Pricing American Options with Reinforcement Learning

Ashwin Rao

ICME, Stanford University

November 13, 2018

Overview

1. Review of Stopping Time

2. Optimal Stopping and American Pricing as MDPs

Stopping Time

- Stopping time τ is a "random time" (random variable) interpreted as time at which a given stochastic process exhibits certain behavior
- Stopping time often defined by a "stopping policy" to decide whether to continue/stop a process based on present position and past events
- Random variable τ such that $Pr[\tau \leq t]$ is in σ-algebra $\mathcal{F}_t$, for all t
- Deciding whether $\tau \leq t$ only depends on information up to time t
- Hitting time of a Borel set A for a process X_t is the first time X_t takes a value within the set A
- Hitting time is an example of stopping time. Formally,

$$T_{X,A} = \min\{t \in \mathbb{R} | X_t \in A\}$$

eg: Hitting time of a process to exceed a certain fixed level

Optimal Stopping Problem

- Optimal Stopping problem for Stochastic Process X_t:

$$V(x) = \max_{\tau} \mathbb{E}[G(X_\tau)|X_0 = x]$$

  where τ is a set of stopping times of X_t, $V(\cdot)$ is called the Value function, and G is the Reward function.
- Note that sometimes we can have several stopping times that maximize $\mathbb{E}[G(X_\tau)]$ and we say that the optimal stopping time is the smallest stopping time achieving the maximum value.
- Example of Optimal Stopping: Optimal Exercise of American Options
  - X_t is stochastic process for underlying security's price
  - x is underlying security's current price
  - τ is set of exercise times corresponding to various stopping policies
  - $V(\cdot)$ is American option price as function of underlying's current price
  - $G(\cdot)$ is the option payoff function

Optimal Stopping Problems as Markov Decision Processes

- We formulate Stopping Time problems as Markov Decision Processes
- *State* is a suitable function of the history of Stochastic Process X_t
- *Action* is Boolean: Stop or Continue
- *Reward* always 0, except upon Stopping (when it is $= G(X_\tau)$)
- *State*-transitions governed by Underlying Price Stochastic Process

Optimal Exercise of Path-Dependent American Options

- RL is an alternative to Longstaff-Schwartz algorithm for Pricing
- *State* is [Current Time, History of Underlying Security Prices]
- *Action* is Boolean: Exercise (i.e., Payoff and Stop) or Continue
- *Reward* always 0, except upon Exercise (= Payoff)
- *State*-transitions governed by Underlying Price's Stochastic Process
- Optimal Policy $\Rightarrow$ Optimal Stopping $\Rightarrow$ Option Price