Lecture 4: Cooperation in Finite Games

Albert Banal-Estanol

May 2006

Today's Lecture

- Agents may be able to sustain cooperation if, by cooperating, they can mislead the others about their objectives
- Model: repeating a prisoner's dilemma potentially against an irrational player
- Preliminary analysis: one and two-period repetitions
- Decreasing level of cooperation in three-period repetitions
- Cooperation in *almost* every period in large number of repetitions

Cooperation in Finite Games

- Assume that (a) the following game is finitely repeated

	C	NC
C	$1,\ 1$	$-a, b$
NC	$b, -a$	$0, 0$

where $a > 0, b > 1$

- (b) Player 2 is either
  rational (maximise as usual) (Prob $1 - \theta$)
  "mechanical" (plays C as long as 1 has played C in the past) (Prob θ)

- If the game is played only once...
  both player 1 and the rational player 2 play NC

- If the game is played more than once...
  player 1 may play C because 2 may be mechanical
  player 2 (rational) may play also C pretending to be mechanical

Two repetitions

- In the second period:
  player 1 and rational 2 play NC (for any beliefs and any first period play)
  mechanical player 2 play C iff player 1 played 1 in the first period

- In the first period:
  rational 2 play NC (knows that player 1 will play NC in 2nd)
  mechanical player 2 plays C
  player 1 knows that rational 2 plays NC and mechanical 2 plays C in 1st
  if player 1 plays C in the first: payoffs $\theta(1+b)-(1-\theta)a=\theta(1+b+a)-a$
  if player 1 plays NC in the first: payoffs θb
  Hence, player 1 plays C in the first iff $\theta \geq \frac{a}{1+a}$

Three repetitions

- Notation: μ 1's prob. assessment that 2 is mechanical at beginning of 2^{nd} round
- Continuation after 1 played C in 1st period:
  game is as in the two-repetition above replacing θ with μ
- Continuation after 1 played NC:
  mechanical 2 plays NC in any circumstance
  outcome is (NC,NC) in the third and also in the second
- Can we achieve equilibrium full cooperation (all play C) in 1st period?
  Notice that no information would be revealed ($\theta = \mu$)

Cooperation in 1st Period?

- Suppose that $\theta \geq \frac{a}{1+a}$. Total payoffs after 1st period cooperation:
  player 1: $1 + \theta(1 + b + a) - a$ (she plays C in 2nd)
  rational 2: $1 + b$ (she plays NC in 2nd)

- For player 1...
  by deviating to NC her total payoffs are b
  C is an equilibrium provided that $\theta \geq \frac{a+b-1}{a+b+1}$ ($\in (0, 1)$)

- For rational player 2...
  by deviating to NC her total payoffs are b
  indeed, there is full revelation and $\mu = 0 < \frac{a}{1+a}$
  deviation is not profitable since $1 + b \geq b$

- Hence, if $\theta \geq \max\{\frac{a+b-1}{a+b+1}, \frac{a}{1+a}\}$ then there is an equilibrium with full cooperation in the first period

Remarks

- Rational 2 imitates the mechanical to maintain possibility in player 1's mind that she is mechanical (and not rational)

- Outcome: cooperation in the first, possible cooperation in the second and no cooperation in the third, as observed in experiments

- Again, this is only one of the equilibria (and it is an equilibrium when certain parametric assumptions hold)

Large Number of Repetitions (T)

- Theorem: in any SE, the number of stages where someone plays NC is bounded above by a constant that depends on θ, but is independent of T
- Step 1: If the type of the rational player 2 becomes known to 1 prior to some period t then both players select NC in round t and thereafter
- Proof by induction on the number of stages left in the game
  At $t = T$ this is clearly true
  Assume that it is true for some t. In $t - 1$, they know that they can't affect successive periods' outcomes
  Hence they look for a static best response in $t - 1$, which is NC
- Step 2: If both have played C until $t - 1$ and the rational 2 plays NC in t, then both play NC in all succeeding rounds
  Proof follows from the fact that this reveals 2's type and step 1

- Step 3: If 1 plays NC in $t-1$, both types of 2 play NC in all suceeding rounds
  Clear for the mechanical type
  For the rational one, if she plays C in $t' > t-1$ she will be revealed as rational and both rational players will play NC thereafter (from step 1)
  If she plays NC in $t' > t-1$ she will get a higher period payoff and cannot do worse thereafter (e.g. by playing always NC)

- Step 4: If 1 plays NC in $t-1$, then 1 play NC in all suceeding rounds
  From step 3, 1 knows that both types of 2 will play NC (for any of her actions)
  Best response is NC

- Define $M \equiv \frac{b+(1-\theta)a}{\theta}$
- Step 5: If no player has selected NC up to and including $t' < T - M$, then 1 should select C in $t' + 1$
- Proof by contradiction: suppose that player 1 selects NC in $t' + 1$ with prob>0
  She receives at most b from there and thereafter (from steps 3 and 4)
  If she deviates to: "C until 2 plays NC and plays NC thereafter"
  No information revealed until t', her payoff from there on is no worse than $\theta(T - t') - (1 - \theta)a > \theta M - (1 - \theta)a = b$
  Profitable deviation and contradiction

- Step 6: If no player has selected NC up to and including round t' where $t' < T - M - 1$ then rational 2 selects C in round $t' + 1$

- Proof:

  Suppose not, i.e. there was an equilibrium in which rational 2 selects NC with strictly positive probability in round $t' + 1$

  By steps 2 and 5, her continuation payoff is b (player 1 will play C in round $t' + 1$ and NC thereafter).

  If instead 2 deviates to play C in $t' + 1$ and NC in all subsequent rounds

  By step 5, she receives 1 in $t' + 1$ and b in $t' + 2$ (since $t' + 1 < T - M$, 1 plays C in $t' + 2$) and by step 2, 0 thereafter

  Thus, her total payoff is $1 + b > b$ and this is a profitable deviation

Remarks

- We have ruled out equilibria in which either player selects NC with positive probability in any round $t < T - M$. Given that a SE exists, we have that there exists a SE in which the players select C in these rounds
- Equilibrium structure difficult to characterize. Involves players cooperating early on, with cooperation breaking down probabilistically as the end of the game approaches, accurately describing experimental outcomes
- For any θ (even small) the fraction of cooperating rounds goes to unity as $T \to \infty$. Players almost always cooperate with long horizons, even if facing an irrational player is only a remote possibility
- Here, we have a particular form of irrationality. If one allows for all conceivable forms, one obtains folk-like theorems (anything can happen with sufficiently long finite horizons and arbitrarily small probabilities of irrationality provided that one does not restrict the form of irrationality)