Machine Learning: Project
Multi-Agent Learning in Canonical Games and Knights Archers Zombies

March 23, 2025

Dimitrios Mystriotis - Jan Cichomski
r1027781 - r1026448

---

1 Task 1:

2 Task 2:

2.1

- A game is in Nash equilibrium when no player can improve their outcome by changing their strategy, if the other player doesn't change their's.
- A game is in Pareto Optimal when it is impossible to make a player better off without making the total payoff worse.

(a) Stag Hunt:

- Nash Equilibria: (Hare,Hare) and (Stag,Stag)
- Pareto Optimal: (Hare,Hare) and (Stag,Stag)

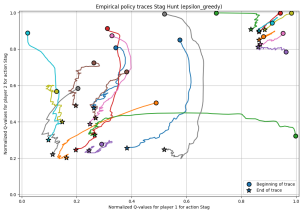
(b) Subsidy game:

- Nash Equilibria: (Subsidy 2,Subsidy 2)
- Pareto Optimal: (Subsidy 1,Subsidy 1) and (Subsidy 2,Subsidy 2)

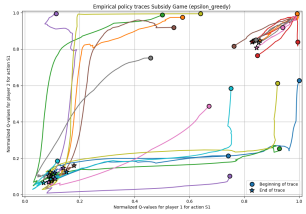
(c) Matching Pennis:

- Nash Equilibria: There is no Nash Equilibria for a pure strategy. For a mixed strategy, the Nash Equilibria is picking heads or tails with probability 0.5 each.
- Pareto Optimal: Every outcome is Pareto Optimal

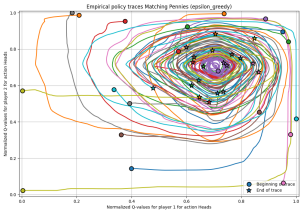
(d) Prisoner's Dilemma:



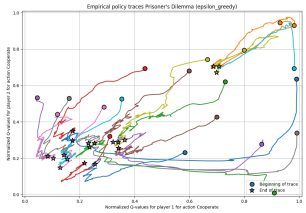
(a) e-Greedy Stag Hunt



(b) e-Greedy Subsidy Game



(c) e-Greedy Matching Pennies



(d) e-Greedy Prisoner's Dilemma

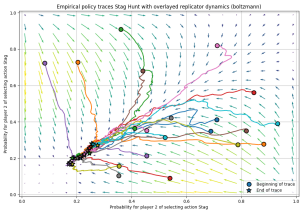
- Nash Equilibria: (Defect,Defect)
- Pareto Optimal: (Cooperate,Cooperate)

2.2

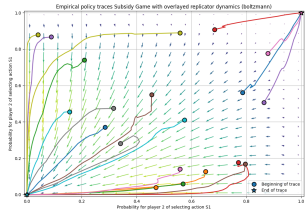
The e-Greedy algorithm converges to the Nash Equilibrium for all games. Even when an algorithm converges to the Pareto Optimal solution, it starts to move towards the Nash Equilibrium as the number of iterations increases. This is because the Nash Equilibrium is the most stable solution, which random choices will favor. The values plotted are the q-values of the agents and so each of the quarters represents a game outcome as the agents pick the action with the highest q-value deterministically.

The Boltzmann algorithm converges to either the Nash Equilibrium or the Pareto Optimal solution. The Boltzmann algorithm is more likely to converge to the Pareto Optimal solution compared to the e-Greedy algorithm, and it doesn't deviate from that point. The algorithm will basically find an optimal solution (either Nash or Pareto) and stick to it. Which on is favored depends on the starting distance from the point and the rewards of the game. It seams to favor the Nash Equilibrium slightly more than the Pareto Optimal solution.

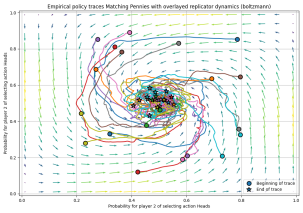
For the Lenient Boltzmann Q-Learning algorithm, the results are similar to the Boltzmann Q-Learning algorithm. The main advantage of the Lenient Boltzmann Q-Learning algorithm is that it allow for more exploration, by allowing for some suboptimal choices. This can be seen in the plots, where the algorithm is more likely to explore the state space and find the Pareto Optimal solution. The existence of the k value allow for the algorithm to explore more or less, depending on the value of k.



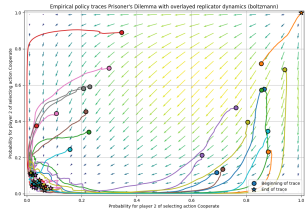
(a) Boltzmann Q-Learning Stag Hunt



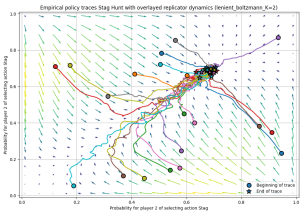
(b) Boltzmann Q-Learning Subsidy Game



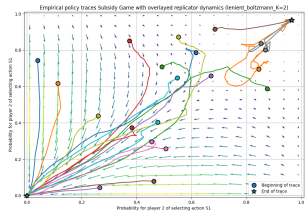
(c) Boltzmann Q-Learning Matching Pennies



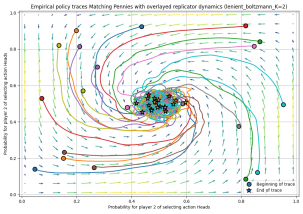
(d) Boltzmann Q-Learning Prisoner's Dilemma



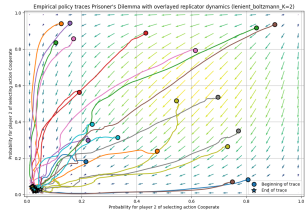
(a) Lenient Boltzmann Q-Learning Stag Hunt



(b) Lenient Boltzmann Q-Learning Subsidy Game



(c) Lenient Boltzmann Q-Learning Matching Pennies



(d) Lenient Boltzmann Q-Learning Prisoner's Dilemma

2.3

The learning trajectories are behaving as expected and do follow the replicator dynamics. For each of the games learning follow the dynamics, especially for the Lenient Boltzmann Q-Learning algorithm were the changes of the k value affects both dynamics and trajectories making the match between the two very visible.

3 Task 3:

4 Task 4: