Machine Learning in Finance

Exercise sheet 11

Exercise 11.1 (American Option) We consider discrete price process $S = (S_0, S_1)$ consisting of a risk-free bond $S_t^0 = (1+r)^t$ and a risky asset $(S_t^1)_{t=1}^T$. Assume the market is complete and no arbitrage, and let Q be the equivalent martingale measure. Let θ be the self-financing strategy and $V(\theta)$ be the corresponding value process:

$$V_t(\theta) = V_0(\theta) + \sum_{u=1}^{t} \theta_u \cdot \Delta_u. \tag{1}$$

Consider the American option with payoff $f_t = (K - S_t^1)_+$ and fair price $V^*(0) = x$

(a) Assume that there exist θ^* and stopping time τ^* s.t. $V_{\tau^*}(\theta^*) = f_{\tau^*}$, prove that

$$x = \sup_{\tau \in \mathcal{T}} \mathbb{E}_Q[(1+r)^{-\tau} f_\tau]. \tag{2}$$

by optional stopping theorem. (We don't know if such θ^* and τ^* exist for now but we prove they do exist in (b)(c)(d))

(b) Let V^* be the fair price process of the American option, prove that

$$V_{t-1}^* = \max\{f_{t-1}, (1+r)^{-1}\mathbb{E}_Q[V_t^*|\mathcal{F}_{t-1}]\}, \quad t = 1, \dots, T. \tag{3}$$

(c) (Snell Envelope) If Z and X satisfies that

$$Z_T = X_T, \quad Z_{t-1} = \max\{X_{t-1}, \mathbb{E}_Q[Z_t|\mathcal{F}_{t-1}]\}, \quad t = 1, \dots, T, \tag{4}$$

then prove that Z is a supermartingale and

$$Z_0 = \sup_{\tau \in \mathcal{T}} \mathbb{E}[X_\tau]. \tag{5}$$

(d) Use (c) (with $X_t = (1+r)^{-t} f_t$ and $Z_t = (1+r)^{-t} V_t^*$) and the martingale representation theorem to prove that there exist θ^* and stopping time τ^* s.t. $V_{\tau^*}(\theta^*) = f_{\tau^*}$.

(e) Use (3) and write conditional expectation as a function of random variable to replicate the Longstaff and Schwartz method.

Exercise 11.2 (Reinforcement Learning)

(a) Assume a time discrete finite time horizon setting and prove the Dynamic Programming Principle:

$$V^*(t,x) = \sup_{\pi \in \Pi} E_{t,x}\big[V^*(t+1, x^\pi(t)) + c(t, x^\pi(t), \pi_t)\big] \tag{6}$$

(b) Consider the following optimal stopping problem:

$$\sup_{\tau \in \mathcal{T}} E\big[c(\tau, x_\tau)\big] \tag{7}$$

and formulate it as a reinforcement learning problem.

(c) Prove that under policy iteration

$$V^{\pi^{(n+1)}} \leq \mathcal{T} V^{\pi^{(n)}} \tag{8}$$

where $\mathcal{T}$ is the Bellman operator (see lecture note) and use (8) to further prove the linear convergence of policy iteration:

$$\|V^{\pi^{(n)}} - V^*\| \leq C\gamma^n \tag{9}$$

(d) Does value iteration/ policy iteration/ Q iteration need the knowledge of the underlying dynamic (the transition probability)? What can we do if we can only draw samples from doing experiment under this dynamic? What is the difference between Monte Carlo method and Temporal Difference method in reinforcement learning? What is the difference between SARSA and Q-learning?

References

[1] Dimitri Bertsekas. *Reinforcement learning and optimal control*. Athena Scientific, 2019.

[2] Robert J. Elliott and P. Ekkehard Kopp. *Mathematics of Financial Markets*. Springer, 2005.

[3] Calypso Herrera, Florian Krach, Pierre Ruyssen, and Josef Teichmann. Optimal stopping via randomized neural networks. *arXiv preprint arXiv:2104.13669*, 2021.

[4] Francis A Longstaff and Eduardo S Schwartz. Valuing american options by simulation: a simple least-squares approach. *The review of financial studies*, 14(1):113–147, 2001.

[5] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[6] John N Tsitsiklis and Benjamin Van Roy. Regression methods for pricing complex american-style options. *IEEE Transactions on Neural Networks*, 12(4):694–703, 2001.