Generating Realistic Stock Market Order Streams

Junyi Li, Xintong Wang, Yaoyang Lin, Arunesh Sinha, Michael P. Wellman
University of Michigan, Ann Arbor

Motivation

Computational Models of Financial Markets

• **Analytic Models**
  Extract highly stylized models for tractability to provide key insights;

• **Agent-based Models**
  Simulate markets as complex multi-agent systems to reproduce “stylized facts”;

• **Generative Models**
  Learn from historical market data to produce realistic and high-fidelity data to facilitate analysis;

Our goal is to generate streams of limit orders that are close in aggregate to the real market of a single stock.

Stock-GAN

Limit Order Representation

• Limit price
• Quantity
• Type (buy/sell; submit/withdraw)
• Elapsed time
• Best buy price in market
• Best sell price in market

A Conditional Wasserstein Generative Adversarial Network

Evaluation

How to Define Realistic?

• Price - distribution over price for limit orders by order type.
• Quantity - distribution over quantity for limit orders by type.
• Intensity - number of orders for a consecutive period of time.
• Inter-arrival time - distribution over inter-arrival duration for limit orders by order type.
• Best bid/ask evolution - changes in the best bid and ask over time as new orders arrive.

Synthetic Data

GOOG Data

Source: real limit-order streams obtained from OneMarketData.