MS&E 336 *Lecture 3: Reputation and payoff bounds*
Ramesh Johari April 9, 2007

In the last lecture, we saw an example of reputation analysis in repeated interaction (a sequential entry game) with one long-lived player and many short-lived players, where the short-lived players are unsure of the long-lived player's type. In this lecture, we generalize that analysis by giving an overview of the key results on reputation in repeated games with one long-lived player and many short-lived players. Our development draws on Chapter 15 of the book by Mailath and Samuelson [5].

1 An Example

1.1 Complete Information Static Game

Consider a two player simultaneous move game with the following payoff matrix:

		Player 2: a	Player 2: b
Player 1	A	(2,3)	(0,2)
	B	(3,0)	(1,1)

This game has a unique Nash equilibrium, (B, b).

1.2 Complete Information Repeated Game

Now suppose that a single Player 1 plays the game against an infinite sequence of short-lived players with the payoffs of Player 2; number the stages by $0, 1, 2, \ldots$. At each stage both players move simultaneously, and all past actions are perfectly observed. Denote the (possibly mixed) action played at time t by player i as a_i^t. Player 1 discounts payoffs, so he maximizes the following payoff in the repeated game is:

$$(1-\delta)\sum_{t\geq 0}\delta^t \Pi_1(a_1^t, a_2^t), \tag{1}$$

where the discount factor δ lies in $(0, 1)$, and $\Pi_1(\cdot)$ is player 1's stage game payoff function. Each Player 2 is short-lived, and so acts to maximize only his single period payoff at each stage.

Consider the following pair of strategies. After any history h^t where b has never been played, player 1 plays A; and otherwise player 1 plays B. Similarly, after any history h^t where B has never been played, player 2 plays a; otherwise player 2 plays b. It is straightforward to show this is a subgame perfect Nash equilibrium if $\delta \geq 1/2$. The strategy of player 2 is clearly a single stage best response to the strategy of player 1 after any history. Suppose player 1 deviates and plays B instead of A after any history where only (A, a) has been played; in order that this is not profitable, we must have:

$$(1-\delta)(3)+\delta(1) \leq 2.$$

The previous expression reduces to $\delta \geq 1/2$.

1.3 Incomplete Information Repeated Game

Now suppose that prior to beginning play in the repeated game, Nature chooses a type for player 1. With probability p, player 1 is a "commitment type" t^*, who plays A in every period regardless of the history. With probability $1-p$, player 1 is a "normal type" t that maximizes the discounted payoff in (1). We assume $p > 0$.

We will reason about the *Nash* equilibria of this game. Consider a pair of strategies analogous to those discussed in the previous section. The normal type plays A if b has never been played before, and otherwise plays B; similarly, each player 2 plays a if B has never been played before, and otherwise plays b.

Is this a Nash equilibrium? Let the *belief* $\mu(h^t)$ of player 2 be the probability that player 2 assigns to the commitment type of player 1, after the history h^t. After any history where B has never been played, we have $\mu(h^t) = p$; and if B has been played at least once, then $\mu(h^t) = 0$. In the latter case, given player 1's strategy, clearly b is the best response. If $\mu(h^t) = p$, then player 1 will play A regardless of his type, and so a is the best response for player 2.

It remains to be checked that the strategy above is optimal for the normal type. But this calculation is identical to the complete information case in the previous section. We conclude that the proposed strategy pair is a Nash equilibrium of the repeated game with incomplete information.

Note that in equilibrium, regardless of the type of player 1, (A, a) is observed in every period; and the belief of player 2 remains constant at p. Thus there exist equilibria of the repeated game where the normal type is *indistinguishable* from the commitment type. As we will see below, this feature rests on the fact that all past actions are perfectly observed by all players (also called *games with perfect monitoring*).

In the incomplete information setting, it is possible to show that the payoff of the normal type of player 1 in *any* Nash equilibrium is bounded below, and in particular bounded away from 1 if the discount factor is large. The intuition is that if the normal type of player 1 chooses to play A for the initial stages, even at the risk of a lower payoff, eventually player 2 must have high probability that player 1 is of the commitment type, and play a as a best response. But then player 1 will play A as a best response to player 2, sustaining high payoffs indefinitely.

In our example, we can lower bound *pure strategy Nash equilibrium* payoff of player 1 easily. Given any pure strategy Nash equilibrium, let t be the first stage on the equilibrium path where player 1 plays B. If $t = \infty$, then player 1 always plays A, the best response of player 2 is a every stage, and so (1) is 2.

Suppose $t < \infty$. Suppose player 1 deviates and plays A instead of B at time t, as well as at all subsequent stages. Player 2 expects only the commitment type to play A at time t, and thus has belief $\mu(h^t) = 1$ after player 1 deviates at time t; as a result, player 2 will play a in every future stage. The payoff under this strategy to player 1 must be at least as high as the payoff to player 1 under the Nash strategy. Thus the Nash equilibrium payoff of player 1 is at least:

$$\begin{aligned}(1-\delta)[(1+\delta+\cdots+\delta^{t-1})(2)+(\delta^t)(2)+(2)(\delta^{t+1}+\delta^{t+2}+\cdots)] &= 2(1-\delta^t)+2\delta^{t+1} \\ &= 2-2\delta^t(1-\delta) \\ &> 2\delta.\end{aligned}$$

Thus for any $\varepsilon > 0$, for all large enough δ *any* pure strategy Nash equilibrium payoff of player 1 is lower bounded by $2-\varepsilon$.

2 The Payoff Bound

In this section we generalize the payoff bound established for our example. Note that 2 is also the maximum payoff player 1 could guarantee himself (using a pure action) in a *Stackelberg* game, where he moves first and player 2 moves second: player 1 would play H, and subsequently player 2 would play h. For this reason 2 is also called the *Stackelberg payoff*, and H is called the *Stackelberg action*.

Consider a general repeated game with long-lived player 1, and short-lived player 2's; player i has pure action set A_i. For simplicity, assume all action sets and the type space are finite. Define:

$$v_1^* = \max_{a_1 \in A_1} \min_{a_2 \in BR_2(a_1)} \Pi_1(a_1, a_2).$$

Here BR_2 is the set of (pure and mixed) best responses of player 2 to a_1. The value v_1^* is called the *pure action Stackelberg payoff* of player 1, and the maximizer a_1^* is the *pure Stackelberg action* of player 1. Among the types of player 1, we assume there is a *Stackelberg commitment type* t^* that always plays a_1^*, and a *normal type* t that (rationally) maximizes the expected payoff (1).

Fudenberg and Levine [3] prove the following theorem.

Theorem 1 *Suppose that the prior p over player 1's type has $p(t^*), p(t) > 0$. Then for any $\varepsilon > 0$, there exists $d \in (0,1)$ such that if $d \leq \delta < 1$, the expected discounted payoff* (1) *to player 1 in* any *Nash equilibrium of the game is at least $v_1^* - \varepsilon$.*

Note that there may be many more possible types than just the Stackelberg and normal types; the theorem only depends on these types having nonzero prior probability.

3 Imperfect Monitoring

The game discussed in Section 1 had *perfect monitoring*: all players perfectly observed all past actions. Now suppose we modify the game so that player 2 only sees a *signal* of player 1's action. Formally, we assume that given (a_1^t, a_2^t), a signal s^t is realized according to a probability distribution $\pi(s|a_1, a_2)$, independent of all history. (Assume the signal space is finite.) All the player 2 actions are perfectly observable, so player 1 has the history $h_1^t = (a_1^0, a_2^0, s^0, \ldots, a_1^{t-1}, a_2^{t-1}, s^{t-1})$, while player 2 only has the history $h_2^t = (a_2^0, s^0, a_2^1, s^1, \ldots, a_2^{t-1}, s^{t-1})$.

An analogous payoff bound holds here, but by appropriately modifying the notions of Stackelberg payoff and action. Define the *confirmed best responses* $BR_2^c(a_1)$to a (possibly mixed) action a_1 as the set of all (pure and mixed) actions a_2 such that a_2 is a best response to *some* a_1' with $\pi(\cdot|a_1, a_2) = \pi(\cdot|a_1', a_2)$. Thus if $a_2 \in BR_2^c(a_1)$, then a_2 is a best response to *some* action of player 1 that induces the same distribution over signals as a_1. The point is that since a_2 is a best response to a_1', and the signal distributions are identical, player 2 can rationalize playing a_2 even if the true action of player 1 is a_1.

The *mixed action Stackelberg payoff* in this game is:

$$v_1^{**} = \max_{a_1 \in \Delta(A_1)} \min_{a_2 \in BR_2^c(a_1)} \Pi_1(a_1, a_2).$$

Here $\Delta(A_1)$ is the set of all mixed actions for player 1. Note that $v_1^{**} \geq v_1^*$ for a game with perfect monitoring. As above, let a_1^{**} be a (possibly mixed) action achieving v_1^{**}; we assume such an action exists, though the subsequent theorem can be proven with weaker assumptions. Denote by t^{**} be commitment type that always plays a_1^{**}, and let t continue to represent the normal type. Fudenberg and Levine prove the following result [4].

Theorem 2 *Suppose that the prior p over player 1's type has $p(t^{**}), p(t) > 0$. Assume that for fixed a_1, the signal distributions $\pi(\cdot|a_1, a_2)$ (over all mixed a_2) are linearly independent.*

Then for any $\varepsilon > 0$, there exists $d \in (0, 1)$ such that if $d \leq \delta < 1$, the expected discounted payoff (1) *to player 1 in* any *Nash equilibrium of the imperfect monitoring game is at least $v_1^{**} - \varepsilon$.*

The intuition for the result is quite similar to the perfect monitoring case. The linear independence condition is needed to ensure that with enough samples, any two action profiles of player 2 can be distinguished.

4 Comparing Perfect and Imperfect Monitoring

Note that in the perfect monitoring version of our example, we constructed a Nash equilibrium (in fact a sequential equilibrium) where the normal type plays A forever on the equilibrium path, exactly mimicking the commitment type; player 2 always plays a. This happens despite the fact that (A, a) is not a NE of the one-shot game.

Intuitively, it seems that the beliefs of player 2 should "converge" over time, since they are collecting an increasing amount of data about player 1's behavior. Once the beliefs have converged, then it would seem nothing player 1 does in one time step can alter the player 2 beliefs built on an essentially infinite amount of accumulated past data; and in turn, this suggest it might be plausible for player 1 to deviate from playing A occasionally, and play B instead.

This intuition is incorrect in the perfect monitoring case. With perfect monitoring, any player 2 knows that if the past history contains B, then player 1 *must* be the normal type. This condition holds regardless of how much past data has been accumulated. As a result, even if beliefs have converged, they can be substantially altered by a single stage play of B by player 1. Indeed, in the NE above in the perfect monitoring case, note that the beliefs of player 2 remain constant at p for all time on the equilibrium path—they never know with certainty whether they are playing a normal or commitment type, exactly because the normal and commitment type are indistinguishable on the equilibrium path.

The imperfect monitoring case is quite different. Here, player 2 imperfectly observes the past actions of player 1; instead, only the signals are recorded. If all signals have positive probability under all action pairs, then player 2's Bayesian updating will lead not only to converged beliefs, but beliefs that are not altered by any single action taken by player 1.

Consider the following imperfect monitoring analog of the 2×2 game constructed above. As one example, suppose that the signal space is $\{y_1, y_2\}$, with the following distribution:

$$\pi(y_1|(a_1, a_2)) = 1 - \pi(y_2|a_1, a_2) = \begin{cases} p, & \text{if} a_1 = A; \\ q, & \text{if} a_1 = B. \end{cases}$$

Here we assume $0 < q < p < 1$; then all signals have positive probability under any action profile. As the history grows, player 2's beliefs must converge, say to μ^∞. But then in the limit, the normal type of player 1 has no incentive to play A: by deviating and playing B, he *does not* substantively alter the belief of player 2, since monitoring is imperfect; and B is a strict dominant strategy in the stage game. But if this happens often enough, eventually player 2 must discover player 1 is of the normal type, contradicting the fact that beliefs had converged.

Note that in this example, the best response of player 2 to the Stackelberg action A is a; and the best response of player 1 to a is B. Our intuitive argument suggests that in this situation, the only possibility is that in the limit player 2 must discover player 1's true type. This is the content of the following theorem of Cripps et al. [1].

Theorem 3 *In a repeated game with imperfect monitoring: assume that all signals have positive probability under any action profile. Assume that for fixed a_1, the signal distributions $\pi(\cdot|a_1, a_2)$ (over all pure a_2) are linearly independent; and the same if we fix a_2, and vary over all pure a_1.*

Let $\hat{a}_1$ be the (possibly mixed) action of the commitment type of player 1; assume that player 2 has a unique (possibly mixed) best response $\hat{a}_2$ to $\hat{a}_1$ in the one shot game, and that $\hat{a}_1$ is not a best response to $\hat{a}_2$ in the one shot game. Then if player 1 has normal type, $\mu(h_2^t) \to 0$ as $t \to \infty$ (with probability 1).

The key step in the proof is a lemma showing that when player 1 has normal type, $\mu(h_2^t)\|\hat{a}_1 - E[s_1(h^{t+1})|h_t]\|$ converges to zero with probability 1. (The norm is the sup norm.) In other words, either the belief converges to zero, or in the limit player 1 is playing (in expectation) like the commitment type.

How is this lemma used? Suppose that when player 1 has normal type, there exists a positive probability set of histories where $\mu(h_2^t)$ remains bounded away from zero. Since the belief converges, in the limit we have $\mu(h_2^t) \to \mu^\infty > 0$. (We show rigorously below that beliefs converge.) From the previous lemma, on these histories the normal type player 1 must play like the commitment type eventually. On such histories, player 2 eventually comes to believe that with high probability, that she will be playing the best response $\hat{a}_2$ to $\hat{a}_1$ in the future. But in turn, it can be shown that with positive probability, player 1 will eventually prefer to play a best response to $\hat{a}_2$ along these histories, which is not $\hat{a}_1$; this contradicts the assumption that player 1 plays like the commitment type along these histories.

The moral of the theorem is that ultimately, in imperfect monitoring games, reputations are temporary—a normal type cannot masquerade forever as a commitment type. There are several approaches to preventing such a result. First, one could assume the type of player 1 varies stochastically over time; some references using this approach are presented at the end of the introduction of Cripps et al. [1].

Alternatively, one could design a *reputation system* that only allows player 2 access to a limited amount of history, e.g., perhaps only a fixed finite number of stages into the past. This ensures that the beliefs of player 2 never converge, and then reputation can be sustained indefinitely. This is a clever approach to motivating the use of reputation systems, and was pursued by Ekmekci [2].

5 Convergence of Beliefs

We conclude with a simple martingale argument that shows that beliefs must converge in the repeated games considered (with or without imperfect monitoring). Given an equilibrium $\boldsymbol{s} = (s_1(\cdot), s_2(\cdot))$, Nature's initial selection of the type of player 1 together with the strategy profile $\boldsymbol{s}$ induce a distribution over histories. Let χ be a random variable that is 1 if player 1 is the commitment type, and 0 if player 1 is the normal type. Then the belief of player 2 after history h^t is the conditional expectation of χ given the history; i.e.:

$$\mu(h^t) = \mathbb{E}[\chi | h^t].$$

But then we have:

$$\mu(h^t) = \mathbb{E}[\chi | h^t] = \mathbb{E}[\mathbb{E}[\chi | h^{t+1}] | h^t] = \mathbb{E}[\mu(h^{t+1}) | h^t].$$

(This follows by the computation rule for nested conditional expectations.) The computation implies that the beliefs are a *martingale* with respect to the history; further, since $0 \leq \mu(h^t) \leq 1$, this martingale is bounded, and so converges (with probability 1) by the martingale convergence theorem.

References

[1] M. W. Cripps, G. J. Mailath, and L. Samuelson. Imperfect monitoring and impermanent reputations. *Econometrica*, 72:407–432, 2004.

[2] M. Ekmekci. Sustainable reputations with rating systems. 2005. Working paper.

[3] D. Fudenberg and D. K. Levine. Reputation and equilibrium selection in games with a patient player. *Econometrica*, 57:759–778, 1989.

[4] D. Fudenberg and D. K. Levine. Maintaining a reputation when strategies are imperfectly observed. *Review of Economic Studies*, 59:561–579, 1992.

[5] G. J. Mailath and L. Samuelson. *Repeated games and reputations: long-run relationships*. Oxford University Press, Oxford, United Kingdom, 2006.