

Rafael Zimmer

**Abordagens Numéricas para Simuladores de Processos Estocásticos de
Chegada de Ordens em Mercados Financeiros**

Orientador:

Prof. Dr. **OSWALDO LUIZ DO VALLE COSTA**

Tese para obtenção do título de Bacharel em Ciências
da Computação

**São Paulo
2024**

RESUMO

Abordagens Numéricas para Simuladores de Processos Estocásticos de Chegada de Ordens em Mercados Financeiros

O presente trabalho propõe uma técnica de modelagem para simuladores de mercado financeiro, especificamente das dinâmicas de livros de ordens limite. A utilização de simuladores se revela de extrema importância para testar estratégias de negociação sem depender de grandes volumes de dados históricos, especialmente para estratégias cujo impacto no estado do livro de ordens não é desprezível e também não é refletido ao se utilizar ordens históricas e estáticas.

O simulador a ser desenvolvido é fundamentado na teoria de modelos ocultos de Markov (*Hidden Markov Chains*, ou *HMC*, em inglês) para modelar o comportamento dinâmico do livro de ordens limite. Além disso, serão comparadas diferentes distribuições para os processos de chegada de ordens e suas intensidades, especificamente os Processos de Hawkes, Cadeias de Markov Autoregressivas e Modelos de Hawkes Dependentes do Estado. Por fim, a principal contribuição do trabalho introduzir o uso de algoritmos genéticos como técnica para aproximar as distribuições dos processos de chegada de ordens dependentes dos estados, utilizando-se estimativas por máxima verossimilhança para comparar com outras técnicas utilizadas na literatura.

A abordagem proposta visa melhorar a precisão das simulações e fornecer insights mais robustos sobre o comportamento do mercado financeiro em diferentes cenários. O uso de algoritmos evolutivos tem como objetivo contribuir no desenvolvimento e na melhor compreensão de modelos simuladores do mercado financeiro, assim como o desenvolvimento de estratégias de negociação mais eficientes e resilientes a diferentes regimes de mercado.

Palavras-chave: Hidden Markov Chains, Limit Order Book Modelling, High-Frequency Data, Markov Chain Monte Carlo methods, Genetic Algorithms

JEL: C45; C15; G17

1 INTRODUÇÃO

O livro de ordens limite representa o estado atual de todas ordens aguardando execução. Essencialmente fornece uma representação dinâmica das intenções de compra e venda de ativos financeiros em um determinado mercado por meio de uma lista de ordens, ordenadas por preço e tempo de chegada, onde cada entrada representa o desejo de um participante do mercado de comprar ou vender um certo ativo (Abergel et al., 2020; Avellaneda e Stoikov, 2008). Os intervalos de tempo entre a chegada de duas ordens subsequentes são comumente representados por processos estocásticos e o estudo desses processos é uma abordagem fundamental para a modelagem e representação da dinâmica dos mercados financeiros (Shi e Cartlidge, 2022; Guilbaud e Pham, 2013; Liu et al., 2021). Tradicionalmente, os modelos de chegada de ordens têm se baseado em processos de contagem, onde os incrementos entre eventos são distribuídos de acordo com distribuições de Poisson ou distribuições exponenciais (Cont et al., 2022; Ponta et al., 2012). Recentemente surge grande interesse na utilização de processos de Hawkes para modelar os processos de chegada, devido a natureza autoexcitável dos mesmos (Abergel et al., 2020; Morariu-Patrichi e Pakkanen, 2022; Toke, 2011).

O uso de uma distribuição com parâmetros fixos para estimar os tempos de chegada não considera as diferentes propriedades dentre os diversos regimes financeiros existentes do mercado, tanto em escala macro, como por exemplo regimes de inflação (Krause et al., 2022), como em escalas de curto prazo, como por exemplo regimes de alta ou baixa volatilidade, liquidez e volume de negociação (Guilbaud e Pham, 2013), que exibem comportamentos distintos entre si. Esses regimes podem ser influenciados por uma variedade de fatores, incluindo eventos econômicos, políticos e sazonais e resultam em alterações nas características estatísticas dos processos de preços, e chegada de ordens (Krause et al., 2022).

Uma abordagem promissora que considere os efeitos dos diferentes regimes é o uso de modelos ocultos de Markov (*Hidden Markov Chains*, em inglês, ou *HMC*) (Cont et al., 2010). Um *HMC* é um modelo estocástico que assume a existência de estados ocultos, não observáveis diretamente, mas que podem ser inferidos a partir de observações de variáveis externas (Baum e Petrie, 1966). Nesse contexto, os diferentes estados ou regimes implícitos do livro de ordens e as diferentes intensidades de chegada de ordens podem ser representados como estados ocultos da cadeia de Markov e as distribuições ou parâmetros associados com os respectivos estados (Morariu-Patrichi e Pakkanen, 2022). Essa abordagem permite modelar a transição entre diferentes regimes de mercado de forma dinâmica, considerando uma matriz de transição de estados.

Na literatura de simulação de livros de ordem foram utilizados alguns modelos para a identificação e simulação dos diferentes estados do livro, como por exemplo modelos de *Markov Switching* (Guilbaud e Pham, 2013) ou algoritmos de *Thinning* para simulação de processos dependentes de estados (Ponta et al., 2012). Além disso, as distribuições utilizadas para modelar o processo de chegada de ordens são mantidas fixas durante a simulação, com apenas seus parâmetros sendo ajustados e obtidos a partir dos estados atuais. Em suma, essas abordagens requerem que sejam feitas assunções sobre o funcionamento intrínscio do mercado, além de limitar as características estatísticas da simulação apenas ao período observado devido ao ajuste dos parâmetros não capturarem todas características do conjunto de dados (Zare et al., 2021).

Considerando tais problemas, o presente trabalho propõe o uso e comparação de duas abordagens numéricas para otimizar a amostragem das distribuições dos processos de chegada, cujo formato é tomado como desconhecido como ponto de partida inicial:

- Algoritmos genéticos (AGs) têm se destacado como uma ferramenta poderosa para otimização e modelagem em uma variedade de domínios, incluindo finanças (Oesch e Maringer, 2013; Katoch et al., 2021). Esses algoritmos podem ser aplicados para aproximar as possíveis distribuições dos processos de chegada de ordens, utilizando estimativas por máxima verossimilhança (Boonthiem

et al., 2023; Colla et al., 2010) para ajustar os parâmetros das populações aos dados históricos observados;

- Métodos de Monte Carlo para Cadeias de Markov (*Markov chain Monte Carlo* em inglês, ou MCMC) também vêm ganhando destaque como uma abordagem eficaz para amostragem de distribuições complexas e multidimensionais (Rasmussen, 2013; Rousseau et al., 2018). Os métodos de MCMC, como o algoritmo de Metropolis-Hastings e o Gibbs sampling, são particularmente úteis quando se trata de amostrar a partir de distribuições de probabilidade desconhecidas ou difíceis de obter diretamente (Glasserman, 2004). Ao construir uma cadeia de Markov com distribuição estacionária igual à distribuição alvo, esses métodos podem possivelmente gerar amostras que representam fielmente a distribuição desejada.

Em suma, este trabalho propõe uma abordagem computacional alternativa para modelar e simular o livro de ordens limite e os processos de chegada de ordens no mercado financeiro, utilizando Hidden Markov Chains e otimizadores de amostragens por meio de bibliotecas de computação em unidades de processamento gráficas (*GPUs*). As abordagens propostas tem o potencial de melhorar significativamente a precisão das simulações e fornecer propriedades estatísticas mais robustas sobre o comportamento do mercado em diferentes regimes para teste de estratégias de negociação algorítmicas.

2 OBJETIVOS DO TRABALHO

Este trabalho visa comparar o desempenho de Algoritmos Genéticos (AGs) e Métodos de Monte Carlo para Cadeias de Markov (MCMC) na otimização da amostragem das distribuições dos processos de chegada de ordens no mercado financeiro, com foco nos resultados computacionais das simulações e de treino, especificamente os tempos de execução para teste, geração de dados e tempo para realização dos testes estatísticos, tanto em CPU, quanto em unidades de processamento gráficas (*GPU*) utilizando bibliotecas de computação vetorial. A primeira abordagem utiliza AGs para ajustar os parâmetros das distribuições aos dados históricos, enquanto a segunda emprega MCMC, como Metropolis-Hastings e Gibbs sampling, para amostrar diretamente das distribuições. O objetivo é avaliar a precisão e eficácia de cada método na replicação das características estatísticas dos processos de chegada em dados históricos, fornecendo insights para estratégias de negociação algorítmicas mais robustas.

3 METODOLOGIA

Para atingir os objetivos propostos, adotaremos uma metodologia em duas etapas. Primeiramente, será implementado um simulador de mercado com dados históricos estáticos disponibilizados pela bolsa de Nova York (NASDAQ) (Nagy et al., 2023). Em sequência, o simulador será modelado para utilizar um estimador das distribuições de chegada de eventos, especificamente com duas abordagens a serem comparadas: Algoritmos Genéticos (AGs) e Métodos de Monte Carlo para Cadeias de Markov (MCMC) como modelos dos processos de chegada de ordens. Utilizaremos linguagens de programação adequadas para a implementação eficiente desses métodos, em Python com bibliotecas de processamento vetorial, especificamente NumPy, PyMC3 e JAX (Bradbury et al., 2018; Oriol et al., 2023).

1. Na primeira etapa, aplicaremos AGs para otimizar a representação das distribuições dos processos de chegada de ordens. Isso envolverá a definição de uma função de aptidão (fitness) que quantifique a adequação de uma determinada distribuição aos dados históricos. Os parâmetros das distribuições serão ajustados iterativamente através de operadores genéticos, como seleção, recombinação e mutação, para maximizar a função de aptidão.
2. Na segunda etapa, implementaremos e utilizaremos métodos de MCMC, como Metropolis-Hastings e Gibbs sampling, para amostrar diretamente das distribuições dos processos de chegada. Faremos isso construindo cadeias de Markov com distribuição estacionária igual à distribuição alvo. Avaliaremos a convergência das cadeias de Markov e a qualidade das amostras geradas em relação aos dados históricos.

Por fim, conduziremos experimentos computacionais para comparar o desempenho e a eficácia dos dois métodos em termos de precisão na representação dos processos de chegada de ordens. Analisaremos métricas relevantes, como erro médio, tempo de convergência e robustez em diferentes cenários de mercado.

4 CRONOGRAMA

A execução da pesquisa para o trabalho de conclusão do curso será realizada de acordo com o seguinte cronograma, para o Projeto de Graduação I, e será continuada no Projeto de Graduação II no primeiro semestre de 2025.

Etapa	1. Julho	2. Agosto	3. Setembro	4. Outubro	5. Novembro
1. Pesquisa Bibliográfica	X				
2. Análise Bibliográfica	X	X			
3. Obtenção de Dataset		X			
4. Definição dos modelos		X	X	X	
5. Implementação em código			X	X	X
6. Análise das métricas				X	X
7. Escrita do artigo			X	X	X

REFERÊNCIAS BIBLIOGRÁFICAS

- Abergel, F., Hure, C., e Pham, H. (2020). Algorithmic trading in a microstructural limit order book model. *QUANTITATIVE FINANCE*, 20(8):1263–1283.
- Avellaneda, M. e Stoikov, S. (2008). High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224.
- Baum, L. E. e Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563.
- Boonthiem, S., Sutikasana, C., Klongdee, W., e Ieosanurak, W. (2023). Parameter estimations of normal distribution via genetic algorithm and its application to carbonation depth. *WSEAS TRANSACTIONS ON MATHEMATICS*, 22:184–189.
- Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., e Zhang, Q. (2018). JAX: composable transformations of Python+NumPy programs.
- Colla, V., Nastasi, G., e Matarese, N. (2010). Gadf — genetic algorithms for distribution fitting. *Proceedings of the 2010 10th International Conference on Intelligent Systems Design and Applications, ISDA '10*, pages 6–11.
- Cont, R., Cucuringu, M., Kochems, J. A., e Prenzel, F. (2022). Dynamic calibration of order flow models with generative adversarial networks. In *3RD ACM INTERNATIONAL CONFERENCE ON AI IN FINANCE, ICAIF 2022*, pages 446–453, 1601 Broadway, 10th Floor, NEW YORK, NY, UNITED STATES. Assoc Comp Machinery; J P Morgan Chase & Co; U S Bank, ASSOC COMPUTING MACHINERY. 3rd ACM International Conference on AI in Finance (ICAIF), New York, NY, NOV 02-04, 2022.
- Cont, R., Stoikov, S., e Talreja, R. (2010). A stochastic model for order book dynamics. *Operations Research*, 58(3):549–563.
- Glasserman, P. (2004). *Monte Carlo methods in financial engineering*. Springer, New York.
- Guilbaud, F. e Pham, H. (2013). Optimal high-frequency trading with limit and market orders. *QUANTITATIVE FINANCE*, 13(1):79–94.
- Katoch, S., Chauhan, S. S., e Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 80(5):8091–8126.
- Krause, S. M., Jungblut, E., e Guhr, T. (2022). Two price regimes in limit order books: liquidity cushion and fragmented distant field. *JOURNAL OF STATISTICAL MECHANICS-THEORY AND EXPERIMENT*, 2022(2).
- Liu, Y., Xing, Y., Yang, X., Wang, X., Shi, J., Jin, D., e Chen, Z. (2021). Learning continuous-time dynamics by stochastic differential networks.
- Morariu-Patrichi, M. e Pakkanen, M. S. (2022). State-dependent hawkes processes and their application to limit order book modelling. *QUANTITATIVE FINANCE*, 22(3):563–583.
- Nagy, P., Calliess, J.-P., e Zohren, S. (2023). Asynchronous deep double dueling q-learning for trading-signal execution in limit order book markets. *FRONTIERS IN ARTIFICIAL INTELLIGENCE*, 6.

- Oesch, C. e Maringer, D. (2013). Portfolio optimization under market impact costs. In *2013 IEEE CONGRESS ON EVOLUTIONARY COMPUTATION (CEC)*, pages 1–7, 345 E 47TH ST, NEW YORK, NY 10017 USA. IEEE; IEEE Computat Intelligence Soc, IEEE. IEEE Congress on Evolutionary Computation, Cancun, MEXICO, JUN 20-23, 2013.
- Oriol, A.-P., Virgile, A., Colin, C., Larry, D., J., F. C., Maxim, K., Ravin, K., Jupeng, L., C., L. C., A., M. O., Michael, O., Ricardo, V., Thomas, W., e Robert, Z. (2023). Pymc: A modern and comprehensive probabilistic programming framework in python. *PeerJ Computer Science*, 9:e1516.
- Ponta, L., Scalas, E., Raberto, M., e Cincotti, S. (2012). Statistical analysis and agent-based microstructure modeling of high-frequency financial trading. *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING*, 6(4):381–387.
- Rasmussen, J. (2013). Bayesian inference for hawkes processes. *Methodology and Computing in Applied Probability*, 15.
- Rousseau, J., Donnet, S., e Rivoirard, V. (2018). Nonparametric bayesian estimation of multivariate hawkes processes. *Annals of Statistics*, 48.
- Shi, Z. e Cartlidge, J. (2022). State dependent parallel neural hawkes process for limit order book event stream prediction and simulation. In *PROCEEDINGS OF THE 28TH ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, KDD 2022*, pages 1607–1615, 1601 Broadway, 10th Floor, NEW YORK, NY, UNITED STATES. Assoc Comp Machinery; ACM SIGKDD; ACM SIGMOD, ASSOC COMPUTING MACHINERY. 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Washington, DC, AUG 14-18, 2022.
- Toke, I. M. (2011). “market making” in an order book model and its impact on the spread. In Abergel, F., Chakrabarti, B., Chakraborti, A., e Mitra, M., editors, *ECONOPHYSICS OF ORDER-DRIVEN MARKETS*, New Economic Windows, pages 49–64, 233 SPRING STREET, NEW YORK, NY 10013, UNITED STATES. Saha Inst Nucl Phys, Ctr Appl Math & Comp Sci; Indian Stat Inst; Ecole Cent Paris, SPRINGER. 5th Econophys-Kolkata Conference, Saha Inst Nucl Phys, Kolkata, INDIA, MAR, 2010.
- Zare, M., Naghshineh Arjmand, O., Salavati, E., e Mohammadpour, A. (2021). An agent-based model for limit order book: Estimation and simulation. *INTERNATIONAL JOURNAL OF FINANCE & ECONOMICS*, 26(1):1112–1121.