SCHOOL OF COMPUTING & DATA SCIENCE
The University of Hong Kong

MARKET MAKING WITH DEEP REINFORCEMENT LEARNING

Aidan Wong Weng Seng (3035868918)
Supervisor: Dr. Huang Zhiyi

fyp25009

INTRODUCTION

Financial markets depend on market makers to provide liquidity by continuously quoting buy and sell prices.

Traditional models, such as the Avellaneda-Stoikov framework, have long served as benchmarks but rely on overly simplistic assumptions like treating price movements as independent of order flow and ignoring the time-dependent nature of market dynamics.

With the rise of deep learning, there is an opportunity to build adaptive, data-driven agents that learn directly from the complexity of the Limit Order Book (LOB).

This project develops a Deep Reinforcement Learning (DRL) agent trained in a realistic LOB simulation environment, combining a Hawkes process model with a Soft Actor-Critic (SAC) agent enhanced by an LSTM attention mechanism — a combination not previously explored in the literature.

OBJECTIVE

- Model the arrival of orders in the LOB using a Non-linear Multivariate Hawkes Process, capturing mutual excitation and inhibition effects between order types
- Construct a weakly consistent, realistic LOB simulation environment using the OpenAI Gymnasium library
- Develop a DRL agent using Soft Actor-Critic (SAC) enhanced with an LSTM attention mechanism for temporal pattern recognition
- Benchmark agent performance against a base top-level quoting model and evaluate improvements from sequential memory

METHODOLOGY

ENVIRONMENT

HAWKES PROCESS LOB SIMULATION

The Limit Order Book (LOB) records all outstanding bid and ask orders, forming the core data structure for market-making decisions. Order arrivals are modeled using a Non-linear Multivariate Hawkes Process — a self-exciting point process where each arrival triggers future arrivals, capturing the clustering behaviour of real markets:

$$\lambda_i(t) = \phi_i\left(\mu_i + \sum_{j=1}^{d}\int_0^t h_{ij}(t-s)\,dN_j(s)\right)$$

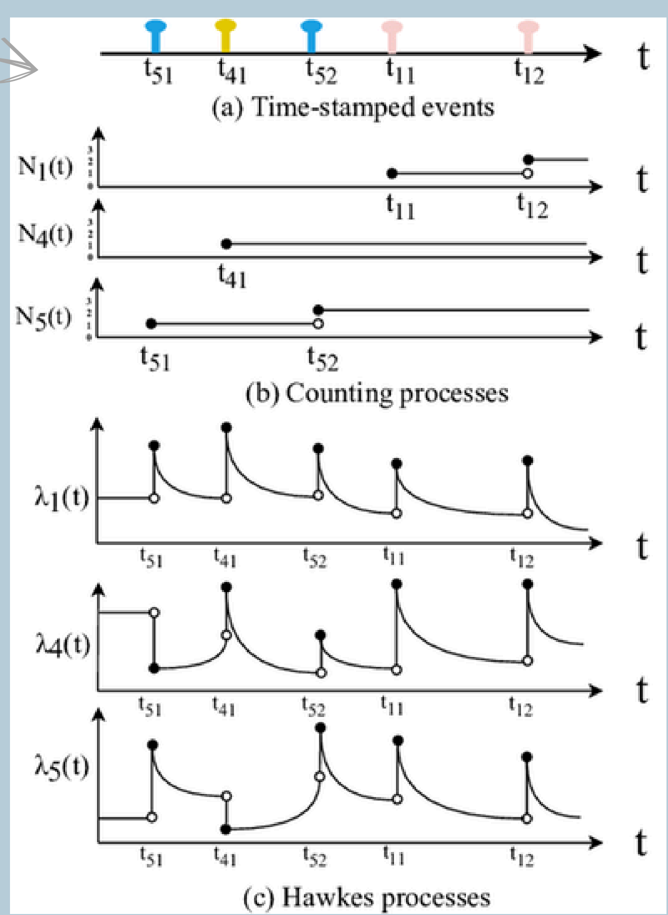
Non-linear Multivariate Hawkes Process

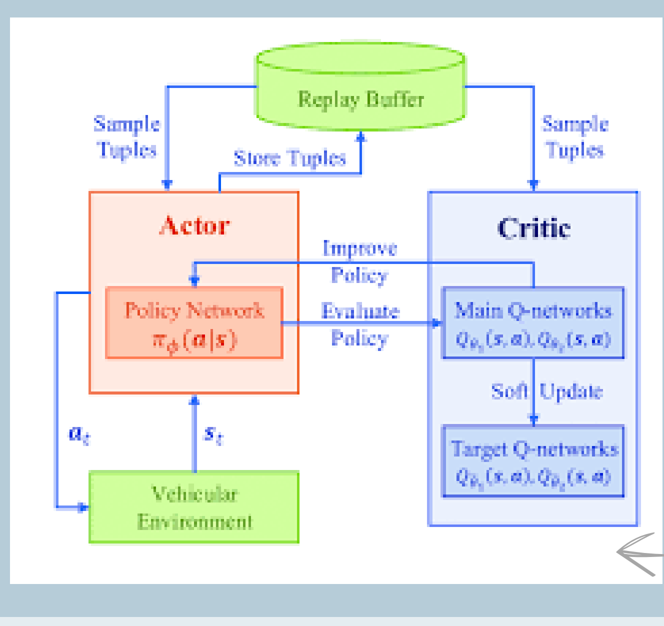
$$h_{ij}(t) = \alpha_{ij}e^{-\beta_{ij}t}\cdot \mathbf{1}_{\{t\geq 0\}}$$

Exponential kernel

$$P_t = P_0 + \left(\sum_{e\in E_{\text{inc}}} J_e - \sum_{e\in E_{\text{dec}}} J_e\right)\cdot\frac{\delta}{2}$$

Jump process of price from aggressive orders





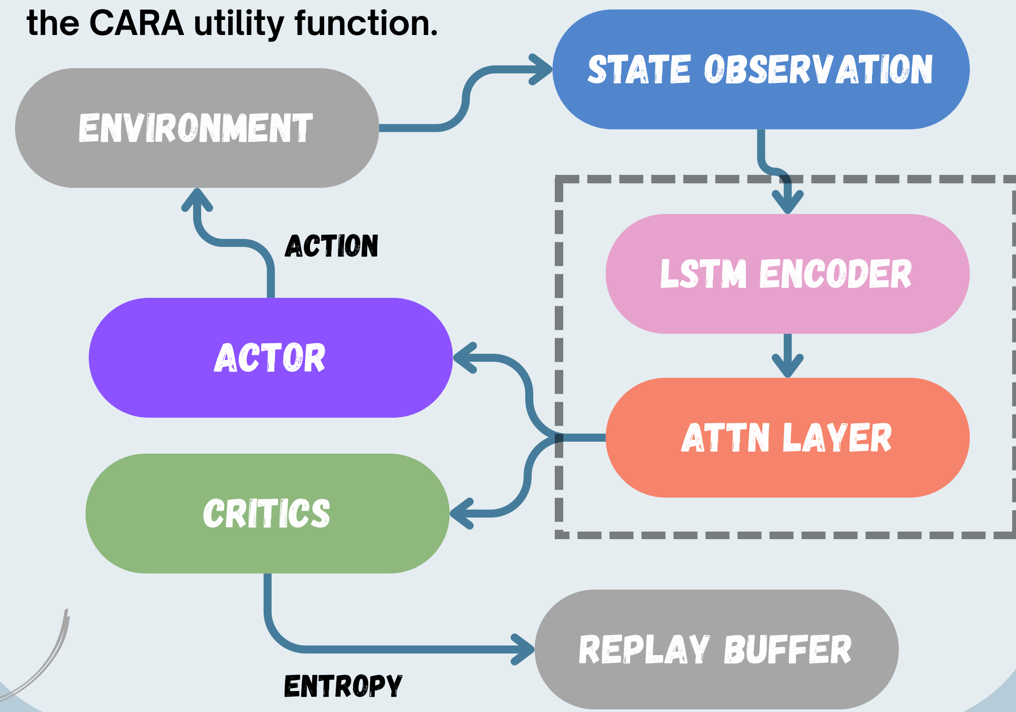
AGENT

SAC & SAC WITH LSTM (ATTN)

The agent uses Soft Actor-Critic (SAC), an off-policy DRL algorithm balancing reward maximisation and policy entropy to encourage robust exploration.

An LSTM with Attention Mechanism captures temporal dependencies in order flow — the LSTM retains memory of past market states while attention filters out high-frequency noise, focusing the agent on the most relevant signals.

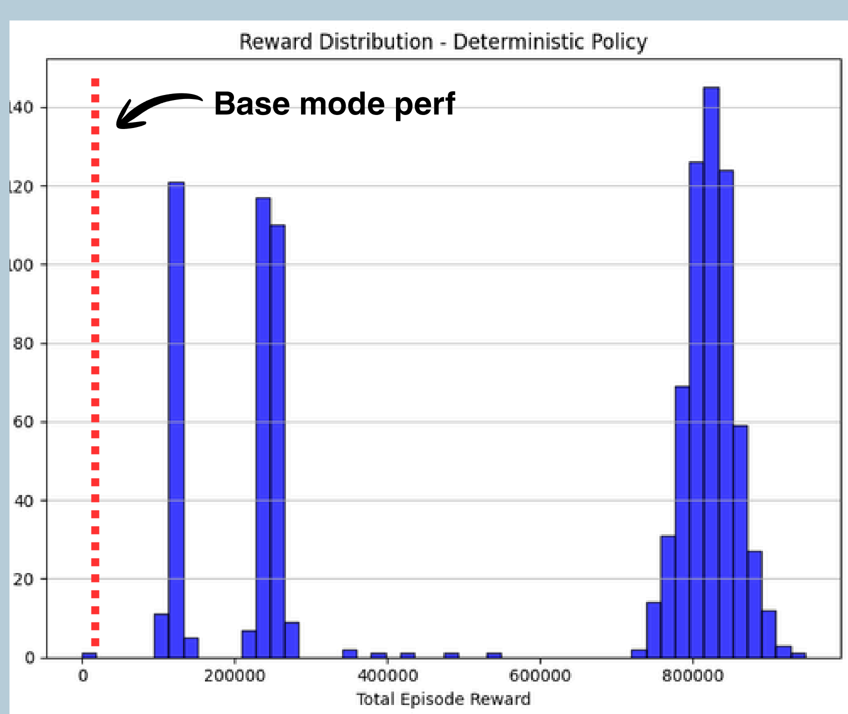
The reward function incorporates PnL, inventory risk penalties, and fill rates, with risk sensitivity modelled via the CARA utility function.



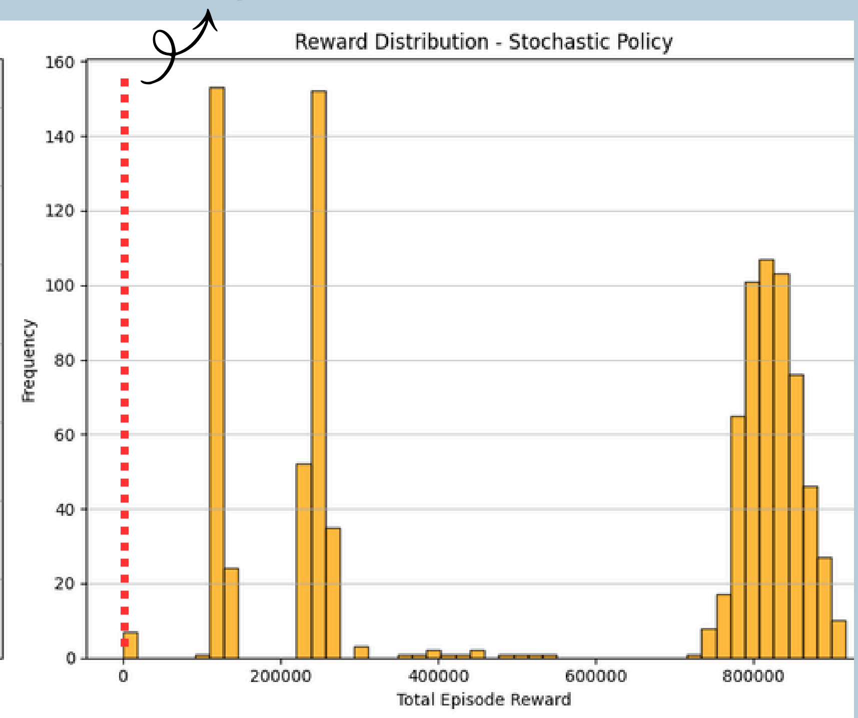
RESULTS

*Base Model: Quoting on top bid and ask

SAC MODEL RESULTS VS BASE MODEL



Base mode perf



CONCLUSION

This project develops a DRL-based market-making agent trained in a realistic Hawkes process LOB environment. The standalone SAC agent has been successfully implemented and demonstrates strong performance against the base quoting model.

Work is currently ongoing to integrate the LSTM attention mechanism into the SAC framework, with the goal of further improving performance by capturing temporal dependencies in order flow. Looking ahead, future work will explore full order book depth modeling beyond the best bid/ask, and multi-agent settings where competing market makers interact — more accurately reflecting real-world trading dynamics.