Real-world Derivatives Hedging with Deep Reinforcement Learning

Ashwin Rao

ICME, Stanford University

October 28, 2019

Classical Pricing and Hedging of Derivatives

- Classical Pricing/Hedging Theory assumes frictionless markets
- Frictionless := continuous trading, any volume, no transaction costs
- Assumptions of arbitrage-free and completeness lead to (dynamic and exact) replication of derivatives with *basic* market securities
- Replication strategy gives us the pricing and hedging solutions
- This is the foundation of the famous Black-Scholes formulas
- However, the real-world has many frictions $\Rightarrow$ *Incomplete Market*
- ... where Derivatives cannot be exactly replicated

Pricing and Hedging in an Incomplete Market

- In an incomplete market, we have multiple risk-neutral measures
- So, multiple derivative prices (each consistent with no-arbitrage)
- The market/trader "chooses" a risk-neutral measure (hence, price)
- Based on a specified preference for trading risk versus return
- This preference is equivalent to specifying a Utility function
- Maximizing "risk-adjusted return" of the derivative plus hedges
- Reminiscent of the Portfolio Optimization problem
- Likewise, we can set this up as a stochastic control (MDP) problem
- Where the decision at each time step is: *Trades in the hedges*
- So what's the best way to solve this MDP?

Deep Reinforcement Learning (DRL)

- Dynamic Programming not suitable in practice due to:
  - Curse of Dimensionality
  - Curse of Modeling
- So we solve the MDP with *Deep Reinforcement Learning* (DRL)
- The idea is to use real market data and real market frictions
- Developing realistic simulations to derive the optimal policy
- The optimal policy gives us the (practical) hedging strategy
- The optimal value function gives us the price (valuation)
- Formulation based on Deep Hedging paper by J.P.Morgan researchers
- More details in the prior paper by some of the same authors

Problem Formulation

- We will simplify the problem setup a bit for ease of exposition
- This model works for more complex, more frictionful markets too
- Assume time is in discrete and finite steps $t = 0, 1, \ldots, T$
- Given a portfolio D of m derivatives, all expiring within time T
- Portfolio-aggregated *Contingent Cashflows* at time t denoted $X_t \in \mathbb{R}$
- Assume we have n basic market securities as potential hedges
- Hedge positions (units held) at time t denoted $\alpha_t \in \mathbb{R}^n$
- We assume initial hedge positions are 0 ($\alpha_0 = 0$)
- Cashflows per unit of hedges held at time t denoted $Y_t \in \mathbb{R}^n$
- Prices per unit of hedges at time t denoted $P_t \in \mathbb{R}^n$
- Problem is to determine hedging strategy and valuation of portfolio D

States and Actions

- Denote state space at time t as $\mathcal{S}_t$, state at time t as $s_t \in \mathcal{S}_t$
- Among other things, s_t contains t, α_t, P_t (perhaps also past prices)
- s_t will include any market information relevant to trading actions
- Denote action space at time t as $\mathcal{A}_t$, action at time t as $a_t \in \mathcal{A}_t$
- a_t represents units of hedges traded (positive for buy, negative for sell)

$$\alpha_t = \alpha_{t-1} + a_{t-1} \text{ for } t = 1, \ldots, T$$

- Trading restrictions (eg: no short-selling) define $\mathcal{A}_t$ as a function of s_t
- Force-liquidation at termination T means $a_T = -\alpha_T$
- State transitions available from a *simulator*, whose internals are estimated from real market data and realistic assumptions

Rewards, Utility, Hedging/Valuation

- Cashflows materialize at time t in three ways:
  - Cashflows emitted from holding positions $= X_t + \alpha_t \cdot Y_t$
  - Trading PnL $= -a_t \cdot P_t$
  - Transaction Costs, as an example could be $= -\gamma P_t \cdot |a_t|$ for some $\gamma > 0$
- Rewards r_t at time t is the Utility of the sum of above quantities, i.e.

$$r_t = U(X_t + \alpha_t \cdot Y_t - a_t \cdot P_t - \gamma P_t \cdot |a_t|)$$

- For an appropriate concave Utility function U (based on risk-aversion)
- Hedging Strategy for Portfolio D at time t given by Optimal Policy $\pi_* : \mathcal{S}_t \rightarrow \mathcal{A}_t$
- Valuation of Portfolio D at time t given by Optimal Value Function $V_* : \mathcal{S}_t \rightarrow \mathbb{R}$

DRL Approach a Breakthrough for Practical Trading?

- The industry practice/tradition has been to start with *Complete Market* assumption, and then layer ad-hoc/unsatisfactory adjustments
- There is some past work on pricing/hedging amidst frictions
- But they are mostly theoretical and not usable in real trading
- My opinion: This DRL approach a breakthrough for practical trading
- Key advantages of this DRL approach:
    - Algorithm for pricing/hedging independent of market dynamics
    - Computational cost scales efficiently with size m of derivatives portfolio
    - Enables one to faithfully capture practical trading situations/constraints
    - Deep Neural Networks provide great function approximation for RL