalable stochastic bayesian inference methods for multivaariate hawkes processes jiang and rodriguez Bouchaud et al, Trades quotes and prices