Unsupervised Generative Framework for Event-Driven Financial Time Series

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Mihir Agarwal, Aaditya Barve

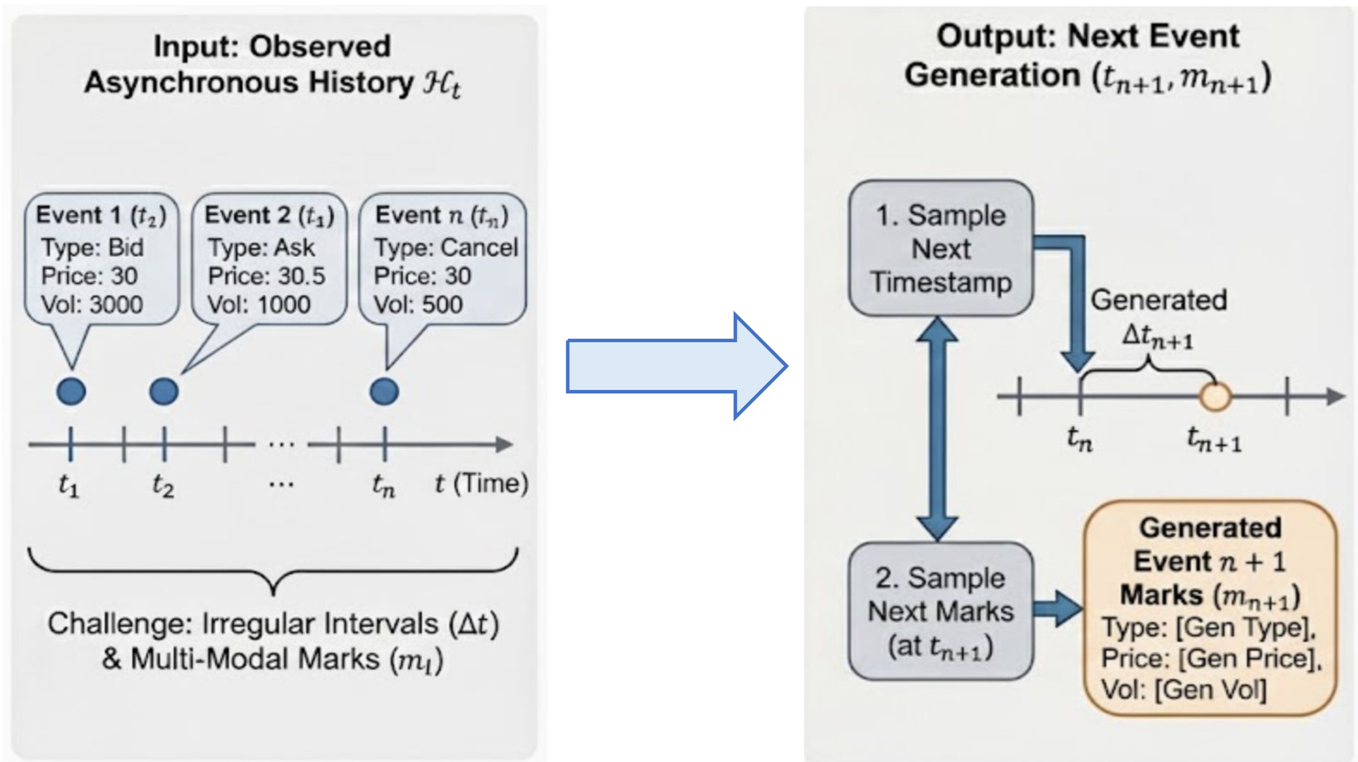
COMS 4774: Unsupervised Learning

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Problem Statement and motivation

Asynchronous Event-Driven Systems

- Many real-world systems evolve via discrete, irregular events rather than fixed clock ticks.
- Examples:
  - Finance: Limit orders arriving at nanosecond intervals.
  - Seismology: Earthquake aftershocks sequences.
  - Healthcare: Patient vital monitoring and drug administration.
-



Literature Survey

CLASSIC ALGORITHMS

- **Hawkes Process:** Models "self-exciting" events where order arrivals increase the probability of future orders.[1]
- **Queue-Reactive Models:** Order placement probability is influenced by the current queue size.[2]
- **Stochastic Differential Equations (SDEs):** Approximates the LOB as a system of Partial Differential Equations. [3]
- **Zero Intelligence (ZI) Agents:** A "null hypothesis" testing if market laws emerge from random order placement.[4]

STATE-OF-THE-ART (SOTA) ALGORITHMS

- **"Painting the Market" (2024):** Uses Denoising Diffusion Probabilistic Models (DDPMs) to generate LOB volume shapes.[5]
- **TRADES (2025):** A Transformer-based Denoising Diffusion engine handling time-series dependencies and complex joint distributions.[6]
- **State Space Models (SSMs):** Uses S5/Mamba architecture, which is faster than Transformers for long sequences.[7]

BENCHMARK DATASETS

- **LOBSTER:** Reconstructed NASDAQ data (Industry Gold Standard). [8]
- **LOB-Bench (2025):** Open-source benchmark suite for standardizing evaluation. [9]
- **Crypto-LOBs:** High volatility datasets from Binance/ Coinbase, crucial for modeling extreme events. [10] **(Our Dataset)**

Baseline Architectures

Neural Point Process (NPP):

Dual-Head Prediction

Continuous Head (Regression): Predicts time, price, and volume.
Discrete Head (Classification): Predicts the event type (e.g., buy/sell/cancel).

$$\hat{p}_{type} = \text{softmax}(\phi_{cls}(c))$$

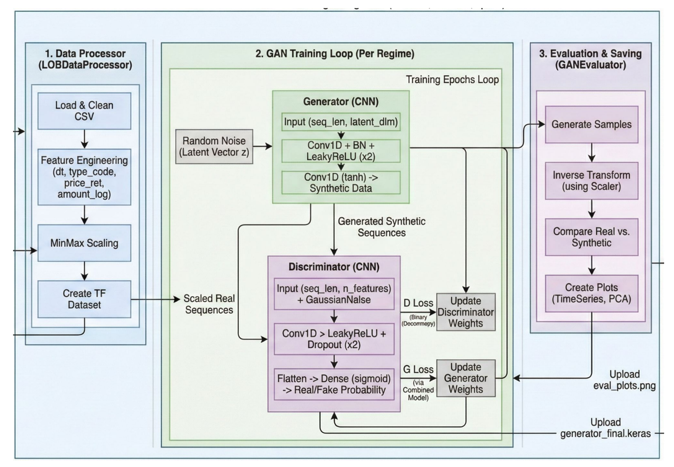
Optimization & Loss

(Mean Squared Error + Cross-Entropy) $\mathcal{L}_{total} = \|x_{cont} - \hat{x}_{cont}\|^2 - \sum_k y_k \log(\hat{p}_k)$

Generative Adversarial Network(GAN):

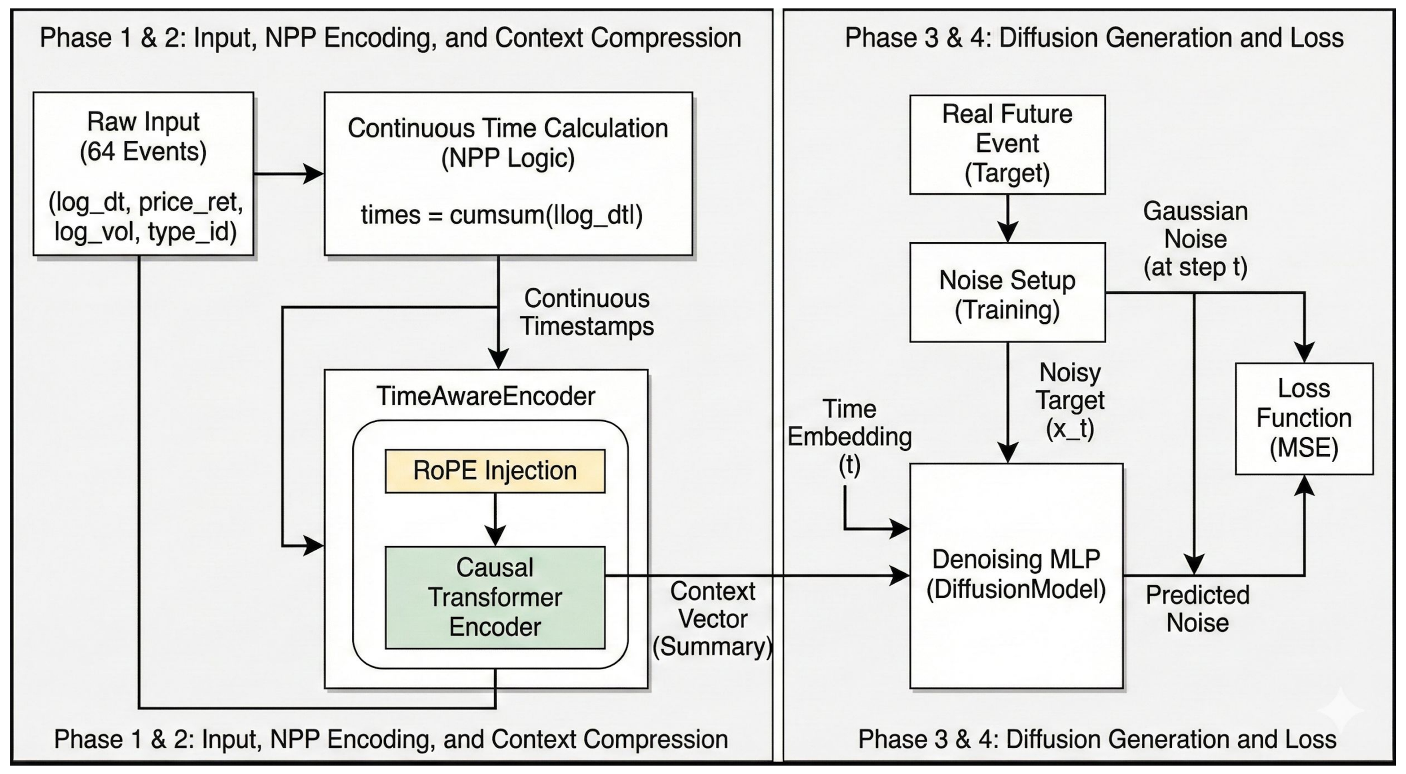
The input vector x_t combines temporal and order flow features. 1D-Convolutional GAN

$$x_t = [\ln(\Delta t), \text{OneHot}(\text{type}), r_t, \ln(v_t)]$$



Adversarial Training Objective is optimized via a minimax game using Binary Cross-Entropy (BCE). The Discriminator minimizes the classification error, while the Generator maximizes the probability of D making a mistake: $\min_G \max_D V(D,G) = \mathbb{E}_x[\ln D(x)] + \mathbb{E}_z[\ln(1 - D(G(z)))]$

Marked Temporal Point Process



Event Representation: Each event x_i is a multi-dimensional vector comprising inter-arrival time Δt, price return Δp, volume v, and event type k:

$$x_i = [\ln(\Delta t_i), \Delta p_i, \ln(v_i), \text{OneHot}(k_i)]$$

Continuous-Time Encoding: To capture irregular arrivals, we compute cumulative timestamps $\tau_i = \sum \Delta t_j$ and inject them directly into the attention mechanism using Rotary Positional Embeddings (RoPE):

$$\text{RoPE}(q, \tau) = (q \cos \tau) + (\text{rotate}(q) \sin \tau)$$

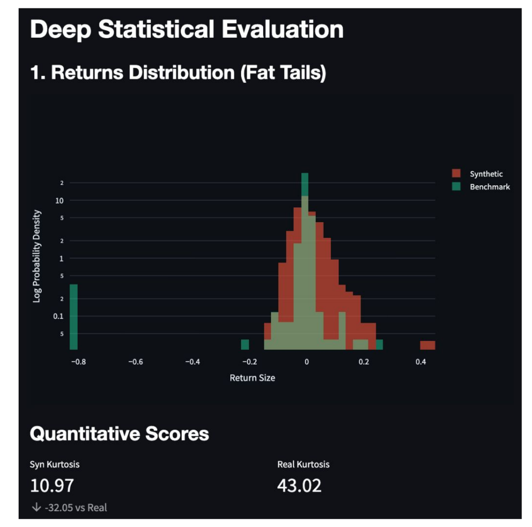
Context Aggregation: A Causal Transformer Encoder processes the historical sequence $H = \{x_{1:\ell}\}$ and time τ to produce a compact context vector c representing the market state.

Diffusion Generation: We model the distribution of the next event x□ex□ by reversing a Gaussian diffusion process. The forward process adds noise $\epsilon \sim \mathcal{N}(0, I)$ at step t:

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon$$

Training Objective: A Denoising MLP predicts the added noise, conditioned on the history context c. We optimize the standard variational lower bound via Mean Squared Error: $\mathcal{L}_{\text{simple}} = \mathbb{E}_{t, x_0, \epsilon} \left[\|\epsilon - \epsilon_\theta(x_t, t, c)\|^2 \right]$

Results



SCAN ME

Generate Custom Data!!

	Marked Temporal Point Process	NPP	GAN	Actual Data
KS Score	0.16	0.438	0.279	0.0

Conclusion and Future Scope

Summary: We successfully demonstrated that integrating Neural Point Processes with Diffusion Models captures the irregular, event-driven dynamics of Limit Order Books better than deterministic baselines.

Key Findings: A hybrid architecture faithfully reproduces complex market stylized facts, such as volatility clustering and heavy-tailed distributions, avoiding the "regression to the mean" issues seen in pure NPPs.

Future Directions:

- Downstream Applications
- Scalability

References

1. Bacry, E., Mastromatteo, I., & Muzy, J. F. (2015). Hawkes processes in finance. *Quantitative Finance*, 15(2), 195-212.
2. Huang, W., Lehalle, C. A., & Rosenbaum, M. (2015). Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509), 107-122.
3. Lasry, J. M., & Lions, P. L. (2007). Mean field games. *Japanese Journal of Mathematics*, 2(1), 229-260.
4. Gode, D. K., & Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, 101(1), 119-137.
5. Coletta, A., et al. (2024). Painting the Market: Generative Diffusion Models for Financial Limit Order Book Simulation and Forecasting. *arXiv preprint arXiv:2404.06556*.
6. Berti, L., et al. (2025). TRADES: Generating Realistic Market Simulations with Diffusion Models. *arXiv preprint arXiv:2502.07071*.
7. Gu, A., & Dao, T. (2023). Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *arXiv preprint arXiv:2312.00752*.
8. Huang, W., & Polak, T. (2011). LOBSTER: Limit Order Book Reconstruction System. *SSRN Electronic Journal*.
9. Nagy, P., et al. (2025). LOB-Bench: Benchmarking Generative AI for Finance -- an Application to Limit Order Book Data. *Proceedings of the 42nd International Conference on Machine Learning*.
10. Ntakaris, A., Magris, M., Kanniainen, J., Gabbouj, M., & Iosifidis, A. (2018). Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods. *Journal of Forecasting*, 37(8), 852-866.