Game Theory and Algorithms*
Lecture 9: Extensive Games: Simultaneous Moves and (Infinite) Repetition

March 28, 2011

Summary: The goal of this lecture is to give a rigorous model in which "always cooperate" is a plausible outcome for the Prisoners' Dilemma: namely, the model of infinitely repeated strategic games with discounting. To get to this model, we introduce the ability to have simultaneous moves in extensive games, and we describe an alternate "single deviation" characterization of subgame perfect equilibria.

1 Simultaneous Moves in Extensive Games

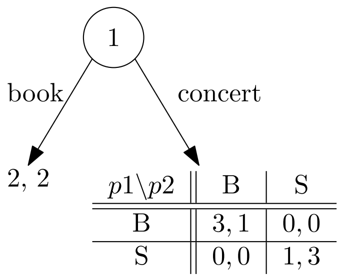
Previously, we defined extensive games in such a way that for every internal node h of the tree, exactly one player made a choice. (We also think of h as a *state* or a *history* or a *decision node*.) We now extend this as follows, for the model of *extensive games with simultaneous moves*:

- Each internal node h specifies a set $P(h)$ of one *or more* players who move at h.
- For each player $i \in P(h)$, they have a set $A_i(h)$ of valid actions at h.
- For each outcome in $\prod_{i \in P(h)} A_i(h)$, either it is a *terminal history* giving utilities to each player, or else the game continues with more decisions.

The players $i \in P(h)$ are thought of as making their choices simultaneously.

Remark. The special case where $|P(h)| = 1$ for all h is the model of extensive games without simultaneous moves described last class. Strategic games, on the other hand, correspond to any game where there is exactly one non-terminal history (namely, the root).

* Lecture Notes for a course given by David Pritchard at EPFL, Lausanne.



$p1\backslash p2$	B	S
B	$3,1$	$0,0$
S	$0,0$	$1,3$

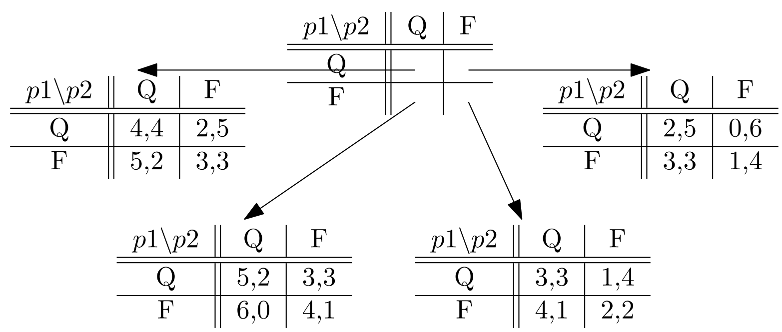
An example: Bach, Stravinsky or Book. First, player 1 decides either that both players will stay home and read a book, or they will go see a concert. If they decide to see a concert, then they proceed to play a game of Bach or Stravinsky. Both players prefer reading a book to listening their un-preferred composer.

1.1 Finitely Repeated Prisoners' Dilemma

Recall the example of a *Prisoners' Dilemma* from the first lecture:

p1 \p2	Q	F
Q	2,2	0,3
F	3,0	1,1

Our first example of a finitely repeated strategic game is the following: the players get to play two "rounds" of the Prisoners' Dilemma, and their utility will be the sum of their payoffs in the two rounds. For example if their actions in the first round are (Q, F) and their actions in the second round are (F, F), then for this history $h = ((Q, F); (F, F))$ we have $u_1(h) = 0 + 1 = 1$ and $u_2(h) = 3 + 1 = 4$.



$p1\backslash p2$	Q	F
Q	4,4	2,5
F	5,2	3,3

$p1\backslash p2$	Q	F
Q	2,5	0,6
F	3,3	1,4

$p1\backslash p2$	Q	F
Q	5,2	3,3
F	6,0	4,1

$p1\backslash p2$	Q	F
Q	3,3	1,4
F	4,1	2,2

Q: How many pure strategies does each player have in this game?
A: Each player has a total of 5 possible decisions in the game "tree." So, the total number of strategies for each player is $2^5 = 32$.

Let us try to use backwards induction on this game to predict the outcome. (It is a little informal but matches the formal descriptions to come later.)

- We need to start from the bottom of the game. For example, consider the subgame starting from history (Q, Q). If both players stayed quiet in the first round, what do we predict in the second round? Notice that the second round is still a Prisoners' Dilemma, the only difference is that both players each won 2 units of utility in the first round. So for player 1, in this subgame strategy F should strictly dominate strategy Q (indeed, $5 > 4$ and $3 > 2$). As we assume player 1 is greedy and rational, they will fink in the second round, and likewise for player 2.

- Similarly we deduce that no matter what happens in the first round, both players will fink in the second round.

- Now we can analyze what happens at the start of the game. Based on our analysis of the second round, we deduce that the effect of the first round on the payoffs is

p1 \p2	Q	F
Q	3,3	1,4
F	4,1	2,2

  or in other words that the players will get a constant utility of 1 unit each in the second round (from both finking). Again, this is the same as the original Prisoners' Dilemma up to an additive constant and the same calculations can use strict dominance to predict that both players will fink in round 1.

So backwards induction (together with strict dominance) predicts the strategy profile where both players will fink no matter what, leading to the outcome $((F, F); (F, F))$. This strategy profile will indeed turn out to be the only subgame perfect equilibrium of this game.

1.2 Subgame-Perfect Equilibria and Backwards Induction

The subgame-perfect equilibrium concept applies without modification to games with simultaneous moves:

Definition 1. A strategy profile s is a *subgame-perfect equilibrium* if for all players i, all strategies s'_i, and all histories h,

$$u_i(h(s'_i, s_{-i})) \leq u_i(h(s)).$$

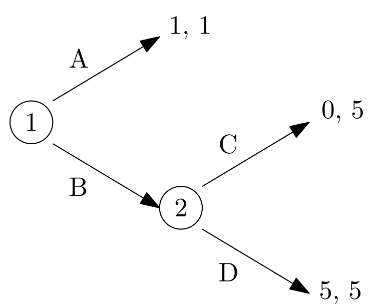
As before $h(s)$ means the leaf node reached starting from h when players play according to s. **Remark**: The same definition works too when s, s' are *mixed*. The definition also works for games of infinite duration.

Now it is a good time to go back to the concept of backwards induction, for which we gave no formal proof last time. The backwards induction algorithm can be described in the following way, which is actually forward-recursive instead of backwards-inductive:

// finds all subgame-perfect equilibria of a finite game G
Procedure BackwardsInduction(h) // G_h is a subgame
If h is a leaf node this game is trivial, return $\varnothing$
Call BackwardsInduction on each child $\{(h; a) \mid a \in \prod_{i \in P(h)} A_i(h)\}$ of h
for each combination $\widetilde{s}$ of SPEs returned by the recursive calls, **do**
  Consider the one-round strategic game with players $P(h)$ and action sets $A_i(h)$ with utilities $u(a) = u((h; a)(\widetilde{s}))$
  for every Nash equilibrium a^* of this game **do**
    output the combination of a^* and $\widetilde{s}$ as an SPE
Output BackwardsInduction($\varnothing$) //call it on root node/whole game G

This is the same as the algorithm we introduced last class, but generalized to handle ties and simultaneous moves. Be aware that we are intentionally using the same notation in a couple of different ways to make the algorithm description more succinct.

Remarks. The same algorithm works to find all *mixed* SPEs, but it is inefficient in general since one subroutine is *find all mixed NEs of a strategic game*. Even considering only pure equilibria, once ties are allowed there can be exponentially many SPEs, all of which are found by this algorithm, so it may not run in polynomial time.



Example 1. This is a game introduced in the last lecture. There is only one nontrivial subgame other than the whole game, $h = (B)$. (We name each internal node according to the sequence of actions which lead to it.) In this subgame, player 2 has two optimal actions. (The BI algorithm will construct a one-player game with two actions, which gives two Nash equilibria.) So the subgame has two SPEs, C and D. For the whole game each SPE of the subgame leads to a new case: if we take SPE C in the subgame then the root node gives optimal action A, and if we take SPE D in the subgame, the root node gives optimal action B. So there are two SPEs: AC and BD.

Example 2. Now consider finding all (pure) SPEs of Bach, Stravinsky or Book. The subgame has two NE, (B, B) and (S, S). Each SPE of the subgame gives one case when computing the SPEs of the whole game. In the (B, B) case the optimal starting action for player 1 is *concert* and otherwise it is *book*. So there are 2 SPE: $((\textit{concert}, B), B)$ and $((\textit{book}, S), S)$.

Example 3. The argument before for the Prisoners' Dilemma generalizes as follows: in *any strategic game with only one (pure/mixed) NE, for any finite k the k-repeated version has only one (pure/mixed) SPE*, namely the one in which all players always play according to the NE.

We eventually want to get to infinitely-repeated games to contrast with this last example. We digress a moment to another useful concept.

1.3 The One-Deviation Property

We haven't yet formally proved that backwards induction gives all SPEs. We will prove this now. As a side benefit it lets us mention a property which can be used to replace backwards induction once we talk about infinite games. (Clearly, if the game has infinite duration, backwards induction cannot be applied.)

Definition 2. A strategy profile s satisfies the *one-deviation property* if for all histories h, for all players $i \in P(h)$, and all strategies s_i' *which differ from* s_i *only at* h,

$$u_i(h(s_i', s_{-i})) \leq u_i(h(s)).$$

If we delete *which differ from* s_i *only at* h then we recover the definition of a SPE. Therefore, the following claim is trivial:

Claim 3. *Every SPE satisfies the one-deviation property.*

To prove that the backwards induction algorithm is correct, we will prove the converse of the above claim, and use the following:

Claim 4. *The backwards induction algorithm outputs s if and only if s satisfies the one-deviation property.*

Proof. Backwards induction precisely finds all strategy profiles satisfying the following: at each h each player $i \in P(h)$ picks an optimal action (locally at h), if we assume their future actions, and the actions of their opponents, are fixed. But this is directly the same as the definition of strategy profiles satisfying the one-deviation property. □

Claim 5. *In a finite game (extensive with simultaneous moves), if s satisfies the one-deviation property, then s is a SPE.*

In other words, we know that every s output by the backwards induction algorithm is locally optimal; we want to prove that actually there is no way to adjust (deviate) it in *two or more* moves so that any player profits. The proof is a little subtle and here we use an "extremal" argument.

Proof. Suppose that any strategy profile s is not an SPE. We will show that s does not satisfy the one-deviation property, which will complete the proof.

Since s is not an SPE, there is a history h, a player i, and a strategy s_i' so that $u_i(h(s_i', s_{-i})) > u_i(h(s))$. Define h^* as follows: h^* is the lowest node in the subgame rooted at h, of those where i plays, such that $u_i(h^*(s_i', s_{-i})) > u_i(h^*(s))$. In other words, the new value to player i of h^* is greater than its old value. Because all subgames strictly below h^* did not increase in value, it follows that the actions chosen by player i at h^* must be different in s_i and s_i', and that this single deviation increased his utility. Specifically, let s_i'' be the same as s_i, except that it takes the same deviation as s_i' at h^*; then

$$u_i(h^*(s_i'', s_{-i})) \geq u_i(h^*(s_i', s_{-i})) > u_i(h^*(s_i, s_{-i}))$$

where the first inequality holds since subgames strictly below h^* did not increase in value. □

As mentioned above, the three claims together give:

Theorem 6. *The output of the backwards induction algorithm is the set of all SPEs.*

2 Infinitely Repeated Games

We now move to infinitely repeated games, which lie in the generalized extensive-with-simultaneous-moves model wherein the game tree can be infinite. Our earlier definitions of subgame-perfect equilibria and the one-deviation property still make sense, but they are not always the same thing.

Exercise. Give an example of an infinite extensive game (with just one player) and one of its strategies s, such that s satisfies the one-deviation property, but s is not an SPE.

2.1 Discounting

As we alluded to earlier, the naive approach of "add your utilities from each round" does not give a good way to define payoffs in infinitely repeated games, since $1+1+1+\cdots = +\infty = 2+2+2+\cdots$ although clearly one of these sequences is "better" than the other. There are several approaches to fix this but we will focus on one easy, general, and well-motivated approach, which is discounting.

Definition 7. Let $(x_1, x_2, \dots)$ be an infinite sequence of payoffs (x_i is from round i). For $0 < \delta < 1$, the δ-*discounted sum* of the sequence is $x_1 + \delta x_2 + \delta^2 x_3 + \cdots = \sum_{i=1}^{\infty} x_i \delta^{i-1}$.

Provided that the x_i are bounded, this is a geometric series whose value is finite.

- We call this *discounting* since it agrees with the idea that (due to inflation), one dollar today is only worth $\delta < 1$ dollars tomorrow.
- Another equivalent view is that we are playing an infinitely repeated game which could terminate at any moment. Specifically, after each round, the game ends with probability $1-\delta$, otherwise it continues. Then given an infinite sequence of outcomes leading to utility $(x_1, x_2, \dots)$ for some player, its *expected* value if we terminate it in this random way equals its δ-discounted value.

So, we imagine in infinitely repeated games that each player wants to maximize their δ-discounted payoff.
Q: As δ increases, are people more patient, or less patient?
An important fact about this model of discounting is the following:

Theorem 8. *Under δ-discounted payoffs for any fixed $0 < \delta < 1$, for any infinitely repeated finite game, the strategy profile s is a subgame-perfect equilibrium if and only if it satisfies the one-deviation property.*

We give a proof in the appendix to the lecture notes. It uses the fact that the payoffs are "continuous:" on every infinite path, for every $\epsilon > 0$, there is a point on the path so that all payoffs below it differ by at most ϵ for every player.

We want to apply this to the infinitely-repeated Prisoners' Dilemma, in order to give a plausible reason for people to repeatedly cooperate. As we saw last lecture, a subgame-perfect equilibrium is specified by giving a complete *strategy* for each player that says what to do in every round and in every situation.

Exercise. Show that the following strategy profile s is *not* a subgame-perfect equilibrium (not even a Nash equilibrium): each player always cooperates.

To get what we want, the easiest fix is to look at a strategy profile which provides a credible threat.

Definition 9. The *Grim Trigger* (GT) strategy plays Q if your opponent has always played Q, and plays F otherwise. The *Modified Grim Trigger* (MGT) strategy plays Q if both players have always played Q, and plays F otherwise.

Note, the outcome of both (GT, GT) and (MGT, MGT) is that both players stay quiet forever. But (GT, GT) is not quite an SPE:

Exercise. Show (GT, GT) is not a subgame-perfect equilibrium.

Proposition 10. *(MGT, MGT) is a subgame-perfect equilibrium, if δ is large enough.*

Proof. To show this, we use Theorem 8: we need now only to show this strategy profile has the one-deviation property.

Consider a history h and a potential deviation by (WOLOG) player 1 at h. We need to show that if player 1 does the opposite of the action specified by MGT at h, their payoff does not increase.

Notice that the players have already accumulated some payoffs in the previous rounds. Since this value is unaffected by future choices, we may ignore it in our calculations. Also, we may multiply the utilities by $\delta^{-|h|}$ where $|h|$ is the length of h, in order to simplify the calculations.

- If everybody stayed quiet so far in h, and player 1 plays according to MGT, then everyone remains quiet. Player 1's payoff is the δ-discounted average

  $$2 + 2\delta + 2\delta^2 + \cdots .$$

  If player 1 deviates then the next round is (F, Q) and all future rounds are (F, F). So player 1's payoff would be

  $$3 + \delta + \delta^2 + \cdots .$$

  Thus, provided $2/(1-\delta) \geq 2 + 1/(1-\delta)$, i.e. $\delta \geq 1/2$, this is not a profitable deviation.

- The other case is that someone already defected in h. Then it is straightforward to see that everyone will continue to fink according to MGT in h, whereas deviating leads to a loss of utility (0 in one round for staying quiet, rather than 1).

This completes the proof. □

There are other strategies that do well (and in some ways better than MGT) in practice. There was a famous computerized tournament for the iterated Prisoner's dilemma held by Axelrod where *tit-for-tat* was the winner, beating out many more complex strategies:

Tit-for-tat: copy your opponent's last move.

This strategy does very well when $\delta = 1$ and there is a finite ending time, since it never underperforms its opponent by more than an additive constant. Another interesting variant, which we've copied for use in this class in some past years, is to extend the Prisoners' Dilemma to 3 players.

Exercise (Gibbons)**.** Consider the following infinitely repeated extensive game for some fixed $0 < \delta < 1$. Two players want to split \$1. Player 1 moves first, and they propose to a split of the dollar (two nonnegative numbers adding to 1, one for each player). Player 2 can either accept this offer (then each player gets utility equal to the specified amount) or reject it. If player 2 rejects, then the \$1 becomes $\$\delta$, and player 2 makes the next proposal (splitting $\$\delta$), which player 1 can accept or reject. Next player 1 again proposes a split of $\$\delta^2$, etc., with the players alternating proposals. Find a subgame-perfect equilibrium of this game. What are the players' utilities if they play according to this subgame-perfect equilibrium? (Hint: it can be useful to first think about a variant where after 2 rounds, the dollar is automatically split $(c, \delta^2 - c)$.)

Appendix: Proof of Theorem 8

As in finite games, SPEs trivially have the one-deviation property; we need to show the other direction.

Take some s which satisfies the one-deviation property and assume for the sake of contradiction that it is not an SPE. So $u_i(h(s_i', s_{-i})) > u_i(h(s))$ for some h, i, and s_i'. Let $x = u_i(h(s_i', s_{-i}))$ and $y = u_i(h(s))$, so $x > y$.

Since all players but i are fixed, it is helpful to ignore them: thus we get a game tree where the only player making a decision at each node is player i.

Let the *old value* of a node $\widetilde{h}$ be $v(\widetilde{h}) := u_i(\widetilde{h}(s))$ and its *new value* be $v'(\widetilde{h}) := u_i(\widetilde{h}(s_i'))$.

Let $(h = h_0, h_1, h_2, \dots,)$ be the sequence of nodes traversed in the tree starting from h when player i plays according to s_i'. We have $v'(h_0) = v'(h_1) = v'(h_2) = \cdots = x$ since this is the utility to player i using s_i' starting from h.

Since s satisfies the one-deviation property, even if they modify s_i by playing $s_i'(h_0)$ at h_0, their utility does not increase. So $v(h_1) \leq v(h_0)$. Likewise, even if they modify s_i by playing $s_i'(h_1)$ at h_1, their utility does not increase, so $v(h_2) \leq v(h_1)$. Continuing, we get $v(h_j) \leq v(h) = y$ for all j.

Notice that the new and old outcomes $h_j(s_i', s_{-i})$ and $h_j(s)$ are the same for at least the first j rounds. Since we are using δ-discounted payoffs, this implies that as $j \to \infty$, the difference in the utilities between the new and old outcomes tends to zero. But this contradicts that their difference is at least $v'(h_j) - v(h_j) \geq x - y$ which is a fixed positive number.