Fall 2023
14.12 Game Theory
Tomasz Sadzik

1 Infinitely Repeated Prisoners' Dilemma

Consider the Repeated Prisoners' Dilemma

$$
\begin{array}{cc|c|c|}
 & & \multicolumn{2}{c}{2} \\
 & & C & D \\
\hline
1 & C & 1,\ \ 1 & -1,\ 2 \\
\hline
 & D & 2,\ -1 & 0,\ 0 \\
\hline
\end{array}
$$

Suppose that the stage game is repeated infinitely often, and the players discount the later periods with a discount factor δ. That is, if a player get an income stream of $1, 1, 1, \cdots$, then her utility is

$$
\begin{aligned}
u_1 &= 1 + \delta + \delta^2 + \delta^3 + \cdots \\
&= \frac{1-\delta^\infty}{1-\delta} = \frac{1}{1-\delta}
\end{aligned}
$$

1.1 Grim-Trigger

We saw last lecture that the "grim-trigger" strategies

$$
s_1^* = s_2^* = \begin{cases} \text{On path: } C \\ \text{Off path: } D \end{cases}
$$

are a SPE as long as $\delta \geq 1/2$.

1.2 Limited Punishment

Grim-trigger achieves cooperation (in every single stage) by threatening to defect forever if one player should deviate from cooperation. This provides sufficient incentives for cooperation but seems excessively harsh and unforgiving. Imagine that players may make mistakes in playing the game or observing the past play. If such mistakes happen with positive probability in every stage, the interaction would brake down at some stage and the players would defect on each other forever after.

Let's now look at less extreme strategies that will eventually forgive the opponent for having played D

$$s^* = \begin{cases} \text{On path: } C \\ \text{Off path:} \\ \text{- punish by playing } D \text{ for 2 stages} \\ \text{- then revert to } C \end{cases}$$

The idea, of course, is to forgive each other after a period of punishment and to allow each other to return to the socially preferable Cooperation, while at the same time threatening to punish a defector for a time, e.g. 2 stages, that is long enough to make cooperation worthwhile in the first place.

Let's assume that both players adopt this strategy $s_1^* = s_2^* = s^*$ and call the 2 stages after an initial D the "punishment phase" (off the equilibrium path) and all other stages the "cooperation phase" (on the equilibrium path).

By the one-deviation principle we need to verify that player 1, say, does not want to deviate from the prescribed action in either the cooperation phase or the punishment phase.

1.2.1 Deviations in the punishment phase

In the 1st stage of the punishment phase a player knows that her opponent will play D for 2 stages and then revert to C. Her own strategy prescribes her to Defect in the coming 2 stages and then to revert to Cooperation.

Thus her payoff from Defecting now is given by

0 immediately

0 for the second punishment stage

1 for each stage after reversion to cooperation

for a total of $u(D) = \delta^2 (1 + \delta + \delta^2 + ...) = \frac{\delta^2}{1-\delta}$.

Her payoff from Cooperation now is given by

-1 immediately

0 for the second punishment stage

1 for each stage after reversion to cooperation

for a total of $u(C) = -1 + \delta^2 (1 + \delta + \delta^2 + ...) = \frac{\delta^2}{1-\delta} - 1$.

Thus she is better off by playing D.

The same argument goes through in the second punishment stage.

1.2.2 Deviations in the cooperation phase

In the cooperation phase a player has to trade off the incentive from defecting and getting a higher payoff immediately, and from cooperating and getting a higher payoff in the next 2 stages.

Her payoff from Defecting now is given by

2 immediately, i.e. in stage $t = 0$

0 for the coming 2 stages of punishment

1 for each stage after reversion to cooperation in stage $t = 3$.

Her payoff from Cooperation now is given by

1 immediately, i.e. in stage $t = 0$

1 in stage $t = 1$...

As both choices lead to the same continuation from stage 3 onwards, the incentives to Cooperate are given by $(1 + \delta + \delta^2) - 2 = \delta + \delta^2 - 1$ which is positive for $\delta \geq \delta^*$ where $\delta^* \approx 0.61$.

Remember that the harshest possible punishment in the "grim-trigger" strategies made the players indifferent between C and D for exactly $\delta = 0.5$. If the punishment is less harsh, as is the case here where players forgive each other after 2 stages, they need to be somewhat more patient to support Cooperation.

1.3 Tit for tat

There was a big interest in how people actually play in the repeated prisoners' dilemma and the political theorist Axelrod organized tournaments where scientists submitted strategies that subsequently played against each other. The winner of two of these tournaments was "tit-for-tat".

Consider the following strategy (called "tit for tat"):

$$s_1^* = s_2^* = \begin{cases} \text{stage 1: Play } C \\ \text{stage } t+1\text{: Play what your opponent played in stage } t \end{cases}$$

So, on the equilibrium path nobody ever defects and we get cooperation in every stage. However, tit-for-tat is more forgiving than grim-trigger in that any player could return to cooperation (by changing his strategy), if somebody "mistakenly" played D at some point.

Let us check whether it is an equilibrium for both players to play "tit for tat". The strategies specify different behavior after each action profile (C, C), (C, D), (D, C) and (D, D). By the one-step deviation principle we have to check four cases:

1.3.1 After (C, C)

Does player 1 want to deviate from playing C in the first stage?

C leads to (C, C) forever and a payoff of

$$1 + \delta + \delta^2 + \cdots$$

D leads to $(D, C), (C, D), (D, C), \cdots$ and a payoff of

$$2 - \delta + 2\delta^2 - \cdots$$

Deviation incentive

$$\begin{aligned} & 1 - 2\delta + \delta^2 - 2\delta^3 + \\ = \; & (1 + \delta^2 + \delta^4 + \cdots) - 2\delta(1 + \delta^2 + \delta^4 + \cdots) \\ = \; & (1 - 2\delta)(1 + \delta^2 + \delta^4 + \cdots) \end{aligned}$$

This is positive, i.e. 1 wants to deviate, iff $1 \geq 2\delta$, i.e. if $\delta \leq 1/2$

By symmetry, player 2 would not deviate either when play is still on path.

1.3.2 After (C, D)

Does player 1 want to Defect after player 2 Defected?

D leads to $(D, C), (C, D), (D, C), \cdots$ and a payoff of

$$2 - \delta + 2\delta^2 - \cdots$$

C leads to (C, C) forever and a payoff of

$$1 + \delta + \delta^2 + \cdots$$

Deviation incentive

$$(2\delta - 1)(1 + \delta^2 + \delta^4 + \cdots)$$

This is positive, i.e. 1 wants to deviate, iff $1 \leq 2\delta$, i.e. if $\delta \geq 1/2$! This constraint is unusual. Previous deviations we considered traded off a short term gain against a long term loss; such deviations are unprofitable if the long term is important, i.e. if the discount factor is large. Here, the off path deviation to forgive and C instead of following tit-for-tat and play D involves a short term loss and a long term gain. To prevent this deviation, the discount factor must thus be small.

1.3.3 After (D, C)

Player 1 is supposed to play C, leading to a sequence of action profiles $(C, D), (D, C), (C, D), \cdots$ for a payoff of

$$-1 + 2\delta - \delta^2 + 2\delta^3 - \cdots = \frac{-1 + 2\delta}{(1 - \delta)(1 + \delta)}.$$

If he rather deviates and plays D, the players will play (D, D) forever after for a total payoff of 0. Thus, the prescribed action C is better than D if $-1 + 2\delta \geq 0$, that is, if $\delta \geq 1/2$.

1.3.4 After (D, D)

Player 1 is supposed to play D, leading to (D, D) forever after for a total payoff of 0.

a sequence of action profiles $(C,D),(D,C),(C,D),\cdots$ for a payoff of

$$-1+2\delta-\delta^2+2\delta^3-\cdots=\frac{-1+2\delta}{(1-\delta)(1+\delta)}.$$

If he rather deviates and plays C, this will lead to a sequence of action profiles $(C,D),(D,C),(C,D),\cdots$ for a payoff of

$$-1+2\delta-\delta^2+2\delta^3-\cdots=\frac{-1+2\delta}{(1-\delta)(1+\delta)}.$$

Thus, the prescribed action D is better than C if $0 \geq -1+2\delta$, that is, if $\delta \leq 1/2$.

1.3.5 Summary

To summarize, tit-for-tat is not a SPE, unless $\delta \leq 1/2$ and $\delta \geq 1/2$, that is, if $\delta = 1/2$.

1.4 Asymmetric Grim Trigger

Both the plain "Grim-Trigger" strategy and the version with limited punishment are symmetric strategy profiles. This seems natural becuase the stage game, i.e. the PD, is symmetric as well. We will now see that there are also asymmetric equilibria in the infinitely repeated PD game. Consider the following strategy profile (s_1^*, s_2^*):

- On path
  - Player 1 always plays C
  - Player 2 plays C in stages $1,2,4,5,7,8,10...$ and D in stages $0,3,6,9,...$
- Off path, i.e. if player 1 ever defects or player 2 defects in any additional stage (other than $t=0,3,6,9,...$) both players defect forever after.

This is like the grim-trigger strategy with the exception that every third round player 2 cheats on player 1 and player 1 tolerates this as long as it only happens every third round. To check that these strategies constitute a SPE we will only check deviations by player 1 (because player 2 is getting a higher payoff from this strategy profile than he did from the symmetric grim-trigger, and thus will prefer to stick to s_2^* as long as $\delta \geq \frac{1}{2}$).

1.4.1 Off path

This is easy: Off the equilibrium path both players expect each other to defect forever after, so defecting in every stage is an optimal response

1.4.2 On path

Subgame $0, 3, 6, ...$ In any subgame on the equilibrium path starting at stage $t = 0, 3, 6, 9$: Player 1's discounted present payoff of these strategies is

$$\begin{aligned} u_1(s_1^*, s_2^*) &= -1 + \delta + \delta^2 - \delta^3 + \delta^4 + \delta^5 - \delta^6 + \cdots \\ &= \left(-1 + \delta + \delta^2\right)\left(1 + \delta^3 + \delta^6 + \cdots\right) \end{aligned}$$

His payoff from choosing D is 0.

Thus, he prefers to stick to (s_1^*, s_2^*) as long as

$$\delta + \delta^2 \geq 1$$

which is satisfied as long as $\delta \geq \delta^* \approx 0.62$.

Subgame $1, 4, 7, ...$ In any subgame on the equilibrium path starting at stage $t = 1, 4, 7, 10, ...$ player 1's payoff from cooperation is

$$\begin{aligned} u_1(s_1^*, s_2^*) &= \delta + \delta^2 - \delta^3 + \delta^4 + \delta^5 - \delta^6 + \cdots \\ &= \left(\delta + \delta^2 - \delta^3\right)\left(1 + \delta^3 + \delta^6 + \cdots\right) \\ &= \frac{\delta + \delta^2 - \delta^3}{1 - \delta^3} \end{aligned}$$

His payoff from D is $2\delta + 0 + 0... = 2\delta$, and he will prefer to cooperate iff

$$\begin{aligned} \frac{\delta + \delta^2 - \delta^3}{1 - \delta^3} &\geq 2\delta \\ 1 + \delta - \delta^2 &\geq 2(1 - \delta^3) \\ \delta - \delta^2 + 2\delta^3 &\geq 1. \end{aligned}$$

Numerically, we can find that this is the case when $\delta \geq \delta^* \approx 0.74$.

Subgame $2, 5, 8, ...$ In any subgame on the equilibrium path starting at stage $t = 2, 5, 8, ...$ player 1's payoff from cooperation is

$$\begin{aligned} u_1(s_1^*, s_2^*) &= \delta^2 - \delta^3 + \delta^4 + \delta^5 - \delta^6 + \cdots \\ &= (\delta^2 - \delta^3 + \delta^4)(1 + \delta^3 + \delta^6 + \cdots) \\ &= \delta^2 \frac{1 - \delta + \delta^2}{1 - \delta^3} \end{aligned}$$

His payoff from D is $2\delta^2 + 0 + 0... = 2\delta^2$, and he will prefer to cooperate iff

$$\begin{aligned} \frac{1 - \delta + \delta^2}{1 - \delta^3} &\geq 2 \\ -\delta + \delta^2 + 2\delta^3 &\geq 1. \end{aligned}$$

Numerically, we can find that this is the case when $\delta \geq \delta^* \approx 0.83$.

Assembling the cases, we find that the asymmetric grim-trigger strategy (s_1^*, s_2^*) is an equilibrium as long as $\delta \geq 0.83$. The binding incentive constraint is the one in rounds $2, 5, 8, ...$ Intuitively, for C to be an optimal action for player 1, the short-term gain of 1 ($2 - 1$ in rounds $0, 3, 6, ...$ and $0 - (-1)$ in all other rounds) must be overcome by the benefits of future cooperation. These benefits of future cooperation are smallest in round $2, 5, 8, ...$ just before player 1 anticipates player 2 to play D. If player 1 was to deviate from s_1^* at some point, the natural time to do this would be in rounds $2, 5, 8....$