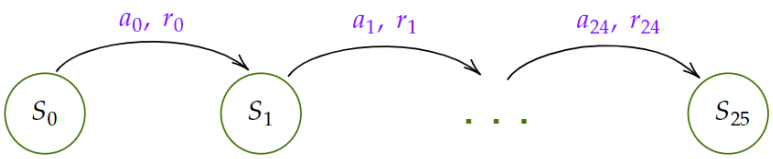
Plicy Gradient methods in the Evolutionary Pricing Game

March 30, 2023

Abstract

The deep reinforcement learning model of an agent in duopoly multi-round pricing game and applying policy gradient methods to find the optimal strategy.

1 Policy Gradient methods



f1equi

Figure 1 environment model

In *action-value* methods such as Q-learning, the value of of each state or state-action pair is learned and based on these values, the policy would be determined. However, in *Policy Gradient* methods, the policy is learnt directly without using value functions for action selection.

Since the number of states in our model is large, we need a function approximator to to parametrize the action preferences. We use a artificial neural network for this purpose and we show the vector of connection weights in our network as θ.

The actions at each state are chosen using in a way that action with higher valuation is more likely to be chosen

2 Monte Carlo Policy Gradient

3 Reinforcement with Baseline

4 Actor-Critic method