Modelling Ecological Populations

Game Theory Project

Ganga & Deddy
EE15B025 & EE15B125

Dynamic Games : Theory and Applications
IIT Madras

April 27, 2019

Schedule

Overview

Reinforcement Learning

- Provides a formalism for behavior
- Obtained from behavioral psychology
- Helpful for modelling ecological populations

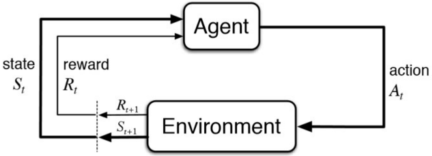

Figure: Reinforcement Learning

Methods in Reinforcement Learning

- Policy Gradient
  - Players have policies (actions)
  - Optimize in the policy space
  - Gradient Ascent
  - Episodic reward
  - $\pi(a_i, \theta)$ = Policy parameterized by θ.
    θ represents the parameters of our neural network.
  - $$\Delta\theta = \alpha_t r_r \frac{d}{d\theta}\pi(a_t, \theta_t)$$
- Multi-Armed Bandits
  - Players pick from k arms
  - Find the best arm to pull

Hawk-Dove Game

- What is it?
  - Models interaction within same species
  - Sharing of resources
- Pay-off matrix :

	Hawk	Dove
Hawk	$\frac{B-C}{2}, \frac{B-C}{2}$	$B, 0$
Dove	$0, B$	$\frac{B}{2}, \frac{B}{2}$

- B < C ; (B=6, C=10 in our expts)

	Hawk	Dove
Hawk	-2, -2	6, 0
Dove	0, 6	3, 3

- The pay-off of player i is denoted by $u_i(s_i, s_j)$

Nash Equilibria : Hawk-Dove Game

- 3 nash equilibria
- 2 pure + 1 mixed

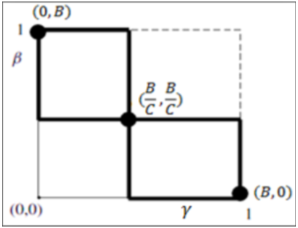

Figure: Nash Equilibria in a Hawk-Dove Game[2]

Modified Hawk-Dove Game

- A population of N players
- Each player can be a hawk or a dove
- Pay-off decided based on interaction with population
- Pay-off of player i in the population is denoted by $u_i(s_i, s_{-i})$

Figure: N-player hawk-dove game

From RL Perspective

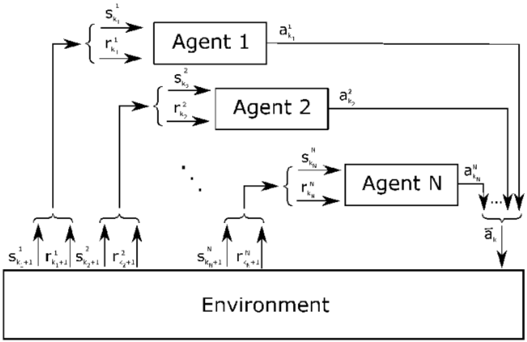

Figure: N-player hawk-dove game (Ref : MARL)

Measuring individual pay-off

1. Playing against the field

$$u_i(s_i, s_{-i}) = \frac{1}{N} \sum_{\forall j \neq i} u_i(s_i, s_j)$$

2. Playing against a group M_j

$$u_i(s_i, s_{-i}) = \frac{1}{|M_j|} \sum_{j \in M_j} u_i(s_i, s_j)$$

3. Pair-wise contest
(Player j chosen randomly by nature)

$$u_i(s_i, s_{-i}) = u_i(s_i, s_j)$$

A better understanding of its significance during convergence

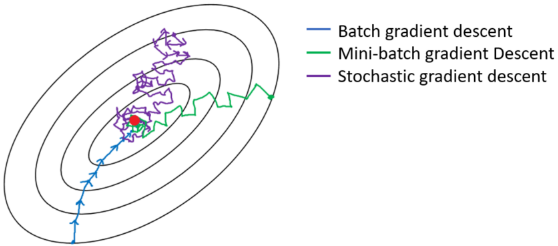

Figure: Convergence comparison of the three methods of calculating individual pay-offs : Playing against the field, Playing against a group and Pair-wise contest respectively

Some Definitions

- **Static Games**: A static game is one in which all players make decisions (or select a strategy) simultaneously, without knowledge of the strategies that are being chosen by other players. Even though the decisions may be made at different points in time, the game is simultaneous because each player has no information about the decisions of others; thus, it is as if the decisions are made simultaneously.
- **Stage Games**: A Stage Game is a game that arises in certain stage of a static game. In other words, the rules of the games depend on the specific stage. The prisoner's dilemma is a classic example of stage game

Models

NN Model Multi-brain

- Selfish agents
- Policy Gradient update
- Players have stochastic strategies, but play pure strategies

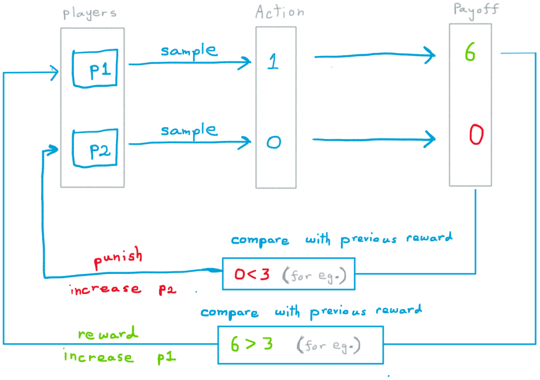

Figure: RL mechanism for pairwise contests

Pairwise contests

1. Players are matched randomly
2. Strategies drawn from Bernoulli distribution

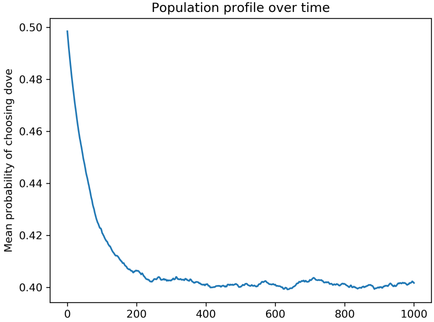

Figure: State of population over time: pairwise contests

Playing against Field

1. Strategies drawn from Bernoulli distribution
2. Payoff obtained against population profile
3. Population converges faster (sort of)

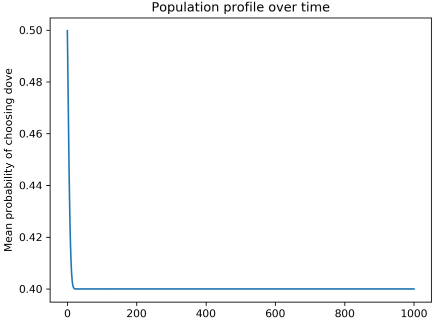

Figure: State of population over time: against the field

NN Model - Code of Conduct

- Players still selfish...
- But agree to a "code of conduct" or Rules of Engagement (RoE)
- Code of conduct updated by each player in turns

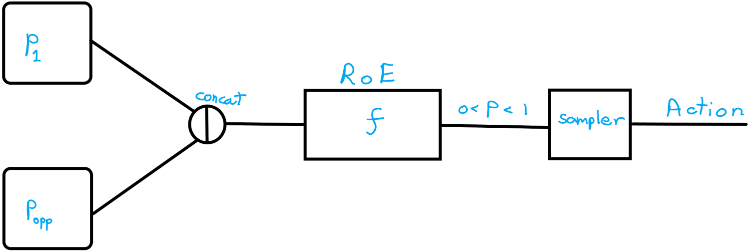

Figure: Depiction of a game with code of conduct

Aritificial Neural Network

- Parameterized function
- Can be tweaked by players

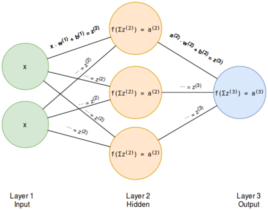

Figure: Example neural network

Playing one vs one

1. Players are matched one on one randomly by Nature
2. Players update RoE and display state through experience

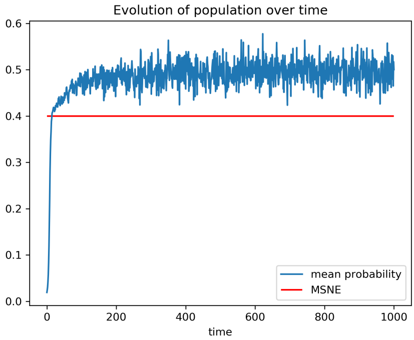

Figure: State of population (RoE) over time: pairwise contests. Some amount of inherent forced cooperation observed resulting in a population pay-off higher than MSNE

Playing against the Field

- Again, final population profile not MSNE.
- Lower variance during steady state.

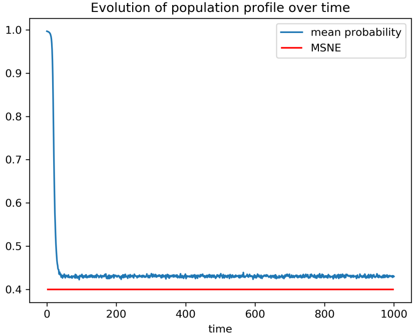

Figure: State of population (RoE) over time: against field. Some amount of inherent forced cooperation observed resulting in a population pay-off higher than MSNE

Multi-Armed Bandits

- A person must choose between multiple actions (originally comes from the idea of slot machines, the "one-armed bandits"), each with an unknown reward.

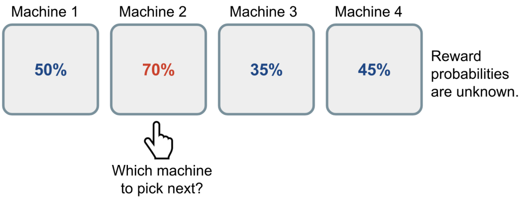

Figure: Multi-Armed Bandit Problem

- Goal : determine the best or most profitable outcome through a series of choices.
- At the beginning of the experiment, when odds and payouts are unknown, the gambler must determine which machine to pull, in which order and how many times.

UCB Model

Upper Confidence Bound Algorithm :
(For a single player)

Initialization: Play each arm once,

For t = K + 1, . . . , n, **repeat**

(1) Play arm $I_t = argmax_{k=1,\ldots,K} UCB_t(k)$, where

$$UCB_t(k) = \hat{\mu}_k(t-1) + \sqrt{\frac{8 \log t}{T_k(t-1)}}$$

(2) Observe sample X_t from the distribution P_{I_t} corresponding to the arm I_t.

Playing against Field

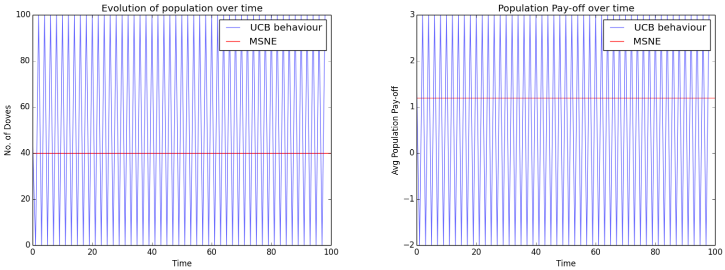

Figure: Evolution of Population over time and the average pay-off of the population over time when the population is initialized randomly with probability 0.5 (Equivalent to individual pay-off over time after convergence in this case). **Each player interacts with everyone else in the population.**

Playing against Field

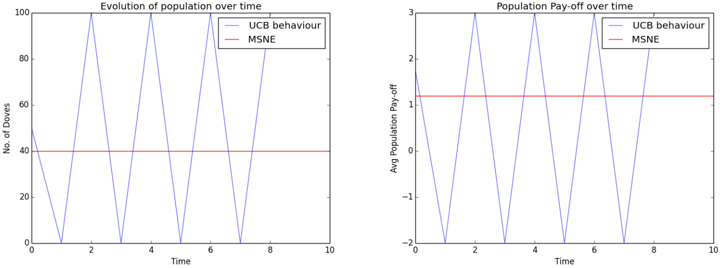

Figure: Evolution of Population over time and the average pay-off of the population over time when the population is initialized randomly with probability 0.5 (Equivalent to individual pay-off over time after convergence in this case). **Each player interacts with everyone else in the population.**

Playing Against a Group

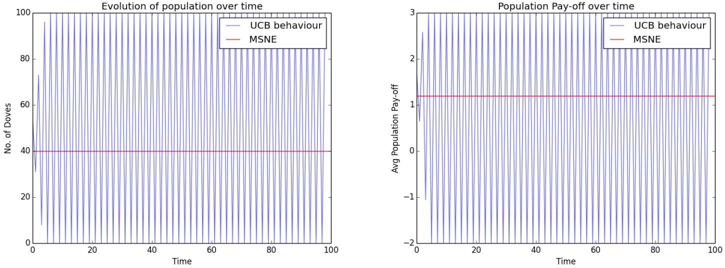

Figure: Evolution of Population over time and the average pay-off of the population over time when the population is initialized randomly with probability 0.5 (Equivalent to individual pay-off over time after convergence in this case). **Each player interacts with m<10% of the population.**

Playing Against a Group

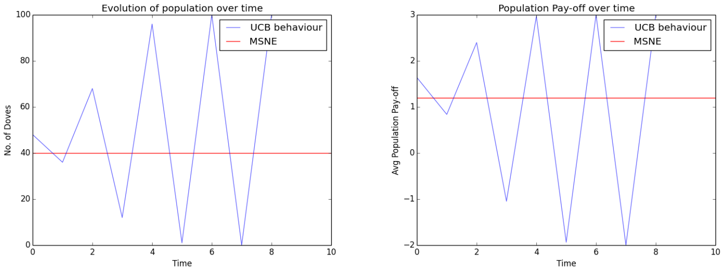

Figure: Evolution of Population over time and the average pay-off of the population over time when the population is initialized randomly with probability 0.5 (Equivalent to individual pay-off over time after convergence in this case). **Each player interacts with m<10% of the population.**

Pair-wise contest

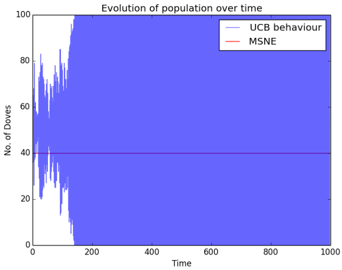

Figure: Evolution of Population over time and the average pay-off of the population over time when the population is initialized randomly with probability 0.5 (Equivalent to individual pay-off over time after convergence in this case). **Every player interacts with another random player - one vs one.**

Pair-wise contest

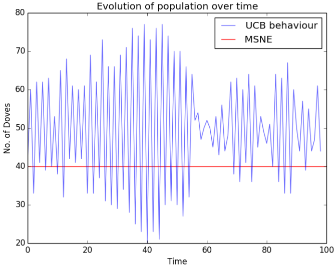

Figure: Evolution of Population over time and the average pay-off of the population over time when the population is initialized randomly with probability 0.5 (Equivalent to individual pay-off over time after convergence in this case). **Every player interacts with another random player - one vs one.**

A better understanding of its significance during convergence

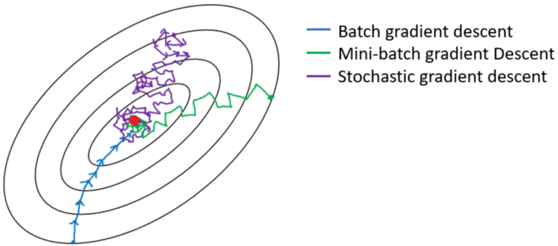

Figure: Convergence comparison of the three methods of calculating individual pay-offs : Playing against the field, Playing against a group and Pair-wise contest respectively

Summarizing UCB experiments

- Observations :
  - In all 3 cases, the population converges to a cycle (either all Hawk or all Dove)
  - In all 3 cases, the average population pay-off converges to a cycle (either -2 or +3)
  - The convergence rate of the 3 methods similar to Full Batch GD, Mini Batch GD and SGD
- Inference :
  - Playing against Field : When majority of the **current** population is Dove(>40%), better to be a Hawk.
  - Playing Against a Group : When majority of the **sampled** population is Dove(>40%), better to be a Hawk
  - Pair-wise contest : When **he**'s a Dove, I'm better off as Hawk
- Reason :
  - **Each player in each iteration chooses best response**

Group Play

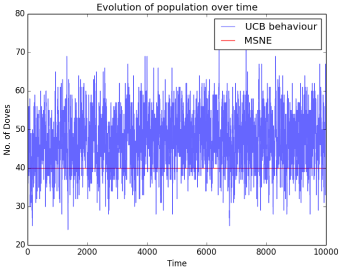

Figure: Evolution of Population over time and the average pay-off of the population over time when the population is initialized randomly with probability 0.5. **A group of m=10% of the population interacts in each interaction.**

Group Play

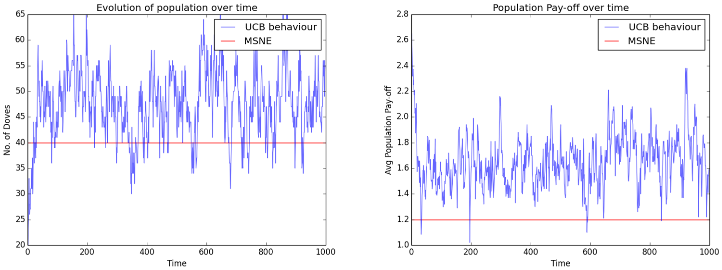

Figure: Evolution of Population over time and the average pay-off of the population over time when the population is initialized randomly with probability 0.5. **A group of m=10% of the population interacts in each interaction.**

Group Play

- Observation :
  - Fairly robust to different population initialization techniques :

Initialization	Avg No. of Dove	Avg population pay-off
All Hawk	48	1.76
All Dove	47	1.78
Random (p=0.5)	47	1.78

  - Average population pay-off better than MSNE pay-off
- Reason :
  - The change in population distribution is minimal

Future Work

Future Work

- Asymmetric Games (eg) Trust-Cooperate
- Strange attractors to analyse chaotic populations
- Quantifying rewards of cooperation
- Informed Reinforcement learners : use communication through revelation schemes

Previous Works

Axelrod - Evolution of Cooperation

- Also used to analyze behavior of populations
- made use of evolutionary programming

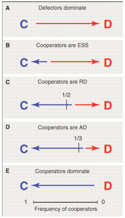

Figure: 5 stages of the evolution of cooperation

Evolutionary Game Theory and Multi-Agent Reinforcement Learning

- Authors : Karl Tuyls and Ann Nowe
- Survey the basics of RL and (Evolutionary) Game Theory
- Multi-Agent Systems
- Mathematical connection between MARL and Evolutionary Game Theory
- Ref : **Paper pdf**

References

References

- Jessica L. Barker. "Robert Axelrod's (1984) The Evolution of Cooperation". In: **Encyclopedia of Evolutionary Psychological Science**. Ed. by Todd K. Shackelford and Viviana A. Weekes-Shackelford. Cham: Springer International Publishing, 2017, pp. 1–8. ISBN: 978-3-319-16999-6. DOI: 10.1007/978-3-319-16999-6_1220-1. URL: https://doi.org/10.1007/978-3-319-16999-6_1220-1.
- Essam EL-Seidy. "On the Behavior of Strategies in Hawk-Dove Game". In: (2016), pp. 1–8. DOI: 10.1007/978-3-319-16999-6_1220-1. URL: https://doi.org/10.5923/j.jgt.20160501.02..
- Richard S. Sutton and Andrew G. Barto. **Reinforcement learning - an introduction**. Adaptive computation and machine learning. MIT Press, 1998. ISBN: 0262193981. URL: http://www.worldcat.org/oclc/37293240.

THANK YOU :)