Research of Parameters in Strategies of the Iterated Continuous Prisoner's Dilemma

Adrian Hossner

Contents

Abstract

1 Introduction

- Interactions
  implication range is wide:
  from two persons in a relationship to countries in alliance
  from two impala's algrooming to zebras warning the herd about a lion

- Key hook:
  helping someone with possible cost – cooperation
  being selfish but breaking trust – defection

- model PD:
  single interaction
  already well analysed
  Nash found solution/equilibrium

- Iterated:
  more interesting, insightful
  strategies

- non-suited variants:
  Axelrod's Tournament
  Evolution

- continuous:
  more complex, accurate

- parameters:
  determines behaviour
  create surface
  the part that has not been experimented

- overall and difference:
  overall: wealth of population
  difference: competence between individuals

2 Theoretical Foundations

- Game Theory (optional, not needed for understanding)
  mathematical framework to investigate interactions

- Prisoner's Dilemma:

The Prisoner's Dilemma (PD) is a famous intellectual game in Game Theory. It was invented by Merrill Flood and Melvin Dresher in 1950. In order to understand the dilemma, a proper context has to be established. Two persons are arrested and are sued for having stolen something. The police, however, has not enough evidence to imprison them. As a consequence of that, the police interrogates the two criminals. They can either say nothing or confess that the other suspect was involved in the crime. They can either cooperate or defect without having the possibility to talk to each other before the decision. A certain tactic is being used by the police officers. They make it profitable for the criminals to confess. They let the criminal out of prison earlier if he/she confesses. But if the other also confesses, both stay long in prison. If he/she says nothing and the other confesses, he/she will stay very long in prison. If, however, both stay silent, both of them go to prison in a relatively short amount of time. The following matrix visualises the pay-off's in this game in a comprehensive way:

	C	D
C	3, 3	10, 0
D	0, 10	5, 5

C stands for cooperation and D stands for defection. Cooperation is defined by remaining silent whereas defection means confessing. The dilemma consists of the following. Confessing seems attractive since the interrogated criminal can walk out freely without going to prison. However, if the second criminal also confesses, both get five years in prison. This is, nevertheless, the worst outcome for both which could have been avoided as both could have stayed silent.

John Nash, a mathematician which made great contributions in Game Theory, has proved that it is the most logical option for both to confess, always. A Nash-equilibrium is introduced. It defines a stable state in which both would not change their decision even though they would know if the other cooperated or defected.

This game composes a one-time interaction very well in a mathematical and analytical perspective. The pay-off's in the matrix can be changed as long as these rules stay valid.

	C	D
C	R, R	S, T
D	T, S	P, P

$$T > R > P > S$$
$$2 * R > T + S$$

- Iterated:

  The Iterated Prisoner's Dilemma (IPD) is a game which extends the PD. As the name suggests, the PD is played a number of times sequentially. After each round the pay-off gets accumulated to the points one strategy has already gained. Since a strategy wants to gain as many points as possible, the pay-off matrix should be altered accordingly.

	C	D
C	3, 3	0, 5
D	5, 0	1, 1

  The Nash-equilibrium in the IPD differs from that in the PD. This means that cooperation can earn out long-term. By cooperating, one is building trust with another. One strategy can only get the maximum points, which can be gained by always getting the 5 points each round, by defecting and the opponent's cooperation. Since the opponent will likely not cooperate after one strategy always defects, the maximum is practically impossible to receive. So, the relative high amount and the possibility of getting it for both is by cooperating.
  The number of round that will be played has to be unknown to both strategies. There is no sense in cooperating the last round since trust does not matter after the game is ended. And because trust will be given up in the last round, there is no sense for both to keep trust the second last round. This will end up in an inductive defection behaviour for both strategies the whole game.
  This extension of the PD allows us to simulate a relationship over time where the two players interact multiple times.

- Strategies:

  A strategy is to be considered as an autonomous player which has its own behaviour. A strategy has the whole history of its own contribution and the ones of the opponent at hand to generate the decision for the next round. These decisions are calculated by using conditions and probabilities on the given data.

- Continuous:

  The continuous variant is considered more accurate to the real world since the decisions are rather rare only either cooperation or defection. The pay-off system needs to be altered accordingly. The pay-off of [sciencedirect.com] is suitable for this project.

  $$p1 = y - c * x$$
  $$p2 = x - c * y$$

  The amount of cooperation is called investment and can range from zero to one (0 to 1) where full defection is represented by 0 and cooperation corresponds to 1. The variable x is the investment of strategy 1 and y is the investment of strategy 2. The parameter c also ranges from 0 to 1 and describes the cost of cooperation.

- Noise:

  Noise in the ICPD is the equivalent to miscommunication or misunderstanding in the real world. In this variant of the PD, noise is essential to trigger interesting outcomes. When a game of the ICPD is started and two strategies who start with a full investment, many strategies only respond with full investments. So noise is necessary to trigger continuous investments. Without noise, in many cases, the ICPD would be equivalent to the IPD.

- Simultaneous vs Alternating:

  Two main differences can be seen when simulating an iterated variant of the PD. On one hand there is the simultaneous form whereby the strategies submit their contribution at the same time. On the other hand, there is the alternating form in which on e strategy submits its contribution and the other strategy can respond to this contribution is that round. Since the alternating form gives and advantage to the strategy which is allowed to respond, this only makes things more complicated than necessary. So, this analysis is only dedicated to the simultaneous form.

- Current Findings:

  In the simpler variants, e.g. the IPD without noise, it has been proved many times over that the strategy Tit-For-Tat is the most successful one. [sources] The success of strategies, however, can vary depending on certain conditions such as the influence of noise or different pay-off systems where continuous investments can be submitted. So, the very specific area of the simultaneous ICPD, has not been very well explored. ...

3   Methods and Implementation

- Parameter-based:

  The main idea of this paper is to use parameter-based strategies. These strategies hold a parameter that defines the behaviour of the strategy. The parameter will always range from 1 to 10 inclusively. [Example]

- Surfaces:

  The end result of this paper will be several surface plots. The data will be generated by letting two parameter-based strategies play against each other the ICPD. The game will be structured so that every parameter came against every other parameter in the ICPD. Like this, a surface can be plotted by having the x-axis being the parameter of strategy 1 and the y-axis being the parameter of strategy 2. The z-axis will indicate the points one strategy gained. That being said, one surface will be shown for each strategy. Further more, the two z-axis of the two surfaces can be added and form a new surface which shows the overall points gained by both strategies. The complement to that would be to subtract the two z-axis to plot a surface showing the difference between the two surfaces. This describes how much better one strategy was than the other. So, in the end there will be four surface plots per one ICPD.

- Simulated Strategies and explanation of parameter:

  My selection:
  In this simulation, strategies have to be defined and selected. There are two main points I want to investigate in this project. First, I want to associate the strategies to specific groups that differ in the characteristic of each member. I suggest the following three characteristics:
  1. Responsive
  2. Rigid
  3. Random
  I chose these three qualities due to the fact that one has to follow one of these three natures. In the real world, every personality can be associated with one of my defined groups. It can also be looked at this from another perspective. Whenever one has to respond to another's action, one has to decide how to answer. The shift applied to the previous investment can be categorised in one of these three groups. Either the shift is conditional to the opponent's last contribution, then this strategy is responsive. Or the shift is always 0 then the behaviour falls into the group of the rigid. This means that the investment will unconditionally stay the same. Or the shift is neither one of the above, meaning it doesn't depend on the past interactions nor stays it always 0. Then there cannot any determinenating aspect be established, meaning it is random.

Second, the question arises if there is a difference between the ICPD and the IPD. So, I propose a discrete and a continuous variant of each group. With the exception of the group 'Rigid', there will be two members of each group. It will be clarified why only one strategy in this group makes sense. Thus, we can analyse the difference between the two members of the group to clarify which variant gained more points in the ICPD. Having two strategies in each group except in the group 'Rigid', means implementing five strategies. And letting all strategies play against every other strategy, including itself, means having enough data to plot $4 \cdot \sum_{a=1}^{5}$ surfaces. In total, this would equal to 60 surface plots.

The following strategies are described by algorithmising them. $i(\theta)$ is a function that corresponds to the investment in the current round and holds the variable θ which is the parameter of the strategy in this round. I showcase what the behaviour of the strategy will be equivalent to when the parameter is 0, 5 and 10 for an easier understanding.

Average:

The strategy Average (Avr) belongs to the group of the responding/responsive strategies. It takes the investments of the opponent in the last θ_{Avr} previous rounds as variables to its function to calculate the mean. This mean will then be submitted in the next round. Since in the first θ_{Avr} rounds the strategy cannot calculate an average, it will submit full cooperation to offer maximum points for both of them.

$$i(\theta_{\text{Avr}}) = \frac{1}{\theta_{\text{Avr}}} \sum_{k=1}^{\theta_{\text{Avr}}} \bar{i}_k$$

The variable $\bar{i}_k$ is the k'th element of the history of the opponent's investments, $k = 1$ being the last round. This strategy, Average, with the parameter 0 is equivalent to since it only calculates the mean of the last opponent's submission.

Adapt:

Adapt (Adp) is an, as the name suggests, adaptive and thus responsive strategy. It starts with full cooperation to offer a constructive relationship within this simulation. After having played one round, it will adapt to the submissions of the opponent by shifting its own investment towards the opponent's.

$$i(\theta_{\text{Adp}}) = i_1 + s(\theta_{\text{Adp}})$$

$$s(\theta_{\text{Adp}}) = \frac{\theta_{\text{Adp}}}{5} \cdot (\bar{i}_1 - i_1)$$

The shift being applied to its own previous investment is notated with the function $s(\theta_{\text{Adp}})$. This strategy with the parameter 0 is identical to AlwaysCooperate since the difference and thus the shift is multiplied by 0. Dividing θ_{Adp} by 5 implies the fact that this coefficient to the difference

is equal to one if the parameter equals five. This means that the strategy will shift its next investment to exactly the previous investment of the opponent. This behaviour corresponds to Tit-For-Tat. $\theta_{\text{Adp}} = 10$ also shows an interesting nature of the strategy. By having the parameter set to ten, it means Adapt will add the shift to its own previous investment twice. Of course, one strategy is not allowed to submit any investment that exceeds the limits that is defined to be from 0 to 1. The implementation of this strategy will simply set its investment to the reached limit if surpassed.

AlwaysSame:
AlwaysSame's (AlwS) parameter calculates the investment very simply. The investment follows the equation:

$$i(\theta_{\text{AlwS}}) = \frac{\theta_{\text{AlwS}}}{10}$$

This means that after each incrementation of the parameter, the investment increases by 0.1. Parameter 0 is equivalent to AlwaysDefect and parameter 1 has the same behaviour as AlwaysCooperate. If the parameter is set to five, the strategy will always submit the investment 0.5 which is identical to Neutral.

Random-Discrete:
Random-Discrete (RndD) is a strategy in the group of the random. It is also discrete, meaning it can only submit either 0 or 1. The parameter in this strategy determines the likelihood of submitting 0 and 1 respectively. The probability of submitting full defection is described in the following equation.

$$i(\theta_{\text{RndD}}) = \Pr(i = 1 \mid \theta_{\text{RndD}}) = \frac{\theta_{\text{RndD}}}{10}$$
$$\text{or}$$
$$i(\theta_{\text{RndD}}) = \Pr(i = 0 \mid \theta_{\text{RndD}}) = 1 - \frac{\theta_{\text{RndD}}}{10}$$

Parameter 0 is equivalent to AlwaysDefect since the probability of submitting 0 is 0%. And on the other hand, parameter 10 is equivalent to AlwaysCooperate since the probability of submitting 1 is 100%.

Random-Continuous:
The other strategy in the group of the random is oppositely continuous, meaning it can additionally submit any number between 0 and 1. The name of this strategy is logically Random-Continuous (RndC). This strategy has a base being at 0.5. The parameter defines the shift being applied to the base, calculated by the equation:

$$i(\theta_{\text{RndC}}) = 0.5 + \epsilon \cdot s(\theta_{\text{RndC}})$$

$$s(\theta_{\text{RndC}}) = \frac{\theta_{\text{RndC}}}{20}$$

$$\Pr(\epsilon = 1) = \Pr(\epsilon = -1) = \frac{1}{2}$$

9

s means the shift. It is divided by 20 so the maximum shift, parameter being 10, can not exceed the range of 0 and 1. There is a 50% probability of subtracting or adding this shift, indicated by the symbol ϵ.

	$\theta = 0$	$\theta = 10$
Random-Discrete	AlwaysDefect	AlwaysCooperate
Random-Continuous	Neutral	Random-Discrete ($\theta = 5$)
Always-Same	AlwaysDefect	AlwaysCooperate
Average	Tit-For-Tat	-
Adapt	AlwaysCooperate	-

- Programming language:

  The simulation, data generation and visualisation is completely written in Python (3).

- Data Generation:
  two strategies
  every parameter against every parameter
  play several times to smooth out noise (average)

- Implemented Variants:
  Prisoner's Dilemma
  Iterated Prisoner's Dilemma
  Iterated Continuous Prisoner's Dilemma
  Parameter-Based Strategies
  only PBS important for MA

4   Results

- Surfaces:
  plots
  surface of strategy 1
  surface of strategy 2
  surface of both (added)
  surface showing difference (subtracted)

- Analysis:
  hills and valleys

10

5 Discussion

- ?

6 Conclusions and Outlook

- Summary and Quintessence
  which strategy 1 better than which strategy 2
  which parameter to be used

- Application in the Real World
  strategies translated to real persons
  giving examples (Politics, economics to simple human beings)
  prospect to the future (which ones will win?)

7 Self Reflection

- Challenges
  estimation of time consumption
  transforming ideas during process
  dedication

- Learnings
  reading professional papers
  writing English
  experimental workflow (usage of neovim, tmux and terminal)

8 Annex