Repeated Games with Perfect Monitoring

Tetsuya Hoshino

March 7, 2022

In this note, we study a **repeated game**, in which a group of players play the same normal-form game repeatedly over a long period of time. In particular, we maintain the assumption of **perfect monitoring**, under which all players perfectly observe past play.

1 Repeated Game with Perfect Monitoring

Example 1. Players play the following prisoners' dilemma repeated over infinite periods $t = 1, 2, \ldots$. Every period t, player i chooses action $a_i^t \in \{C, D\}$ and receives time-t payoff $u_i(a^t)$, for time-t action profile $a^t = (a_1^t, a_2^t)$. There is a discount factor $\delta \in (0, 1)$ common across the players, and an infinite history of play $h^\infty = (a^1, a^2, \ldots)$ gives player i her payoff $\sum_{t=1}^\infty \delta^{t-1} u_t(a^t)$ in terms of time-1 value.

	C	D
C	$1, 1$	$-l, 1+g$
D	$1+g, -l$	$0, 0$

Table 1: a prisoners' dilemma

Assume that player i knows the history of past play $h^t = (a^1, a^2, \ldots, a^{t-1})$ every period t.[1] This assumption is called perfect monitoring. □

1.1 Model

Definition 1. Fix a finite normal-form game $G = \langle I, (A_i, u_i)_i \rangle$ and a discount factor $\delta \in (0, 1)$. An **(infinitely) repeated game $G^\infty(\delta)$ with perfect monitoring** is an infinite-horizon extensive-form game such that:

1. Every period t, each player i simultaneously chooses action $a_i^t \in A_i$.
2. Every period t, each player i observes the history of past play $h^t = (a^1, a^2, \ldots, a^{t-1})$.[2]
   - $h^1 = \varnothing$ denotes that each player i has no information at period 1.
   - $h^\infty = (a^1, a^2, \ldots)$ denotes an infinite history of play.

[1] More precisely, every time-t history h^t is common knowledge among the players.
[2] This is the assumption of perfect recall.

3. Each player i's payoff at history h^∞ is

$$\sum_{t=1}^{\infty} \delta^{t-1} u_i(a^t). \tag{1}$$

The **stage game** of the repeated game $G^\infty(\delta)$ refers to the normal-form game G.

Remark 1. A discount factor δ in an infinitely repeated game has multiple interpretations. A first interpretation is that players literally discount their payoffs over time. This is a direct interpretation. A second interpretation is that a game may end, for an exogenous reason, with probability $1-\delta$ after every period. For example, if two firms engage in repeated price competition (á la Bertrand) they may form a price cartel to keep their prices high, rather than compete literally. Even if this cartel does not have an explicit deadline at which they break up the collusion, either firm may go bankrupt (for an exogenous reason) with probability $1-\delta$ every year. This story fits well the framework of infinitely repeated games. $\square$

Strategies Strategies in an infinite repeated game are defined in the same way as the strategies in a general infinite-horizon extensive-form game.

Definition 2. In a repeated game $G^\infty(\delta)$ with perfect monitoring, each player i's pure strategy is a function $s_i : \bigcup_{t=1}^{\infty} H^t \to A_i$, where H^t denotes the set of period-t histories h^t.[3] Let S_i denote the set of all pure strategies for player i.

Because our repeated game is of perfect recall, Kuhn's Theorem states that mixed strategies and behavior strategies are equivalent to each other.[4] Hence, it is without loss of generality to restrict attention to behavior strategies.

Definition 3. In a repeated game $G^\infty(\delta)$ with perfect monitoring, each player i's behavioral strategy is a function $\sigma_i : \bigcup_{t=1}^{\infty} H^t \to \Delta(A_i)$. Let Σ_i denote the set of all behavior strategies for player i.

Equilibrium Now that strategies have been defined for an infinite repeated game with prefect monitoring, Nash equilibrium and subgame perfect equilibrium for the repeated game can be defined. These equilibrium concepts in the repeated game are defined in the same way as the equilibrium concepts in a general infinite-horizon extensive-form game.

1.2 Preliminary Notions

We introduce preliminary notions that will be useful in later analysis.

[3]By definition, it must be that $H^t = A^t$, but we will use this notation H^t for the sake of exposition.

[4]Recall that the original version of Kuhn's Theorem requires that an extensive-form game not only be of perfect recall but also be finite. An infinitely repeated game is not finite. It turns out, however, that the same argument works for an infinitely repeated game with perfect monitoring.

Feasibility In a finite normal-form game G, a payoff vector v, which is a vector consisting of players' payoffs, is feasible if it is a convex combination of payoff vectors $u(a) \equiv (u_1(a), u_2(a), \ldots, u_{|I|}(a))$ across all action profiles $a \in A$. Note that this definition of feasibility disregards players' incentives.

Definition 4. In a finite normal-form game G, the set of **feasible payoff vectors** is

$$V = \operatorname{co}\Big(\Big\{u(a) : a \in A\Big\}\Big) \subset \mathbb{R}^{|I|},$$

where $\operatorname{co}(\cdot)$ is the convex hull of a given set. A payoff vector $v \in V$ is **feasible**.

Public Randomization It is sometimes natural to assume that players have access to a device for public randomization. We do not formalize this public randomization device, since it is essentially the same as the correlation device, which we have introduced in another note. That is, the public randomization allows players to correlate their choice of actions.

Even when players play a stage game only once, if they have access to a device for public randomization then they can achieve any feasible payoff vector (once we leave their incentives aside). As we will see later, the public randomization will significantly simplify the analysis.

Coffee Break ☕. Public randomization in reality? When the government commissions public works projects to private companies, it often decides which companies to commission by auction (procurement auction). Each company bids its estimate of the project cost, and the company with the lowest bid wins the project. Private companies may form a cartel to illegally raise cost estimates. The cartel secretly decides, in advance, which company wins the projects at a high cost estimate. For instance, sometimes company A wins, and sometimes company B wins. How does the cartel decide which company wins? It is difficult for cartel companies to discuss which company will win each time, and there is a risk of being caught. In a famous case in the 1950s in the U.S., a huge cartel, including GE, used the moon phase (!) to decide who would win a contract. □

Individual Rationality We begin with the worst-case scenario. How much payoff can player i guarantee for herself when players $-i$ take adversarial behavior against her? Her minimax payoff is the payoff that player i can guarantee by suitable choice of action a_i when player $-i$ attempt to minimize player i's payoff (regardless of their own incentives).

Definition 5. In a finite normal-form game G, player i's **minimax payoff** is

$$\underline{v}_i = \min_{\mu_{-i} \in \Delta(A_{-i})} \max_{a_i \in A_i} u_i(a_i, \mu_{-i}).$$

Remark 2. The outer minimization in the minimax payoff is across the set of correlated strategies of players $-i$ (where $\mu_{-i} \in \Delta(A_{-i}) = \Delta(\prod_{j \neq i} A_j)$ but not $\mu_{-i} \in \prod_{j \neq i} \Delta(A_j)$).

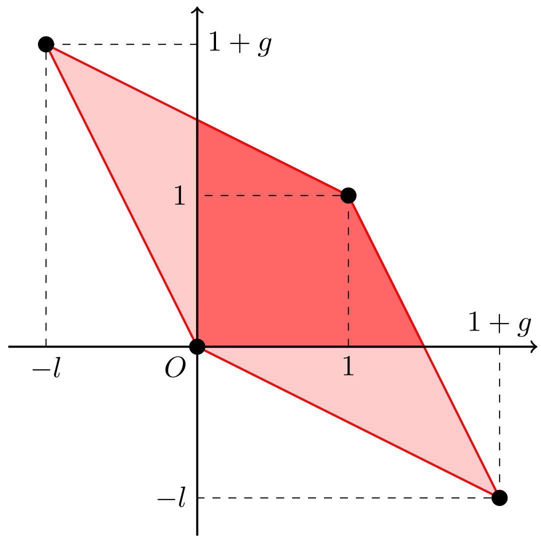

Figure 1: the set V^* of all feasible and individually rational payoff vectors of Example 1

Player i's correlated minimax payoff could be strictly less than her independent minimax payoff (when players $-i$ play independent mixed strategies). This distinction can be ignored if there are only two players. □

Since player i can guarantee herself the minimax payoff $\underline{v}_i$, it is intuitive that in every Nash equilibrium of a repeated game $G^\infty(\delta)$ with perfect monitoring, she will never have a payoff $v_i < \underline{v}_i$. Such a payoff is irrational. Any payoff that is not irrational (in this sense) is called individually rational.

Definition 6. In a finite normal-form game G, a payoff vector $v = (v_i)_i$ is **individually rational** if $v_i \geq \underline{v}_i$ for each $i \in I$ and is **strictly individually rational** if $v_i > \underline{v}_i$ for each $i \in I$. The set of **feasible and strictly individually rational payoff vectors** is

$$V^* = \Big\{ v = (v_i)_i \in V : \forall\, i \in I \;\, v_i > \underline{v}_i \Big\}.$$

Example 2. Figure 1 illustrates the set of feasible and/or individually rational payoff vectors of Example 1. The light red area, the convex hull of the four payoff vectors, is the set of feasible payoff vectors, while the dark red area is the set of feasible and strictly individually rational payoff vectors, where player i's minimax payoff is zero. □

2 Folk Theorem

Average (Stage-Game) Payoffs When discussing repeated games, we are often interested in the "average" (stage-game) payoff. This average payoff is defined such that if player i receives payoff c in every period then she has an "average" payoff c. As is immediate from the following definition, the "average" payoff is just normalization.

Definition 7. In a repeated game $G^\infty(\delta)$ with perfect monitoring, the **average (stage-game) payoff** to player i at an infinite history $h^\infty = (a^1, a^2, \ldots)$ is

$$(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} u_i(a^t). \tag{2}$$

2.1 Efficiency in Repeated Prisoners' Dilemma

Example 3. In the repeated prisoners' dilemma of Example 1, is the efficient average payoff vector $(1,1)$ achieved at some subgame perfect equilibrium when a discount factor δ is high enough?

The efficiency is attained if each player i uses the following **(grim) trigger strategy**:

$$\sigma_i(h^t) = \begin{cases} C & \text{if player } i \text{ has never observed action } D \text{ in history } h^t \\ D & \text{otherwise.} \end{cases}$$

In words, the trigger strategy is such that (i) in period $t = 1$, player i takes C, and (ii) in period $t \geq 2$, player i takes C if neither has played D in periods $1, 2, \ldots, t-1$ but D otherwise. Roughly speaking, a (grim) trigger strategy is a strategy such that a single defection triggers off defection forever.

How does this trigger strategy profile lead to the efficiency? If we assume that both players use the trigger strategies (leaving their incentives aside for a while) then they play (C, C) forever, thereby achieving the average payoff vector $(1, 1)$.

We show that no player has a profitable deviation from the trigger strategy profile. By the one-shot deviation principle, it suffices to show that no player has a profitable one-shot deviation from the trigger strategy profile. There are two cases to study:

1. Suppose that neither player has ever played D in periods $1, 2, \ldots, t-1$, so that both players are prescribed to play C at period t.
   - If player i plays C then her average continuation payoff is
     $$(1-\delta)(1 + \delta \cdot 1 + \delta^2 \cdot 1 + \cdots) = 1.$$
   - If player i plays D then her average continuation payoff is
     $$(1-\delta)(1 + g + \delta \cdot 0 + \delta^2 \cdot 0 + \cdots) = (1-\delta)(1+g).$$

   Player i has no profitable one-shot deviation for any $\delta \geq \frac{g}{1+g}$.

2. Suppose that at least one player has played D in periods $1, 2, \ldots, t-1$, so that both players are prescribed to play D at period t.

- If player i plays D then her average continuation payoff is

$$(1-\delta)(0+\delta \cdot 0+\delta^2 \cdot 0+\cdots)=0.$$

- If player i plays C then her average continuation payoff is

$$(1-\delta)(-l+\delta \cdot 0+\delta^2 \cdot 0+\cdots)=-(1-\delta)l.$$

Player i has no profitable one-shot deviation for any δ.

For any $\delta \geq \frac{g}{1+g}$, there is a subgame perfect equilibrium with the average payoff vector $(1,1)$. □

Remark 3. In the above proof, Case 1 considers a phase in which players "cooperate" each other to reward each player the target payoff level, while Case 2 considers a phase in which players "punish" each other to the minimax payoff level. We often call the former phase the **cooperation phase** and the latter phase the **punishment phase**. □

2.2 Folk Theorem: Informal Statement

What payoff vectors can be obtained in subgame perfect equilibria of a repeated game $G^\infty(\delta)$? The Folk Theorem collectively refers to the following "anything-goes" result:

> *If players are patient enough (with a high discount factor δ) then any feasible and strictly individually rational payoff vector is the average payoff vector of a subgame perfect equilibrium.*

Coffee Break ♨. The Folk Theorem is called after folklore, because it had been widely known that any feasible and strictly individually rational payoff vector can be obtained in a Nash equilibrium. However, a Nash equilibrium may depend on an empty threat. The "modern" Folk Theorem has refined the folklore by employing subgame perfect equilibrium. □

2.3 Folk Theorem: Two-Player Case

Theorem 1 (Fudenberg & Maskin 1986)**.** *In a repeated game $G^\infty(\delta)$ with perfect monitoring, if there are two players, $I=\{1,2\}$, then for each $v^* \in V^*$, there exists some $\bar{\delta} \in (0,1)$ such that for each $\delta > \bar{\delta}$, there exists a subgame perfect equilibrium with payoff vector v^*.*

Proof. Public randomization is not necessary, but since it simplifies the proof, we allow players to publicly randomize their choice of action.

Fix a target payoff vector $v^* \in V^*$. We will now construct a subgame perfect equilibrium with that payoff vector v^*. Assume, without loss of generality, that each player's minimax payoff is zero. Consider the following strategy profile with two phases:

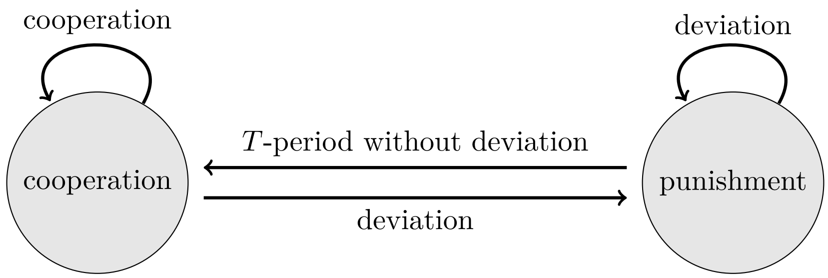

Figure 2: the transition between the two phases

Cooperation phase: Players 1 and 2 publicly randomize their action profiles such that they have the expected payoff vector v^*.

Punishment phase: Players 1 and 2 play the minimax strategies against the opponents, for T periods.

Here is a transition rule between the two phases (Figure 2):

- If at least one player deviates in a cooperation phase, go to a punishment phase.
- If at least one player deviates in a punishment phase, restart a punishment phase.
- If no player deviates for T periods in a punishment phase, go to a cooperation phase.

We introduce useful notations:

$$\bar{v}_i = \max_{a \in A} u_i(a), \qquad \bar{v} = \max\{\bar{v}_1, \bar{v}_2\}.$$

We write $\underline{u}_i$ for player i's stage-game payoff when player 1 plays the minimax action against players 2 and player 2 plays the minimax action against players 1.

It suffices to show that for appropriate choice of length T, this strategy profile is a subgame perfect equilibrium for any high enough discount factor δ. Fix any integer $T \geq 1$ such that $Tv_i^* \geq 2\bar{v}$ for each $i \in I$.

A subgame in a cooperation phase: If player i makes a one-shot deviation, her total payoff is at most

$$\bar{v} + \delta \underline{u}_i + \delta^2 \underline{u}_i + \cdots + \delta^T \underline{u}_i + \frac{\delta^{T+1}}{1-\delta} v_i^*. \tag{3}$$

If player i does not make a one-shot deviation, her total payoff is

$$v_i^* + \delta v_i^* + \delta^2 v_i^* + \cdots + \delta^T v_i^* + \frac{\delta^{T+1}}{1-\delta} v_i^*. \tag{4}$$

Then,

$$(4) - (3) = v_i^* - \bar{v} + \left(\delta + \delta^2 + \cdots + \delta^T\right)(v_i^* - \underline{u}_i)$$

$$\geq -\bar{v} + \left(\delta + \delta^2 + \cdots + \delta^T\right) v_i^*, \tag{5}$$

where the inequality follows from $v_i^* > 0$ and $\underline{u}_i \leq 0$. For any δ close to 1, it must be that $\delta + \delta^2 + \cdots + \delta^T > \frac{T}{2}$ and thus

$$(5) = -\bar{v} + \left(\delta + \delta^2 + \cdots + \delta^T\right) v_i^* \geq -\bar{v} + \frac{T}{2} v_i^* \geq 0,$$

where we recall that $T v_i^* \geq 2\bar{v}$. Hence, player i does not have any profitable one-shot deviation in a subgame in a cooperation phase.

A subgame in a punishment phase: Suppose that play is now at the K-th period in a punishment phase. If player i makes a one-shot deviation, her total payoff is at most

$$0 + \delta \underline{u}_i + \delta^2 \underline{u}_i + \cdots \delta^{T-K} \underline{u}_i + \delta^{T-K+1} \underline{u}_i + \cdots + \delta^T \underline{u}_i + \frac{\delta^{T+1}}{1-\delta} v_i^*. \tag{6}$$

If player i does not make a one-shot deviation, her total payoff is

$$\underline{u}_i + \delta \underline{u}_i + \delta^2 \underline{u}_i + \cdots \delta^{T-K} \underline{u}_i + \delta^{T-K+1} v_i^* + \cdots + \delta^T v_i^* + \frac{\delta^{T+1}}{1-\delta} v_i^*. \tag{7}$$

Then,

$$(7) - (6) = \underline{u}_i + \delta^{T-K+1}(v_i^* - \underline{u}_i) + \cdots + \delta^T (v_i^* - \underline{u}_i). \tag{8}$$

Since $v_i^* > 0$ and $\underline{u}_i \leq 0$, it follows that (8) ≥ 0 for any δ close to 1. Hence, player i does not have any profitable one-shot deviation in a subgame in a punishment phase.

In conclusion, for any high enough discount factor δ, there exists a subgame perfect equilibrium with the average payoff vector v^*. ■

Carrot and Stick The proof for the Folk Theorem (with two players) is involved, but the idea behind is simple. To sustain cooperation, players have to (i) reward themselves with good payoffs for complying with the cooperation and (ii) punish a deviating player by lowering her payoff enough (to discourage her from deviating). This "carrot-and-stick" is the key insight for sustainable cooperation and for the Folk Theorem as well.

2.4 Folk Theorem: Three-or-More-Player Case

The Folk Theorem holds true for a repeated game with perfect monitoring with three or more players. For this extension, we need an additional assumption.

Definition 8. In a repeated game $G^\infty(\delta)$ with perfect monitoring, the set V^* of feasible and strictly individually rational payoff vectors is **full-dimensional** if $\dim V^* = |I|$.[5]

Theorem 2 (Fudenberg & Maskin 1986)**.** *In a repeated game $G^\infty(\delta)$ with perfect monitoring, if the set V^* is full-dimensional then for each $v^* \in V^*$, there exists some $\bar{\delta} \in (0, 1)$ such that for each $\delta > \bar{\delta}$, there exists a subgame perfect equilibrium with payoff vector v^*.*

Proof. Omitted. ■

"Nested" Carrot and Stick If there are three or more players, the carrot-and-stick structure is "nested" in the sense that if there is a player who does not punish a deviating player, that player must also be punished. This idea of "punishing one who does not punish another" is crucial.

References

Fudenberg, D., & Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, *54*(3), 533–554.

[5]This full-dimensionality is equivalent to the set V^* having a non-empty interior.