Fictitious Play and Reinforcement Learning for Computing Equilibria in Repeated Zero-Sum Games

Your Name
Your Institution
your.email@example.com

February 13, 2025

Abstract

This project investigates computational approaches to compute equilibria in repeated zero-sum games. We compare Fictitious Play with several reinforcement learning techniques, including Q-Learning, Minimax Reinforcement Learning, and Belief-Based methods, in the context of two canonical games: stochastic Rock-Paper-Scissors (RPS) and Matching Pennies (MP). The study combines theoretical analysis with extensive simulations. Results are visualized through various performance metrics, such as cumulative scores, convergence behavior, policy evolution, and joint action frequency. Our findings highlight differences in convergence speed, learning stability, and the impact of stochastic environments.

1 Introduction

The computation of equilibria in games is a central problem in game theory and multi-agent systems. Repeated zero-sum games, where the gain of one player is exactly balanced by the loss of the other, provide an ideal test-bed for studying learning dynamics in adversarial settings. In this work, we explore iterative learning algorithms that allow players to converge toward equilibrium strategies without requiring complete analytical solutions. We focus on comparing Fictitious Play (FP) with reinforcement learning (RL) methods—including Q-Learning, Minimax RL, and Belief-Based approaches—in two well-known games: Rock-Paper-Scissors and Matching Pennies.

2 Theoretical Background

2.1 Repeated Zero-Sum Games

Repeated zero-sum games consist of a stage game that is played over several episodes. In each round, both players select actions simultaneously. The payoff for one player is the negative of the other player's payoff, ensuring that the total payoff sums to zero. The repetition of the game allows players to learn from past outcomes, adapt their strategies, and potentially converge to a Nash equilibrium or a minimax solution. These games serve as a foundational model for adversarial interactions in economics, security, and machine learning.

2.2 Learning Agents

Several learning algorithms have been proposed to compute equilibria in repeated games:

- **Fictitious Play (FP):** Players assume that opponents will play according to the empirical frequency of their past actions. They then choose the best response to these beliefs.
- **Q-Learning (QL):** An RL method where agents learn the expected utility of actions using a Q–table and update it iteratively via temporal difference learning.
- **Minimax RL:** A variation of Q–Learning adapted for adversarial settings, where agents aim to maximize the minimum gain against a worst-case opponent.
- **Belief-Based Methods:** Agents update probabilistic beliefs about the opponent's actions and choose their best response accordingly.

3 Implementation

3.1 Environment Setup

Two game environments were implemented in Python:

- **Stochastic Rock-Paper-Scissors (RPS):** In this environment, the game is played in two states with different payoff matrices. A state transition function introduces stochasticity, altering the stage game dynamically.
- **Matching Pennies (MP):** A non-stochastic repeated zero-sum game with a fixed payoff matrix, where the payoff for one player is the negative of the other.

Each environment is encapsulated within a Python class that defines the state, payoff matrices, and state transition rules.

3.2 Experimental Design

Experiments were conducted by pitting every pair of the four agents (FP, QL, Minimax RL, Belief-Based) against each other in both game environments. For each experiment, multiple trials were executed, each spanning thousands of episodes. During these simulations, various performance metrics were recorded, including:

- Moving average rewards.
- Cumulative scores over episodes.
- Q–value evolution and convergence.
- Policy evolution and learning stability.
- Joint action frequency.

The simulation results were exported to CSV files and subsequently visualized using an interactive Python Dash dashboard.

4 Results & Discussion

4.1 Convergence of Different Approaches

The convergence behavior of the different algorithms varied considerably. Fictitious Play exhibited smooth convergence to the Nash equilibrium, while RL methods (Q-Learning and Minimax RL) showed higher variance and occasional oscillatory behavior before stabilization. Belief-Based methods benefited from prior knowledge of the payoff matrix and converged relatively fast.

4.2 Comparative Performance in Stochastic RPS

In the stochastic RPS environment, average cumulative scores indicated that FP and Belief-Based agents performed similarly, consistently approaching equilibrium strategies. In contrast, Q-Learning and Minimax RL displayed slower convergence with higher variance, highlighting the challenges posed by the stochastic transitions between states.

4.3 Comparative Performance in Matching Pennies

For the fixed Matching Pennies game, all agents eventually converged to equilibrium strategies. However, differences in the speed of convergence and the transient performance were observed. RL methods required more episodes to stabilize due to exploration, while FP demonstrated rapid adaptation.

4.4 Policy Evolution and Learning Stability

Visualizations of policy evolution revealed that FP maintained a steadily refining strategy distribution over episodes. In contrast, RL methods often underwent abrupt strategy changes due to the exploration-exploitation trade-off. The convergence charts of Q–values confirmed that RL methods eventually reduce exploitability, although at different rates.

4.5 Joint Action Frequency Analysis

Joint action frequency heatmaps provided insight into the interaction dynamics between agents. In equilibrium, the joint action distributions approached uniformity in RPS, consistent with the theoretical Nash equilibrium. In Matching Pennies, the distribution reflected the symmetric nature of the game, validating the robustness of the learning process.

5 Conclusion

This project demonstrates that iterative learning methods can effectively compute equilibria in repeated zero-sum games. Fictitious Play offers stable and rapid convergence in many cases, while reinforcement learning methods—although initially more volatile—eventually yield robust strategies. The experiments underscore the significance of adapting learning algorithms to the game environment and reveal that stochasticity plays a critical role in determining convergence dynamics.

Key Takeaways

- Iterative learning methods, such as FP and RL, are viable for computing equilibria in repeated zero-sum games.

- Fictitious Play generally converges stably, whereas RL methods require careful tuning of exploration parameters.
- The stochastic nature of an environment can significantly affect the speed and stability of convergence.

Future Work

Future research directions include:

- Extending the analysis to general-sum and multi-agent games.
- Enhancing exploration strategies in RL to improve convergence rates.
- Applying these learning methods to real-world scenarios characterized by high uncertainty and dynamic interactions.

References

1. Fudenberg, D. and Levine, D. K. (1998). *The Theory of Learning in Games*. MIT Press.
2. Shapley, L. S. (1953). Stochastic Games. *Proceedings of the National Academy of Sciences*, 39(10), 1095–1100.
3. Littman, M. L. (1994). Markov Games as a Framework for Multi-Agent Reinforcement Learning. In *Machine Learning Proceedings 1994*, 157–163.
4. Busoniu, L., Babuska, R., and De Schutter, B. (2008). A Comprehensive Survey of Multiagent Reinforcement Learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2), 156–172.