Economics and Computation

ECON 425/563 and CPSC 455/555

Professor Dirk Bergemann and Professor Joan Feigenbaum

Reputation Systems

In case of any questions and/or remarks on these lecture notes, please contact Oliver Bunn at oliver.bunn(at)yale.edu.

1 Reputation Systems

1.1 Introduction

In the design of online marketplaces (of which Ebay is an example), the following three objectives should have superior importance:

- Security

  An online-marketplace should provide maximum security (to be understood in a predominantly financial manner) for the market-participants.

- Liquidity

  In order to increase the efficiency of market-participants' interaction in the marketplace, it is important that transactions can be performed as easily and convenient as possible.

- Trustworthiness

  In order to facilitate or even enable interaction between two participants in a marketplace it is essential to build up a system which measures and reliably displays the trustworthiness of market-participants. Furthermore, this system should be manipulation-proof towards any kind of malevolent action.

This section on **Reputation Systems** centers around the third aspect of trustworthiness.[1] These systems are designed to set incentives for market-participants to act in a mutually beneficent manner. A market-participant's reputation will be built up from objective as well as from subjective sources, which will be exemplified in the following, using the online marketplace Ebay:

- Objective information consists of a comprehensive report about the previous trading activity of a market-participant. This report encompasses the time of the beginning of the account-activity as well as information on the time and the specificities of traded items.

- Subjective information consists of the comments as well as a score on a three-point-scale (positive, neutral, negative) that all those market-participants have submitted that have previously engaged in trading with a particular market-participant.

Reputation systems in general bear the following three advantages:

[1] We will abstract away from potential manipulations of the reputation-systems, for example white-washing.

- Learning about quality

  Reputation systems make a lot of information available for a market-participant before she actually engages in trading activity with another party. Therefore, this market-participant can already draw learning-inference about the quality of a good offered by a potential trading-counterpart.

- Reduction of "moral hazard"

  In more general settings, and predominantly in the economic treatment of insurance-markets, "moral hazard" refers to the phenomenon that insured people change their risk-taking attitude upon having bought insurance. That is, people may take more risk (which can be interpreted as malevolent action from the point of view of the insurance-company) once they know that an insurance-company covers losses in the case of an accident or similar events. Of course, this is something that insurance-companies aim to avoid in order to keep their payments low. This is one of the reasons why insurance-companies have contracts with deductibles, as these entail some participation of the insurees in the case of an accident or similar events.

  In the context of interaction in online marketplaces, market-participants may consider malevolent actions. But it is the reputation-system which introduces intertemporal considerations. Reputation causes people to take into account potential obstructions to their future trading-activity that are caused by receiving a bad reputation from malevolent behavior today. That is, reputation systems alleviate the risk of observing "morally hazardous" behavior by market-participants and, therefore, induces honest and trustworthy behavior.

- Reduction of "adverse selection"

  In more general settings, and predominantly in the economic treatment of insurance-markets, "adverse selection" refers to a specific effect of heterogeneity within the group of potential insurees on the insurance-outcome. Suppose that potential insurees - facing varying degrees of a particular risk - all want to insure this risk fully with an insurance company. The degree of the risk determines the amount that a potential insuree is willing to pay. Importantly, the insurance-company does not observe the type, i.e. the degree of risk, of a potential insuree. In consequence, the insurance-company will only offer a contract for the average potential insuree. The problem that arises stems from the fact that there are now potential insurees who do not want to pay as much for the insurance as it is stated in the contract for the average potential insuree. Hence, the insurance-company will only attract the "more risky" people with the contract for the average potential insuree. But the insurance-company can anticipate this effect and, in consequence, offer a contract that is specifically tailored for the "more risky" people. Unfortunately, this excludes even more potential insurees from taking insurance, and so on. Finally, no insurance will be offered although all potential insurees would have wanted to purchase full insurance.

In the context of interaction in online marketplaces, reputation-systems increase the group of people that want to engage in trading. People that may shy away from trading without the reputation-system because they are too afraid of possible malevolent behavior of their trading-counterparts are caused to participate in the marketplace by the reputation-system which alleviates their fear. Furthermore, on the seller-side, the reputation-system allows to clearly distinguish benevolent sellers from malevolent ones, potentially increasing the group of sellers. That is, some individuals may now decide to participate on the seller-side of the online-mechanism because they can be clearly identified as selling "high-quality" goods. Hence, the reputation-system overcomes a possible segmentation of the group of potential market-participants, overcoming "adverse-selection-problems".

The final two properties are often referred to as the "reduction of **informational risks**".

Some aspects of this section are contained in chapter 27 of the textbook, [NRTV08].

1.2 Repeated Prisoner's Dilemma

In this section, we will investigate an accessible, but at the same time also very simplified, model of a reputation-system, the **repeated prisoner's dilemma**. So, we suppose the existence of two players that play the following normal-form-game:

Column-Player

		Cooperate	Defect
Row-Player	Cooperate	3,3	0,4
	Defect	4,0	1,1

In the following, we will make frequent use of the abbreviation C for "cooperate" and D for "defect".

When first discussing this normal-form-game, we have restricted attention to a single play of this game.[2] In this context, we have, among others, determined the following

[2] We have given different names to the two possible actions of a player. In the first discussion of the

properties:

- The play (D, D) constitutes an equilibrium in dominant strategies. It is optimal for each player to play D irrespective of the opponent's action. In fact, (D, D) is the only equilibrium of the game.

- The pair (C, C) is the efficient outcome of the game, i.e. the outcome which maximizes the sum of the two players' utilities. Hence, it appears to be desirable from a planner's point of view. But, unfortunately the structure of the game renders this outcome impossible in the context of a simultaneous-move game. This is due to the fact that each agent has an incentive to deviate from C by playing D.

Now, we will consider the Prisoner's dilemma in its repeated form, i.e. there are infinitely many periods $t = 0, 1, 2, ...$ and in each period the same two players play the above displayed normal-form game. The modeling-structure is completed as follows:

- There is a common **one-period discount factor** $\delta \in (0, 1)$. That is, the value of a payoff π_{t+1} in period $t + 1$ is given by $\delta\pi_{t+1}$, for all $t = 0, 1, 2, ...$

- A **payoff-stream** $(\pi_{i,0}, \pi_{i,1}, \pi_{i,2}, ...) \in \mathbb{R}^\infty$ for player i, $i = 1, 2$, where payoff $\pi_{i,t}$ is incurred in period t, $t = 0, 1, 2, ...$, is evaluated as

$$\Pi_i = (1 - \delta) \sum_{t=0}^{\infty} \delta^t \pi_i^t.$$

  This means that **agent** i**,** $i = 1, 2$**, evaluates the stream of payoffs as the discounted sum of payoffs**. The factor $1 - \delta$ in front of the sum is simply a normalization-factor. It stems from the desire that a constant payoff-stream $(c, c, c, ...)$, where $c \in \mathbb{R}$, has value c. Without the normalization, one obtains

$$\sum_{t=0}^{\infty} \delta^t c = \frac{c}{1 - \delta}.$$

  So, the multiplication by $1 - \delta$ yields the value c.

- When evaluating choices that a player has available at period t, for $t > 0$, the structure of Π_i allows to simply compare payoffs from the different choices at period t and treat period t as period 0. If a player prefers a certain payoff-stream from period t onwards to another payoff-stream in the sense of obtaining higher discounted value, it is the shape of Π_i which implies that this stream is also preferred when

Prisoner's dilemma we have used the labels "silent" and "confess" which were intended to be close to the "prison-story" that was told to motivate this game. Now, we will use the labels "cooperate" and "defect" in order to match the desired modeling of reputation-systems, to be outlined in more detail below.

compared at period 0. This is due to the fact that period-t-payoffs as well as period-t-discounted sums are transformed into their period-0-analogues by multiplication with δ^t.

- The course of the repeated interaction between the two players is captured by the notion of a **history**. A specific **history at time** t,[3] denoted h_t, contains all the actions that have been taken up to time t, i.e.[4]

$$h_t = ((a_{1,0}, a_{2,0}), (a_{1,1}, a_{2,1}), (a_{1,2}, a_{2,2}), ..., (a_{1,t-1}, a_{2,t-1}),$$

  where $a_{i,t}$ denotes an action taken by player i in period t, for $i = 1, 2$ as well as $t = 0, 1, 2, ...$ The **set of all possible histories at time** t is denoted by H_t, for $t = 0, 1, 2, ...$ Observe that $H_0 = \emptyset$, i.e. at time 0 there is no history of actions.

  Making use of $H_0, H_1, H_2, ...$, we are able to define the **set of histories** H as follows

$$H = \bigcup_{t=0}^{\infty} H_t.$$

  That is, H contains all possible histories at any point of time.

- A **strategy of player i** is given by a mapping[5]

$$s_i : H \rightarrow A_i,$$

  for $i = 1, 2$. That is, at any point of time a strategy prescribes an action conditional on any possible history $h_t \in H_t$. Intuitively, a strategy in a repeated game takes into account any possible contingency that may occur in the course of the repeated game and specifies an action to be taken in response.[6]

The goal of the analysis of the repeated Prisoner's Dilemma is the derivation of so-called **equilibrium strategies**. Instead of a formal definition, the following intuitive, verbal description will be sufficient for our purposes:

---

[3]We will assume that at time t all actions up to time t are perfectly observed. That is, each player knows exactly not only which actions he has taken up to time t, but also the actions that the other player has taken up to time t. This means that from the outcome of the game each player can exactly infer which choice of action the other player has taken. This assumption is far from being obvious. One might think about situations in which agents choose an action and subsequently the outcome of the game is determined stochastically where the probability measure depends on the chosen action.

[4]We will restrict attention to pure strategies in the analysis of the repeated Prisoner's Dilemma.

[5]As mentioned above, we will restrict attention to pure strategies in the analysis of the repeated Prisoner's dilemma.

[6]One might wonder why the strategy-mapping does not involve any time-index. This is due to the fact that the domain H contains all sets H_t, for $t = 0, 1, 2, ...$ Hence, s_i specifies an action at every possible history at time t.

An equilibrium strategy is a strategy where - given that the other player adheres to her strategy - each of the two players does not have a profitable deviation from the strategy at any point of the repeated game.

With respect to the objective to model a reputation-system, the repeated Prisoner's dilemma exhibits the following properties:

- The action "cooperate" refers to acting benevolently, i.e. contributing to a "good" (or "mutually beneficial") trade-outcome. The action "defect" refers to acting malevolently, i.e. causing a "bad" trade-outcome.
- The only reputation-element is the history of plays that is perfectly observed by both players. That is, there is no subjective component in the repeated Prisoner's Dilemma.
- It will always be the same two players that interact with each other. If one thinks about a marketplace like Ebay, where trade occurs between a multitude of pairs and a single participant will - in most cases - interact with many different players, this assumption appears to be highly debatable.
- In each period, both players - about which we can think as a seller and a buyer in an online marketplace - decide whether they want to act benevolently or selfishly. The efficient outcome in one period arises if both players act benevolently. But, any player has an incentive to act selfishly at the cost of the other player. In a repeated game the question arises whether it is possible to sustain cooperation by specifying an equilibrium-strategy that punishes a player sufficiently harsh if they defect.

1.3 Analysis of Strategies

An important role in the analysis of repeated games is played by the so-called **one-shot deviation-principle**, which will also only be displayed in intuitiv terms:

When analyzing potential deviations from a strategy, attention can be restricted to one-period-deviations. If no profitable one-period deviation exists for any of the two players at any stage of the game, then the prescribed strategy is an equilibrium.

1.3.1 Grim Trigger

So, consider the following **grim-trigger-strategy** for $i = 1, 2$:

$$s_i = \begin{cases} C & \text{at} \quad H_0 = \emptyset \\ C & \text{if} \quad h_t = (\underbrace{((CC),(CC),...,(CC))}_{t-1 \text{ times}}) \\ D & \text{at} \quad \text{any other history for } t > 0 \end{cases} .$$

The prescribed strategy has the following characteristics:

- There are only two states in the prescribed strategy, a "cooperation"-state, where both players are supposed to play C, and a "defection"-state, where both players are supposed to play D.
- In period 0, the players start out in the "cooperation-state".
- The repeated game will remain in the cooperation-state until one player decides to play action "defect". This will be the *trigger*-action. In this case, the above described strategies prescribe a switch to the "defection"-state.
- After having entered the "defection"-state, the game will remain in this state forever. This is the *grim*-part of the strategy.
- Although - if both players adhere to the grim-trigger-strategy - the repeated game will never enter the "defection"-state, we have to define the strategy of the player for the case of deviations. On the one hand, this is in accordance with the definition of a strategy as a mapping with domain H. But on the other hand, this procedure is also intuitively appealing because it will be the prospect of a "punishment" that will keep a player on track in playing C.

In order to verify that the prescribed strategy is an equilibrium-strategy, we will need to establish the following properties:

1. If the game is in the "cooperation"-state, none of the players has an incentive to play D instead of the prescribed action C in a particular period, while adhering to the grim-trigger strategy at any other period onwards.
2. If, for whatever reason, the game is in the "defect"-state, none of the players has an incentive to play C instead of the prescribed action D in a particular period, while adhering to the grim-trigger strategy at any other period onwards.

If both of these properties hold, then none of the players has an incentive to make a one-period-deviation from the prescribed strategy. Hence, the one-shot deviation-principle implies that the grim-trigger-strategy is an equilibrium.

Ad 1.
If the game is in the "cooperation-state" and both players adhere to the equilibrium-strategy, then each player will receive 3 as a payoff in every period onwards for infinitely many periods, giving her a discounted payoff of

$$(1-\delta)\frac{3}{1-\delta} = 3.$$

If, on the other hand, one of the players, say player 1, chooses action D, in contrast to what the grim-trigger-strategy prescribes, then in period t the pair of actions (D, C) will be realized (remember that we are always imposing the property that the other player remains true to the prescribed strategy). From period $t+1$ onwards, the game will move to the "defection-state" and both players will play D forever, as prescribed by the strategy. Hence, a player's discounted profit from period t onwards from deviating is given by

$$(1-\delta)4 + (1-\delta)\delta\frac{1}{1-\delta} = (1-\delta)4 + \delta = 4 - 3\delta.$$

So, it is not profitable for any player to choose D in the "cooperation"-state if and only if

$$\begin{aligned} & 3 \geqslant 4 - 3\delta \\ \Leftrightarrow \quad & \delta \geqslant \frac{1}{3}. \end{aligned}$$

Ad 2.
If the game is in the "defection-state" and both players adhere to the equilibrium-strategy, then each player will receive 0 as a payoff in every period onwards for infinitely many periods, giving her a discounted payoff of

$$(1-\delta)\frac{1}{1-\delta} = 1.$$

If, on the other hand, one of the players, say player 1, chooses action C, in contrast to what the grim-trigger-strategy prescribes, then in period t the pair of actions (C, D) will be realized (remember that we are always imposing the property that the other player remains true to the prescribed strategy). From period $t+1$ onwards, the game will stay in the "defection-state" and both players will play D forever, as prescribed by the strategy. Hence, a player's discounted profit from period t onwards from deviating is given by

$$(1-\delta)0 + (1-\delta)\delta\frac{1}{1-\delta} = 0 + \delta = \delta.$$

So, it is not profitable for any player to choose D in the "cooperation"-state if and only if

$$1 \geqslant \delta,$$

which is always satisfied by assumption.

In summary, we need to impose the assumption $\delta \geqslant \frac{1}{3}$ in order to sustain the grim-trigger-strategy as an equilibrium. A lower bound for δ is reasonable, because for very low values of δ the players predominantly care about the present period and not much about the future. This means that we are basically back in the one-period-setting where cooperation has been shown not to be establishable in equilibrium.

1.3.2 Forgiveness

Another strategy which will also sustain cooperation by both players in every period as an equilibrium-strategy (for suitably chosen values of the discount-rate δ to be determined below) is given by the following **forgiveness-strategy** for $i = 1, 2$.

- Play C in period 0, i.e. enter the "coordination"-state:
  - Play C as long as you observe the opponent playing C.
- If any player has played D in a period, enter the following "punishment"-state:
  - Play D in the following period, but return to playing C subsequently and re-enter the "coordination"-state.
  - If any of the players does not play D, play D for another period.

The prescribed strategy has the following characteristics:

- As it already becomes apparent from the description of the strategy, there are only two states in the prescribed strategy, a "cooperation"-state and a "defection"-state. Again, it can be observed that we only need to consider the state of the game in the previous period in order to determine players' actions according to the strategy.
- The repeated game will remain in the cooperation-state until one player decides to play action "defect". According to the forgiveness-strategy, this will trigger a one-period-play of (D, D), after which the repeated game returns to (C, C).

In order to verify that the prescribed strategy is an equilibrium-strategy, we will need to establish the following properties:

1. If the game is in the "cooperation"-state, none of the players has an incentive to play D instead of the prescribed action C in a particular period, while adhering to the grim-trigger strategy at any other period onwards.
2. If, for whatever reason, the game is in the "defect"-state, none of the players has an incentive to play C instead of the prescribed action D in the "punishment-period".

If both of these properties hold, then none of the players has an incentive to make a one-period-deviation from the prescribed strategy. Hence, the one-shot deviation-principle implies that the forgiveness-strategy is an equilibrium.

Ad 1.
If the game is in the "cooperation-state" and both players adhere to the equilibrium-strategy, then each player will receive 3 as a payoff in every period onwards for infinitely many periods, giving her a discounted payoff of

$$(1-\delta)\frac{3}{1-\delta} = 3.$$

If, on the other hand, one of the players, say player 1, chooses action D, in contrast to what the grim-trigger-strategy prescribes, then in period t the pair of actions (D, C) will be realized. Consequently, the strategy determines a one-period-play of (D, D) and afterwards the return to the play of (C, C) in every subsequent period. Hence, a player's discounted profit from period t onwards from deviating is given by

$$(1-\delta)4 + (1-\delta)\delta 1 + (1-\delta)\frac{3}{1-\delta} = (1-\delta)4 + (\delta - \delta^2) + 3\delta^2 = 2\delta^2 - 3\delta + 4.$$

So, it is not profitable for any player to choose D in the "cooperation"-state if and only if

$$3 \geqslant 2\delta^2 - 3\delta + 4$$
$$\Leftrightarrow \quad 2\delta^2 - 3\delta + 1 \leqslant 0.$$

This last inequality will be satisfied for $\delta \geqslant \frac{1}{2}$ (taking into account the assumption $0 < \delta < 1$).

Ad 2.
If the game is in the "defection-state" in period t and both players adhere to the equilibrium-strategy, then each player will receive 1 as a payoff in period t and 3 in every subsequent period from the play of (C, C), i.e. the discounted sum of payoffs is given by

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

If, on the other hand, one of the players, say player 1, chooses action C, in contrast to what the grim-trigger-strategy prescribes, then in period t the pair of actions (C, D) will be realized (remember that we are always imposing the property that the other player remains true to the prescribed strategy). In consequence, the strategy prescribes another round of (D, D) before the play returns to (C, C) for all future periods. Hence, a player's discounted profit from period t onwards from deviating is given by

$$(1-\delta)0 + (1-\delta)\delta \cdot 1 + (1-\delta)\delta^2\frac{3}{1-\delta} = \delta - \delta^2 + 3\delta^2 = \delta + 2\delta^2.$$

So, it is not profitable for any player to choose D in the "defection"-state if and only if

$$\begin{aligned} & 1 + 2\delta \geqslant \delta + 2\delta^2 \\ \Leftrightarrow \quad & 2\delta^2 - \delta - 1 \leqslant 0, \end{aligned}$$

which is always satisfied due to the assumption $0 < \delta < 1$.

In summary, we need to impose the assumption $\delta \geqslant \frac{1}{2}$ in order to sustain the forgiveness-strategy as an equilibrium.

References

[NRTV08] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, 2008.