Prisoner's Dilemma and Donation game

Andreea Patra | CID 01365348

Contents

1 Abstract

This project investigates Iterated Prisoner's Dilemma game and Donation game. Firstly, we want to understand results derived from the Axelrod tournament. Then, we take a more technical approach and analyse the behaviour of TIT FOR TAT, ALWAYS DEFECT and ALWAYS COOPERATE according to Donation game which is a special case of Iterated Prisoner's Dilemma. In the last part, we present the Zero-Determinant strategy and compare it to TIT FOR TAT.

2 Axelrod's tournament[1]

In 1974, Robert Axelrod organised a computer tournament, where he invited 14 scientists to send the strategies for an iterated Prisoner's Dilemma. They would have played for 200 rounds against each other. The frame of the Prisoner's Dilemma game was (T,R,P,S)=(5,3,1,0). If both players would have always cooperated, the total score would be 600 each and if they would constantly defect, the total score would be 200. They would also be competing with a copy of their strategy and with another program, called RANDOM, which defects or cooperates with equal probability. At the end of the tournament, the programs will be ranked according to the total number of scores.

TIT FOR TAT is a strategy that cooperates on the first move and then copies what the other player did on the previous move. It was submitted by Anatol Rapoport and it has won the first place.

On the second place, it was the strategy submitted by T. Nicolas Tideman and Paula Chieruzzi. It begins with cooperation and TIT FOR TAT. However, when the other player finishes his second run of defections, the number of defections is increased by one with each run of the other's defections.

The third place belongs to Rudy Nydegger who plays TIT FOR TAT on the first three moves. However, if it was the only one to cooperate on the first move and the only one to defect on the second one, then it will defect on the third move, regardless of the other's player move. Then, the next move is chosen according to the previous three moves as follows. It initialises A as the sum formed by counting the other's defections as two points a his own defection as one point. It attributes weights 16,4,1 to the previous three moves in chronological order. It will defect if A is 1,6,7,17,22,23,26,29,30,31,33,38,39,45,49,54,55,58 or 61.

The fourth place was attributed to Bernard Grofman whose strategy cooperates with probability $\frac{2}{7}$ when the other player did different things on the previous move, otherwise it will always cooperate.

Martin Shubik won the fifth place with strategy that cooperates until the other defects and then defects once. For each departure from mutual cooperation, the strategy will add one more defection.

On the sixth place, it was the strategy submitted by William Stein, which cooperates on the first 4 moves and then plays TIT FOR TAT every 15 moves. It will check it the other player is playing randomly using a Chi-Squared Test of the transition probabilities.

On the seventh place, it was the strategy submitted by James W. Friedman which cooperates until the other defects and then it will defect until the end of the game.

Winning eighth place, Morton Davis plays similarly to James Friedman, however he will cooperate on the first 10 moves, regardless of the other player's move.

On the ninth place, there was James Graaskamp with his strategy that plays TIT FOR TAT for 50 moves, defects on move 51 and then plays TIT FOR TAT for 5 more moves. Then, it will check if the other player is RANDOM and if it is, it will defect until the end of the game. It will also check for TIT FOR TAT or its own twin and if it is, it will play TIT FOR TAT until the end of the game. Otherwise, it will randomly defect 5 to 15 moves.

Tenth place is a strategy submitted by Leslie Downing whose choice of move aims to maximise its own long term expected payoff on the assumption that the other player's strategy is that it cooperates with a fixed probability which depends only on whether the other player cooperated or defected on the previous move. Probabilities are constantly updated throughout the game and they are initially set to 0.5.

Eleventh place is a strategy submitted by Scott Feld which plays TIT FOR TAT and gradually lowers its probability of cooperating according to the other player's cooperation to 0.5 by the two hundredth move. It will always defect after a defection.

Twelfth place is a strategy submitted by Johann Joss which, after the other player cooperated, it cooperates 90% of the time and always defects after a defection.

Thirteenth place is a strategy submitted by Gordon Tullock which cooperates on the first 11 moves and the it will cooperate 10% less than the other player has cooperated on the previous move.

Fourteenth place is a strategy submitted by a graduate student which has the probability of cooperating, P, updated every 10 moves. It is initially stated to be 30% and then it is adjusted if the other player seems random, very cooperative or very uncooperative. Due to the complexity of the adjustment process, to most of the other players, this strategy seemed random having the probabbility of cooperating in the range $30\% - 70\%$.

In a variant of the preliminary tournament, TIT FOR TAT won first place. These results were sent to the contestants, therefore, they have tried to make improvements on this strategy. However, more complex algorithms did not perform as well as the original TIT FOR TAT. An important rule that separated the top 8 strategies from the other is niceness. A strategy is nice if it is not the first one to defect. Robert Axelrod has introduced the idea of kingmaker which is a strategy that does not perform well for themselves but they determine the rankings among the top contenders. After the tournament, GRAASKAMP and DOWNING proved to be important kingmasters.

As explained above, DOWNING estimates the conditional probability that the other player will cooperate after DOWNING has cooperated and the probability that the co-player will cooperate after it has defected. Therefore, if the probabilities have the same values, DOWNING will defect since the other player seems to play randomly. If the other player seems to cooperate only after a cooperation, then Downing will understand that it is competing with a responsive player and it will cooperate as well. However, because initially, it sets the conditional probabilities to 0.5, it will defect for the first 2 move. This is particularly why DOWNING performed poorly. TFT and TIDEMAN AND CHIERUZZI earned the most points because DOWNING learned that defection will not increase his payoff, therefore there was mutual cooperation. GRAASKAMP performed well with TFT and NYDEGGER. Therefore, TFT ranked first as it performed well with both of the kingmasters and the others came in second and third.

Another rule that was observed during the tournament is forgiveness. Forgiveness is described as the tendency to cooperate in its moves after the other player has defeated. FRIEDMAN performed well with other nice strategies, but did not perform well with the kingmasters because it did not let them recover from the consequences of defecting and it increased the punishment. The main reason why the bottom six strategies underperformed is that most of the rules in the tournament are not forgiving. Because JOSS defects 10% of the time, when playing with TFT, after the move 25, they both constantly defect which substantially lowers their payoff.

Before the tournament, an example of a strategy was sent to prospective contestant to illustrate how to make the submission. The strategy is called TIT FOR TWO TATS and it will have performed better in the tournament that TIT FOR TAT, because it appears to be more forgiving as it will defect only if the other player defected on the previous two moves. As well, if DOWNING would have started with the assumption that the other contestants are responsive rather than unresponsive, it would have won the tournament.

3 Iterated games with Prisoner's Dilemma[2]

3.1 Prisoner's Dilemma Game

It was first formulated in the early fifties of last century as the story of two prisoners accused of joint crime. The state attorney proposed a deal to each of them: They can go free if they betray the accomplice. Then, the accomplice would get 10 years for the crime. If both of them confesses the crime, they would get 7 years each and if they stay silent, each of them gets 1 year.

In the generalized form of Prisoner's Dilemma Game, two players X and Y have to decide if they want to cooperate or defect. If both players cooperate (stay silent), they will earn reward R. On the other hand, if both of them defect (betray), they will each get punishment P. In this set up, reward R is larger than punishment P. Moreover, if one defects and the other cooperates, the defector gets larger payment T and the naive cooperator gets S. The resulting payoff matrix for player X is defined as $\begin{pmatrix} R & S \\ T & P \end{pmatrix}$. We have defined strategy $\mathbf{e_1}$ as cooperate and strategy $\mathbf{e_2}$ as defect.

This game will need to satisfy $T > R > P > S$ in order to guarantee that the Nash equilibrium is mutual defection.

Player X has to choose between 2 strategies, $\mathbf{e}_1$ cooperate and $\mathbf{e}_2$ defect. Suppose now that the player chooses the strategy $\mathbf{e}_i$ with probability x_i. The mixed strategy is given by a stochastic vector $\mathbf{x} = (x_1, x_2)$. The set of all such strategies is defined as Δ. Therefore, if player X uses mixed strategy $\mathbf{x}$ and player Y uses $\mathbf{y}$, the payoff for player X becomes

$$\mathbf{x} \cdot \mathrm{A}\mathbf{y} = \sum_i x_i (A\mathbf{y})_i = \sum_{i,j} a_{ij} x_i y_j$$

where A is defined as before.

Moreover, suppose that player X knows the strategy $\mathbf{y}$ of player Y, then he will need to use a strategy that is best response to $\mathbf{y}$. Then, we can define the best response map as the set of the best responses

$$\mathcal{B}R(\mathbf{y}) = \arg\max_{\mathbf{x}} \mathbf{x} \cdot \mathrm{A}\mathbf{y}$$

We can define the Nash equilibrium as a point $\hat{\mathbf{x}}$ iff $\hat{\mathbf{x}} \in \mathcal{B}R(\hat{\mathbf{x}})$. Explicitly, $\mathbf{x} \cdot \mathrm{A}\mathbf{x} \leq \hat{\mathbf{x}} \cdot \mathrm{A}\hat{\mathbf{x}} \quad \forall \mathbf{x} \in \Delta$, meaning that strategy $\hat{\mathbf{x}}$ cannot be improved by another strategy.

If player X has found a best response to player Y's strategy, then player Y will not change his strategy only if he has also found a best response. Two strategies $\mathbf{x}$ and $\mathbf{y}$ form a Nash Equilibrium pair iff $\mathbf{x} \in \mathcal{B}R(\mathbf{y})$ and $\mathbf{y} \in \mathcal{B}R(\mathbf{x})$.

For player X, the Nash equilibrium is to always defect. The stochastic vector associated to always defect is $\mathbf{e_2}$. Then,

$$\mathcal{B}R(\mathbf{e_2}) = \max_{\mathbf{x}} \mathbf{x} \cdot \begin{pmatrix} R & S \\ T & P \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \max_{\mathbf{x}} \mathbf{x} \cdot \begin{pmatrix} S \\ P \end{pmatrix} = \max_{\mathbf{x}} (x_1 S + x_2 P)$$

Given that $P > S$, $\mathbf{e_2} \in \mathcal{B}R(\mathbf{e_2})$. Similarly, for player Y, the Nash equilibrium will also be to always defect.

$$\mathcal{B}R(\mathbf{e_2}) = \max_{\mathbf{y}} \mathbf{y} \cdot \begin{pmatrix} R & T \\ S & P \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \max_{\mathbf{y}} \mathbf{y} \cdot \begin{pmatrix} T \\ P \end{pmatrix} = \max_{\mathbf{x}} (y_1 T + y_2 P)$$

Given that $T > P$, $\mathbf{e_2} \in \mathcal{B}R(\mathbf{e_2})$. Therefore, $\mathbf{x} = \mathbf{y} = (0, 1)$ represents a Nash equilibrium pair as long as $T > R > P > S$.

The **Donation Game** is a special case of Prisoner's Dilemma where two players have to decide simultaneously if they want to help their co-player. The two strategies will be defined again as cooperate and defect. The payoff matrix for player X with $b > c > 0$ is calculated as

$$\begin{pmatrix} b - c & -c \\ b & 0 \end{pmatrix}$$

3.2 Iterated Games

We can now consider several rounds of Prisoner's Dilemma game or Donation game. Suppose that the two players do not know how many rounds their game will last. We assume that after each round there is a probability z that the game is repeated at least one more round. The probability that the game is repeated at least n times is z^n and the probability that it is repeated exactly $n + 1$ rounds is $z^n(1 - z)$. Then, the number of rounds follow a geometric distribution with probability (1-z). The expected value of this distribution is, therefore, $\frac{1}{1-z}$. If we define A(n) the payoff in the n round, then the expected total payoff becomes

$$\sum_{n=0}^{\infty} z^n(1-z)[A(0)+A(1)+...+A(n)]$$

Using Abel's summation formula, we can reduce it to $A(0)+zA(1)+z^2A(2)+....$ As A(n) is bounded, then the expression will always converge. We define the average payoff per round as $(1-z)A(z)=(1-z)\sum_{n=0}^{\infty} z^n A(n)$.

For an iterated Prisoner's Dilemma, keeping $2R > T+S$, prevents alternating cooperating and defection giving a better outcome that mutual cooperation.

4 The success of TFT strategy in Iterated Prisoner's Dilemma Game or Donation Game[2]

We will investigate the success of TFT strategy compared to ALWAYS DEFECT and ALWAYS COOPERATE. Therefore, $\mathbf{e}_1$=AllC, $\mathbf{e}_2$=AllD and $\mathbf{e}_3$=TFT. We will consider a mixed large population. The frequency of the strategies are denoted by x, y and t which satisfies $x+y+t=1$. We will have P_x, P_y and P_t denote the total expected payoff to players using these strategies. Then, the average payoff is $\bar{\mathrm{P}} = x\mathrm{P}_x + y\mathrm{P}_y + t\mathrm{P}_t$. We will also introduce the concept of replicator equations. The replicator equations defines the growth rate of the frequencies of the strategies. It is defined as:

$$\dot{x} = x(\mathrm{P}_x - \bar{\mathrm{P}})$$
$$\dot{y} = y(\mathrm{P}_y - \bar{\mathrm{P}})$$
$$\dot{t} = t(\mathrm{P}_t - \bar{\mathrm{P}})$$

One important property of the replicator equations is that if we add an arbitrary function to all payoff terms, the replicator equation remains unchanged. Therefore, if we add a constant to a column in the payoff matrix, it will not affect the replicator equations.

If we consider Iterated Donation game, the average payoff per round matrix becomes

$$\begin{pmatrix} b-c & -c & b-c \\ b & 0 & b(1-z) \\ b-c & -c(1-z) & b-c \end{pmatrix}$$

AllC against AllC has $A(n) = b-c$ in every round, then the average payoff per round becomes

$$(1-z)A(z) = (1-z)(b-c)(1+z+z^2+...) = (1-z)(b-c)\frac{1}{1-z} = b-c$$

AllD against AllD has $A(n) = 0$ in every round, then the average payoff per round becomes $(1-z)A(z) = 0$.

TFT against AllC has $A(n) = b-c$ in every round, then the average payoff per round becomes $(1-z)A(z) = b-c$. The same result applies for AllC against TFT.

AllC against AllD has $A(n) = b$ in every round, then the average payoff per round becomes $(1-z)A(z) = b$.

TFT against AllD has $A(0) = b$ and then $A(n) = 0$ for $n \geq 1$. Then the average payoff per round becomes

$$(1-z)A(z) = (1-z)(b + 0 \times z + ...) = b(1-z)$$

AllD againt TFT has $A(0) = -c$ and then $A(n) = 0$ for $n \geq 1$. Then the average payoff per round becomes

$$(1-z)A(z) = (1-z)(-c + 0 \times z + ...) = -c(1-z)$$

TFT against TFT has $A(n) = b-c$ in every round, then the average payoff per round becomes $(1-z)A(z) = b-c$.

If we consider Iterated Prisoner's Dilemma game, the average payoff per round matrix becomes

$$\begin{pmatrix} R & S & R \\ T & P & (1-z)T + zP \\ R & (1-z)S + zP & R \end{pmatrix}$$

AllC against AllC has $A(n) = R$ in every round, then the average payoff per round becomes

$$(1-z)A(z) = (1-z)R(1+z+z^2+...) = (1-z)R\frac{1}{1-z} = R$$

AllD against AllD has $A(n) = P$ in every round, then the average payoff per round becomes $(1-z)A(z) = P$.

TFT against AllC has $A(n) = R$ in every round, then the average payoff per round becomes $(1-z)A(z) = R$. The same result applies for AllC against TFT.

AllC against AllD has $A(n) = T$ in every round, then the average payoff per round becomes $(1-z)A(z) = T$.

TFT against AllD has $A(0) = T$ and then $A(n) = P$ for $n \geq 1$. Then the average payoff per round becomes

$$(1-z)A(z) = (1-z)(T + P \times z + ...) = (1-z)T + (1-z)\frac{zP}{1-z} = (1-z)T + zP$$

AllD againt TFT has $A(0) = S$ and then $A(n) = P$ for $n \geq 1$. Then the average payoff per round becomes

$$(1-z)A(z) = (1-z)(S+ = P \times z + ...) = (1-z)S + (1-z)\frac{zP}{1-z} = (1-z)S + zP$$

TFT against TFT has $A(n) = R$ in every round, then the average payoff per round becomes

$(1-z)A(z) = R$.

For the next part, we will analyse the behaviour of the systems in the Donation game as it is a special case of the Prisoner's Dilemma. We will normalise the payoff matrix so that P_y, the payoff for defectors, is 0.

$$\begin{pmatrix} -c & -c & bz-c \\ 0 & 0 & 0 \\ -c & -c(1-z) & bz-c \end{pmatrix}$$

Therefore, the payoff for TFT and AllC is

$$\begin{aligned} \mathrm{P}_x &= (A\mathbf{x})_1 = -cx - cy + (bz-c)t = -c(x+y+z) + zbt = -c + zbt \\ \mathrm{P}_t &= -cx - c(1-z)y + (bz-c)t = \mathrm{P}_x + zcy \end{aligned}$$

We also have

$$\begin{aligned} \mathrm{P}_t - \bar{\mathrm{P}} &= \mathrm{P}_t - (x\mathrm{P}_x + t\mathrm{P}_t) \\ &= -c + zbt + zcy - \Big(x(-c+zbt) + t(-c+zbt) + tzcy\Big) \\ &= c(x+t) - c + zbt(1-x) + zcy(1-t) - t^2 zb \\ &= -c(1-x-t) + zbt(1-x-t) + zcy(1-t) \\ &= y\Big(zt(b-c) - c(1-z)\Big) \end{aligned}$$

If we have $t=0$ (no TFT players), then $\mathrm{P}_x = -c$ and $\dot{x} < 0$ and $\dot{y} > 0$. Therefore, AllD dominates AllC in this scenario.

If we have $x = 0$ (population only with TFT players and defectors), we have a fixed point $\mathbf{F} = (0, 1-\hat{t}, \hat{t})$ with $\hat{t} = \frac{c(1-z)}{z(b-c)}$. Every attractive fixed point is a Nash equilibrium. A fixed point $\mathbf{x}$ is attractive if for $x_i = 0$, $(A\mathbf{x})_i - \mathbf{x} \cdot A\mathbf{x} \leq 0$.

$$(A\mathbf{F})_1 - \mathbf{F} \cdot A\mathbf{F} \leq 0$$

$$-c + zb\hat{t} \leq \hat{t}\Big(-c + zb\hat{t} + wc(1-\hat{t})\Big)$$

$$(1-\hat{t})(c + zc\hat{t} - zb\hat{t}) \geq 0$$

Therefore, $\mathbf{F}$ is an unstable Nash equilibrium.

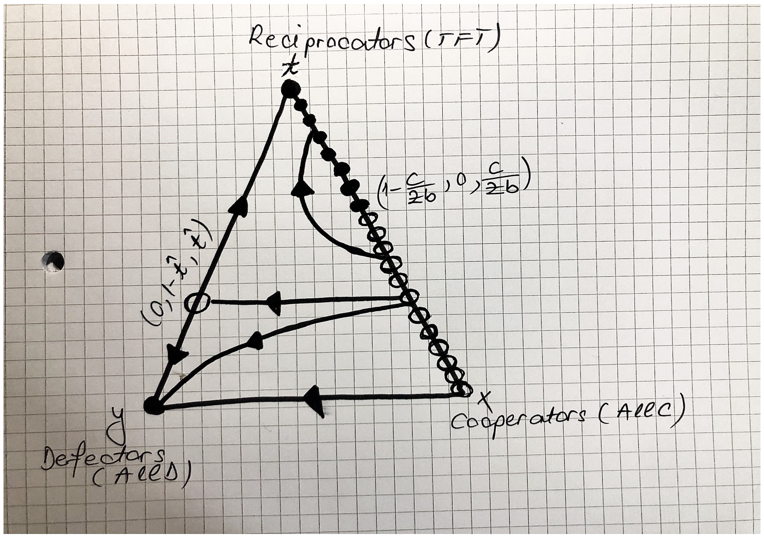
In a game of the form,

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

one strategy $\mathbf{e}_1$ is said to be risk dominant if it provides a higher payoff against a player who would attribute equal probability to both strategies. Hence, $\frac{1}{2}(\alpha+\beta) > \frac{1}{2}(\gamma+\delta)$. In our example, we have $x = 0$, $\alpha = 0$, $\beta = 0$, $\gamma = -c(1-z)$, $\delta = bz - c$. Therefore, TFT is risk dominant if $z > \frac{2c}{b+c}$.

For a sufficiently large population, strategy $\mathbf{e}_2$ is advantageous if $2\gamma + \delta > 2\alpha + \beta$. Then, TFT is advantageous if $z > \frac{3c}{b+2c}$.

If we have y=0 (population only with TFT players and cooperators), we have $\dot{x} = \dot{y} = \dot{t} = 0$. Thus, the segment $< e_1, e_3 >$ consists only of fixed points. To check which ones are Nash equilibrium, we need to check the conditions for $\mathrm{P}_y - \bar{\mathrm{P}} \leq 0$. Therefore, if $t \geq \frac{c}{zb}$, then the points are Nash equilibrium. The point with $y = 1$ is also a Nash equilibrium. There are no interior fixed points as $(A\mathbf{x})_1 \neq (A\mathbf{x})_2 \neq (A\mathbf{x})_3$ and we have $P_t > P_y$ for $y > 0$.



The segment that converges to the saddle point **F** consists of a single orbit that separates the simplex into two parts. As we can see, bellow the line t decreases and eventually, defectors win. On the other hand, above the line, the TFT players win. In the absence of defectors, any mixture of players corresponds to rest points. AllC and TFT players will always cooperate and none of the strategies is favoured. At this point, if we introduce a small number of defectors while $t > \frac{c}{zb}$ then the defectors will lose to TFT players. If the defectors appear while $t < \hat{t}$, then the defectors will win. If the defectors appears while $\hat{t} < t < \frac{c}{zb}$, then the defectors will eventually vanish. They will exploit AllC and dominate them, but they will meet TFT players and will lose, in the end.

5 Zero Determinant strategy introduced by Press and Dyson[3]

In the paper written by William F. Press and Freeman J. Dyson, it is presented a strategy that can help player X set player Y's score independently of his response or strategy and enforce a linear relation between their scores. This strategy is applied to an Iterated Prisoner's

Dilemma game. The strategies are derived assuming that the players have memory of only a single previous move. The four outcomes of the previous move $xy \in (cc, cd, dc, dd)$ are labeled 1,2,3,4 and player X's strategy is $\mathbf{p} = (p_1, p_2, p_3, p_4)$ which represents the probability of cooperating conditioned to each of the previous moves. Therefore $\mathbf{p}$ and $\mathbf{q}$ imply a Markov matrix that represents the transition probabilities

$$\mathbf{M}(\mathbf{p},\mathbf{q}) = \begin{pmatrix} p_1q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) \\ p_2q_3 & p_2(1-q_3) & (1-p_2)q_3 & (1-p_2)(1-q_3) \\ p_3q_2 & p_3(1-q_2) & (1-p_3)q_2 & (1-p_3)(1-q_2) \\ p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4) \end{pmatrix}$$

The dot product of the stationary vector $\mathbf{v}$ of $\mathbf{M}$ with the payoff matrices of each player will result in their specific scores. Any Markov matrix has a unit eigenvalue, then $\mathbf{M}' = \mathbf{M} - \mathbf{I}$ is singular. After some manipulations on $\mathrm{Adj}(\mathbf{M}')$, we get that $\mathbf{v} \cdot \mathbf{f} \equiv \mathbf{D}(\mathbf{p},\mathbf{q},\mathbf{f})$ where

$$\mathbf{D}(\mathbf{p},\mathbf{q},\mathbf{f}) = \det \begin{pmatrix} -1+p_1q_1 & -1+p_1 & -1+q_1 & f_1 \\ p_2q_3 & -1+p_2 & q_3 & f_2 \\ p_3q_2 & p_3 & -1+q_2 & f_3 \\ p_4q_4 & p_4 & q_4 & f_4 \end{pmatrix}$$

If we substitute $\mathbf{f}$ with the payoff matrices of both players we get their expected scores.

$$\begin{aligned} s_X &= \frac{\mathbf{v}\cdot\mathbf{S}_X}{\mathbf{v}\cdot\mathbf{1}} = \frac{\mathbf{D}(\mathbf{p},\mathbf{q},\mathbf{S}_X)}{\mathbf{D}(\mathbf{p},\mathbf{q},\mathbf{1})} \\ s_Y &= \frac{\mathbf{v}\cdot\mathbf{S}_X}{\mathbf{v}\cdot\mathbf{1}} = \frac{\mathbf{D}(\mathbf{p},\mathbf{q},\mathbf{S}_X)}{\mathbf{D}(\mathbf{p},\mathbf{q},\mathbf{1})} \end{aligned}$$

From $\mathbf{D}(\mathbf{p},\mathbf{q},\mathbf{f})$, we can observe that the second column is entirely determined by player X and the third column by player Y. We will name these columns $\bar{\mathbf{p}}$ and $\bar{\mathbf{q}}$.

The linear relation between the scores can be determined from $\alpha s_X+\beta s_Y+\gamma = \frac{\mathbf{D(p,q,\alpha S_X+\beta S_Y+\gamma 1)}}{\mathbf{D(p,q,1)}}$. If player X would choose his strategy so that it satisfies $\bar{\mathbf{p}} = \alpha\mathbf{S}_X + \beta\mathbf{S}_Y + \gamma\mathbf{1}$, then the determinant would be 0 and $\alpha s_X + \beta s_Y + \gamma = 0$. Therefore, William F. Press and Freeman J. Dyson proved that player X can enforce a linear relation between the scores. Now, we need to show how X can set player Y's score.

Suppose we set $\alpha = 0$, then $\bar{\mathbf{p}} = \beta\mathbf{S}_Y + \gamma\mathbf{1}$, which will result in

$$\begin{aligned} p_3 &= \beta S + \gamma \\ p_4 &= \beta P + \gamma \\ p_1 - 1 &= \beta R + \gamma \\ p_2 - 1 &= \beta T + \gamma \end{aligned}$$

Calculating p_2 and p_3 in terms of p_1 and p_4, we get

$$\begin{aligned} p_2 &= \frac{p_1(T-P) - (1+p_4)(T-R)}{R-P} \\ p_3 &= \frac{(1-p_1)(P-S) + p_4(R-S)}{R-P} \end{aligned}$$

With this substitution, we can finally set the score of player Y to

$$s_Y = \frac{(1-p_1)P + p_4 R}{(1-p_1) + p_4}$$

Player X can force a particular score to player Y independent of his strategy when p_1 is close to 1, p_4 is close to 0, p_2 is close to 0 and p_3 is close to 0.

On the other hand, player X cannot set his own score. When he sets $\beta = 0$, the only feasible strategy is given by $\mathbf{p} = (1, 1, 0, 0)$.

Moreover, player X can try and enforce for him a larger share of payoff than mutual defection P. He can do this by setting

$$\bar{\mathbf{p}} = \phi\Big((\mathbf{S}_X - P\mathbf{1}) - \chi(\mathbf{S}_Y - P\mathbf{1})\Big)$$

In this case, the strategy for player X has the following components

$$\begin{aligned}
p_1 &= 1 - \phi(\chi - 1)\frac{R-P}{P-S} \\
p_2 &= 1 - \phi(1 + \chi)\frac{T-P}{P-S} \\
p_3 &= \phi(\chi + \frac{T-P}{P-S}) \\
p_4 &= 0
\end{aligned}$$

Since the range of each p_i is [0,1], we can deduce that the range for ϕ is $\Big(0, \frac{P-S}{P-S+\chi(T-P)}\Big]$. In this case, player X's score depends on the strategy of the co-player and their scores are maximised when player Y fully cooperates. In the conventional frame of Iterated Prisoner's Dilemma that was presented in Axelrod Tournaments, using this strategy and $\chi > 1$ will keep player's X score always greater than 3, with a limit of $\frac{13}{3}$ and player Y's score around 1. If we set $\chi = 1$ and $\phi = \frac{1}{5}$, then the strategy for player X would be reduced to TFT.

5.1 Conclusion[4]

In the last part, we try to understand what would happen if player X would play with an evolutionary player or with a player with 'theory of mind'. An evolutionary player would adjust his strategy using an optimisation scheme designed to maximise his score. A player with 'theory of mind' would adjust his strategy with the intent of making the opponent adjust their strategy. For the evolutionary player, one strategy for him would be to make successive small adjustments in $\mathbf{q}$ so that he can climb the gradient in s_Y. If Y starts fully uncooperative, then X should try to yield a positive gradient in Y's cooperation at this $\mathbf{q}$. The gradient for player Y is calculated as

$$\frac{\partial s_Y}{\partial \mathbf{q}} = \left(0, 0, 0, \frac{(T-S)(S+T-2P)}{(P-S)+\chi(T-P)}\right)$$

This extortionate strategy presented by zero determinant strategy is based on the assumption that your opponent will change his strategy and eventually learning to cooperate. Therefore, player X would be able to exploit an evolutionary player but not a player with a theory of mind. The latter one would understand your strategy and he would start defecting all the time trying to convince you to change your strategies.

The zero-determinant strategy gained success due to the simpleness in each it manages to use the analysis of the replicator equations. After Axelrod tournament, it was proven that nice and forgiving strategies dominate and zero-determinant strategy created excitement because it is based on selfishness. However, zero-determinant and TFT thrive in different environments. TFT accumulates high scores from mutual cooperation against nice strategies. As seen before, TFT produces eventually equal high scores for both players. Essentially, it needs large number of high scoring matches in order to rank first in the tournament.

On the other hand, the extortionate strategy would win individual matches with not nice strategies and score low with nice strategies. In tournament, extortionate strategy came in second to last in number of total points, but second to top in number of matches won. An evolutionary strategy is one that also performs well against itself and zero-determinant strategy would defeat the other strategies, but eventually would become extinct as it is not evolutionary.

6 References

1. Robert Axelrod

Effective Choice in the Prisoner's Dilemma

2. Karl Sigmund

The calculus of selfishness

3. William H. Press and Freeman J. Dyson

Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent

4. Mike Shulman

Zero-Determinant strategies in Iterated Prisoner's Dilemma

Available from *https://golem.ph.utexas.edu/category/2012/07/zerodeterminant_strategies_in.html*