Note MikØk2 - Repeated games

Tobias Markeprand

12. maj 2009

1 One-Stage-Deviation Principle

Recall that we define an extensive form game with perfect information as

$$E = (N, H, A, p, \alpha, \rho, \iota, (u_i)_i)$$

and for each node $x \in H \backslash T$ the subgame E^x is the extensive form game with H^x being the set of sequences of nodes which are sucessors of x, while the remaining functions are restrictions to H^x.

If E is an extensive form game, then we denote by $\ell(E)$ the lenght of E, i.e., the longest possible sequence of nodes in the game,

$$\ell(E) = \sup_N \{(x^n)_{n=1}^N \mid x^n \in H, x^n = x_0, x^n = p(x^{n+1}) \forall n < N\}.$$

A strategy of i in E is a function $s_i \colon H_i \to A$ such that $s_i(x) \in c(x)$ for any $x \in H_i$. A strategy profile is a Nash equilibrium of E if it is the Nash equilibrium of it's normal form representation. A strategy profile is a Subgame Perfect Nash equilibrium (SPE) if for every subgame E^x the restriction of each players' strategy $s_i^\star|_x$ is a Nash equilibrium of E^x.

Proposition 1 (One-Stage-Deviation Principle) *Let E have a finite time horizon. Then the following statements are equivalent*

1 *A strategy profil $s^\star$ is a SPE,*

2 *there exists no player i, a node x' and a strategy s_i in which $s_i^\star(x') \neq s_i(x')$ and $s_i^\star(x) = s_i(x)$ for $x \neq x'$ such that $u_i(s_i, s_{-i}^\star) > u_i(s_i^\star, s_{-i}^\star)$.*

We abbreviate this to be the **OSD** (one-stage-deviation principle).
Proof:

$\Rightarrow$: is obvious from the definition of SPE.

$\Leftarrow$: Assume that $s^\star$ satisfy condition 2, but that it is not a SPE. Let E^x be the subgame in which $s^\star|_x$ is not a NE, and denote by s_i a strategy of i which is a strictly better responds in this subgame. From condition 2 we know that $\ell(E^x) > 1$ and $\ell(E^x) < \infty$. Furthermore, we know that $s_i(x) \neq s_i^\star(x)$ for at least 2 nodes $x', x'' \in H^x$. Let $x^1 \in H^x$ such that $x^1 = \min_{x'} \ell(E^{x'})$ and $s_i(x^1) \neq s_i^\star(x')$.

Then consider the strategy $\tilde{s}_i$ given by $\tilde{s}_i|_{x^1} = s_i^\star|_{x^1}$, and $\tilde{s}_i(p(x^1)) = s_i^\star(p(x^1))$ and $\tilde{s}_i(x') = s_i(x')$ else. We know that by the definition of x^1, $\tilde{s}_i$ is at least as good a responds as $s_i^\star$ and by condition 2 we know that $\tilde{s}_i$ is at least as good as $s_i^\star$ from node $p(x^1)$, thus we can conclude that $u_i(\tilde{s}_i, s_{-i}^\star) > u_i(s_i^\star, s_{-i}^\star)$. We can continue this procedure and eliminate one node in each round, and since the game is finite we will eventually have found a strategy that only deviates from $s_i^\star$ in one node, but is strictly better. Contradiction!

□

The content of the one-stage-deviation principle is that in validating if the strategy profile is indeed a SPE, we only need to consider alternative strategies that deviate in a single stage from the candidate SPE. The full value of the OSD, however, does not present it self until applied within the infinite horizon games.

In the finite horizon case, the one-stage-deviation principle coincide with the Backwards induction principle, which, however, only works whenever there is a final node.

2 Repeated games

Repeated games are ideal to analyse how players can coordinate their behaviour by punishing non-cooperative behaviour.

Definition 1 (Repeated game) *A repeated game is given by* (G, δ, T) *where* $G = (N, (S_i, u_i)_i)$ *is the* **stage game***, a normal form game,* $\delta \in]0, 1[$ *is the* **discountingfactor***, and* T *is the time horizon. An* **outcome** *of* (G, δ, T) *is a sequence* $\xi = (\xi_i)$ *where* $\xi_i = (s_i^t)_{t=0}^T$ *and* $s_i^t \in S_i$ *is an action played in stage* t*, and let* H^{t-1} *denote the set of all outcomes with length* $t-1$ *and payoff is given by*

$$U_i(\xi) = \sum_{t=0}^{T} \delta^t u_i(\xi^t).$$

A **strategy** *in* (G, δ, T) *of* i *is a function* $s_i^t \colon H^{t-1} \to S_i$ *and* $s_i = (s_i^t)_t$*.*

We note that an implicit assumption of a strategy in the definition of a repeated game is that each player can observe each other players' action and can recall this action into the indefinite future.

In repeated games and infinite horizon with discounted utility the one-stage-deviation principle will be valid:

Proposition 2 *Let* T *be arbitrary and* $\delta < 1$*, then* $s^\star$ *is a SPE iff. there exists no player* i*, a node* x' *and a strategy* s_i *such that* $s_i^\star(x') \neq s_i(x')$ *and* $s_i^\star(x) = s_i(x)$ *for* $x \neq x'$ *and* $U_i(s_i, s_{-i}^\star) > U_i(s_i^\star, s_{-i}^\star)$*.*

The intuition is as follows: that with discounted payoff, payoffs into the distant future has a very small effect on the total payoff, and thus we can restrict

attention to deviationens within a finite horizon; but with finite horizon we can apply proposition 1.
Proof:

Consider a strategy that differ in the infinite future, then consider a sequence of finite deviation strategies approximating this infinite deviation, and the payoff of this sequence must all be lower than the equilibrium strategy by the finite horizon OSD, and since the payoff of the approximating sequence converges to the payoff of the infinite deviation, by continuity of the payoff we must have that the equilibrium payoff must be greater than the infinite deviation.

□

Let us illustrate the power of the one-stage-deviation principle. Consider the bimatrix

1/2	F	M
F	(-3,-3)	(0,-7)
M	(-7,0)	(-1,-1)

Figur 1: Prisoners dilemma

where (F, F) is the Nash equilibrium and even in strictly dominated strategies. It is thus a very strong prediction, but both will benefit from coordinating on playing "Mum". But why do we rarely observe criminals rat on each other? One reason could be that there is a day after tomorrow in which the betrayed could make the day a living hell for the betrayer. Let us se how.

Proposition 3 *There exists a SPE in the game 1 as a repeated game such that (M, M) is the outcome in each stage for some $\delta \geq \frac{1}{3}$.*

Proof:

Consider the following strategy of player 1:

$$s_1(\xi^t) = \begin{cases} F & \exists \xi^s, s < t : \xi_2^s = F \\ M & \text{else} \end{cases}$$

Thus: play the cooperative strategy M until anything else happens, and if somebody deviates to F then punish by playing a, the non-cooperative strategy. Consider player 2's strategy

$$s_2(\xi^t) = \begin{cases} F & \exists \xi^s, s < t : \xi_1^s = F \\ M & \text{ellers} \end{cases}$$

We show that this strategy profile is a SPE. By this end we apply the OSD. There are two equilibrium paths, i.e., paths when players stick according to their strategies: one in which (F, F) is the outcome in all stages and one in which (M, M) is the outcome in all stages. It is obvious that $\xi^t = (F, F)$ is a Nash equilibrium, and no player can gain by deviating. Consider the path

$\xi^t = (M, M)$, and assume that player 1 deviates and plays a in stage t. In the remaining game player 2 according to the strategy will play F, thus the outcome in all of the remainning game is (F, F).

The payoff of a path of outcomes (b, B) to player 1 is

$$U_0 = \sum_t \delta^t(-1) = -\frac{1}{1-\delta}$$

while the payoff of deviating on time t is

$$U_1 = \sum_{s=0}^{t-1} \delta^s(-1) + \delta^t(0) + \sum_{s=t+1}^{\infty} \delta^s(-3).$$

The difference between the payoffs are then

$$U_0 - U_1 = \delta^t(-1) + \sum_{s=t+1}^{\infty} \delta^s(2) = \delta^t(-1 + \frac{2\delta}{1-\delta})$$

such that $U_0 \geq U_1$ whenever $\delta \geq \frac{1}{3}$.

□

The requirement that δ should be sufficiently high, is due to the fact that the punishment for the deviation can only be effected by the other player in the future. But this punishment will be discounted and thus ceritus paribus smaller. The more patient, ie. the greater δ, the more likely it is that cooperation can occur.

2.1 Folk theorem

The result that players can maintain a cooperative behaviour in repeated games when the discounting is sufficiently high is also called the Folk Theorem. There is however other payoffs that can be obtained in a SPE. The Folk theorems states the structure of the feasible payoffs of repeated games. It turns out that the set is actually very large.

Definition 2 *We say that an outcome* $(s^t)_{t\in T}$ *is* **feasible** *if* $s^t \in S$ *for all* t

Definition 3 *If* $T < \infty$ *and given a feasible outcome* $(s^t)_{t\in T}$*, define the* **average payoff** *of player* i *as*

$$\pi_i = \frac{1}{T}\sum_{t\in T} u_i(s^t).$$

If $\sup_{s\in S} u_i(s) < \infty$ *and* $T = \infty$ *then we define the average payoff of player* i *as*

$$\pi_i = \lim_{T\to\infty} \frac{1}{T}\sum_{t\in T} u_i(s^t).$$

Definition 4 (Minimax payoff) *We define the* **minimax payoff** *of player i by*

$$\underline{\pi}_i = \inf_{s_{-i} \in S_{-i}} \sup_{s_i \in S_i} u_i(s_i, s_{-i})$$

.

In the game 1, we have that $\underline{\pi}_i = -3$ for both players $i = 1, 2$. The point is that the player can guarantee himself an payoff of at least -3 regardless of what the other play do. Thus, the minimax payoff is a kind of insurance: no matter what happens, the player can always achieve this payoff.

In the game

1/2	A	B	C
a	(4,3)	(5,1)	(6,2)
b	(2,1)	(8,4)	(3,6)
c	(3,0)	(9,6)	(2,8)

Figur 2: Minimax values

we have that $\underline{\pi}_1 = 4$ and $\underline{\pi}_2 = 2$.

Definition 5 (feasible, rational payoff) *A payoff* $(\pi_i)_i$ *is* **feasible, rational** *if there exists a feasible outcome* $(s^t)_t$ *such that the average payoff is* π_i *and* $\pi_i \geq \underline{\pi}_i$ *for all* i*.*

A feasible, rational payoff takes into account the fact that any player can atleast obtain a value equal to his minimax value; and thus, a player cannot be forced to obtain an average less than this value.

We are now ready to state the Folk theorem in the case of discounting payoff.

Proposition 4 (Folk Theorem) *Let* (π_1, π_2) *be feasible, rational then there exists* $\delta < 1$ *and* $s^\star$ *a SPE, such that the average payoff is* π_i*.*

The prove uses the trigger strategies as we applied in the example of Prisoners Dilemma above, to show that (M, M) was sustainable as a SPE. These trigger strategies are very effective. basically, they tell us that cooperative behaviour can be sustained be a very harsh threat: "either you behave nice and cooperatively, or we will punish you forever!". This is a behaviour that we also observe in reality, or perhaps only a punishment in a finite period, both with humans and animals.

One can find graphically the set of payoffs feasible and rational in the case of finite strategy sets. First, we note that the set of feasible payoffs is the convex hull of the payoffs of the stage game. This is intuitively clear: since there is a finite set of strategy profiles there is a finite set of payoffs. But then for every outcome ξ, for every T we can find numbers $n_T(s)$ for each $s \in S$ such that $\sum_s n_T(s) = T$ and

$$\sum_{t=0}^{T} u_i(\xi^t) = \sum_{s \in S} n_T(s) u_i(s)$$

and thus the average payoff is

$$\lim_{T\to\infty}\frac{1}{T}\sum_{s\in S} n_T(s)u_i(s) = \sum_{s\in S}\alpha(s)u_i(s)$$

with $\alpha(s) = \lim_T \frac{n_T(s)}{T}$ and $\sum_s \alpha(s) = 1$. This is the content of the next proposition

Proposition 5 *The closure of the set of feasible payoffs is* $\mathrm{conv}\{(u_i(s))_i \mid s \in S\}$.

Proof:

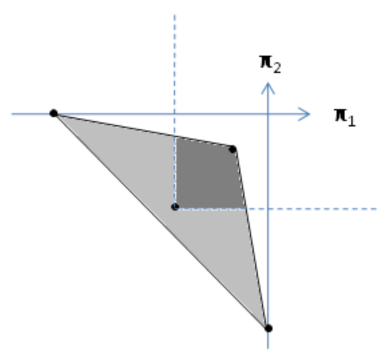
Assume that $\pi_i = \sum_{s\in S}\alpha(s)u_i(s)$ for some $\alpha(s) \geq 0$ rational number and $\sum_s \alpha(s) = 1$. Then $\alpha(s) = \frac{q_s}{p_s}$ for some integers q_s, p_s, and we let $T = \prod_s p_s$ and take $n_T(s) = q_s \prod_{s'\neq s} p_{s'}$, then we have that for every T

$$\frac{1}{T}\sum_{t=0}^{T} u_i(\xi^t) = \frac{1}{T}\sum_{s\in S} n_T(s)u_i(s) = \sum_{s\in S}\frac{q_s}{p_s}u_i(s) = \sum_{s\in S}\alpha(s)u_i(s).$$

But every real number can be approximated by a rational number, and thus the closure is the convex hull.

□

The set of feasible average payoffs is illustrated in figure 3 as the light-gray area, while the set of feasible, rational payoffs is the dark-gray area.



Figur 3: The set of feasible and rational payoffs

The set of feasible, rational payoffs are those feasible payoffs that is not lower than the minimax value of each player.