Cooperation and communication dynamics
Session 4

Olivier Gossner

PSE

Lessons from the Chain Store game

The introduction of "crazy types" in the Chain Store game shows that

- A "small" perturbation of a repeated game can lead to equilibrium outcomes that are drastically different from the original game. Questions the robustness of conclusions wrt. modeling assumptions.
- Reputations can be captured by the introduction of crazy types.

We study reputation games in which

- An infinitely lived long-run player faces a series of short-run players
- The long-run player may either be of a normal type, with known stage payoff function and discount factor δ, or of a commitment type who repeatedly plays a given commitment strategy

Plan

We first review long-run vs. short-run games w/o crazy types

Eq. payoffs are our benchmark

With pure commitment strategies and perfect monitoring

we study Perfect Markov Equilibria

With mixed commitment types and imperfect monitoring

we provide bounds on equilibrium payoffs

Roadmap

An long-run vs. short-run example

Consider the following quality choice game:

	h	l
H	$2, 3$	$0, 2$
L	$3, 0$	$1, 1$

- player 1 is a long-run player, with discount factor δ,
- there is a different short-run player 2 each period.

What can be said about the NE payoffs to player 1?

Another example

Consider the following modification of the quality choice game:

	h	l
H	$2, 3$	$1, 2$
L	$3, 0$	$0, 1$

- Trigger strategies no longer constitute an equilibrium
- We can construct simple strategies that implement Hh forever
- Using self-generation techniques, it is possible to show that the set of perfect equilibrium payoffs to player 1 goes to $[1, 2]$ as $\delta \to 1$.

Roadmap

Reputations in the quality choice game

Consider the quality choice game

	h	l
H	$2,3$	$0,2$
L	$3,0$	$1,1$

- With pba. α, player 1 is a commitment type who always plays H.
- With pba. $1-\alpha$, player 1 is a normal type whose payoffs are given by the matrix, with discount factor $\delta > \frac{1}{2}$.

Reputations in the quality choice game

We look for a Markovian equlibrium, $(\sigma(p), \tau(p))$ where p is the belief that P1 is of type H. Let $v(p)$ be P1's eq. payoff when beliefs are p.

- Note that $v(0) = 1$
- If $\alpha > \frac{1}{2}$, p2 plays H, hence $v(p) \geq 2$

Let $\beta(\alpha)$ be the posterior belief that P1 is of type H after observing H.

- If p2 plays l with positive probability, $\beta(\alpha) \geq 2\alpha$
- If $\alpha > 0$, there exists k s.t. $v(\beta^k(\alpha)) \geq 2$

Assume $v(\beta^{k+1}(\alpha)) \geq 2$. At $\beta^k(\alpha)$, playing H gives at least:

$$(1-\delta)2\tau(\beta^k(\alpha))(h) + \delta.2$$

and L yields

$$(1-\delta)(1 + 2\tau(\beta^k(\alpha))(h)) + \delta.1$$

Assume $\delta > \frac{1}{2}$

- $\sigma(\beta^k(\alpha)) = H$,
- $\tau(\beta^k(\alpha)) = r$,
- $v(\beta^k(\alpha)) \geq 2$.

By induction, for $\alpha > 0$, $v(\alpha) \geq 2$ and $\sigma(\alpha) = H$.

Conclusion

For $\delta > \frac{1}{2}$, there exists a unique Perfect Markovian Equilibrium:

1. $\sigma(0) = 0$, $\tau(0) = l$, $v(0) = 1$.
2. For every $\alpha > 0$, $\sigma(\alpha) = H$, $\tau(\alpha) = h$, $v(\alpha) = 2$.

Roadmap

The basic learning question

- Z finite space of agent's observations
- P law of a process $(z_t)_t$ with values in Z
- Q agent's belief on $(z_t)_t$

Next stage predictions following $z_1 \dots z_{t-1}$

- $p_t = P(z_t | z_1 \dots z_{t-1})$ stage t's law conditional on the past
- $q_t = Q(z_t | z_1 \dots z_{t-1})$ agent's prediction at stage t
- p_t, q_t are r.vs. in $\Delta(Z)$

Does the agent eventually make accurate predictions?

When and in what sense does q_t "converge" to p_t?

Example: P iid. coin tosses, $p \in [0, 1]$. Q puts uniform probability on p

Entropy of a distribution

- X finite set, $p \in \Delta(X)$
- Amount of "surprise" in seeing a realization x

$$\log \frac{1}{p(x)}$$

- Expected amount of surprise, entropy of p

$$H(p) = \sum_{x} p(x) \log \frac{1}{p(x)}$$

  log is $\log_2$ by convention, $0 \log(\infty) = 0$ by continuity
- $H(p)$ measures the "randomness" of a r.v. with distribution p, or equivalently the amount of information contained in its observation

Relative entropy

- $p, q \in \Delta(X)$: p real distribution, q agent's belief

- Expected amount of surprise of the agent with belief q

$$\sum_x p(x) \log \frac{1}{q(x)}$$

$$\sum_x p(x) \log \frac{1}{q(x)} \geq H(p)$$

- The relative entropy is the difference

$$d(p\|q) \quad = \quad \sum_x p(x) \log \frac{p(x)}{q(x)}$$

- It is an information theoretic measure of the agent's prediction error

Fundamental property: Chain Rule

- (x, y) drawn in $X \times Y$ with law P
- Agent's belief on (x, y) is Q

Relative entropy at once

The error in predicting (x, y) is $d(P\|Q)$

Relative entropy in two stages: Assume x is observed, then y

- P_X, Q_X marginals of P, Q on X
- The total expected error in predicting x, then y, is

$$d(P_X\|Q_X) + \mathrm{E}_{P_X}\, d(P(\cdot|x)\|Q(\cdot|x))$$

Chain Rule

$$d(P\|Q) = d(P_X\|Q_X) + \mathrm{E}_{P_X}\, d(P(\cdot|x)\|Q(\cdot|x))$$

Consequences of the Chain Rule

Relative entropy under grain of truth, $Q = \mu P + (1-\mu)P'$

$$d(P\|Q) \leq -\log \mu$$

Total expected prediction error under grain of truth

Let $Q = \mu P + (1-\mu)P'$ on $Z^{\mathbb{N}}$, for every $T \geq 1$

$$\sum_{t=1}^{T} \mathrm{E}_P\, d(p_t\|q_t) \leq -\log \mu$$

Expected δ-discounted prediction error

$$(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1}\, \mathrm{E}_P\, d(p_t\|q_t) \leq -(1-\delta)\log \mu$$

Roadmap

Reputations (Fudenberg Levine 1989, 1992)

Model

- Long-run player 1 facing short-run players 2
- Action spaces A_1, A_2, payoff functions $g_i \colon A_1 \times A_2 \to \mathbb{R}$
- Long-run of behavioral type with probability μ, or normal type
- Commitment type repeats $\hat{s}_1 \in \Delta(A_1)$
- Normal type uses discount factor δ
- Signal spaces Z_1, Z_2, probability of signals $q(z|a) \in \Delta(Z_1 \times Z_2)$
- Each player 2 knows the history of past signals to previous players 2

Questions

- Can the long-run player build a "reputation" for playing $\hat{s}_1$?
- Bounds on NE payoffs to player 1: asymptotic? explicit?

Example: Quality choice game

A long-run (cook, firm) may exert effort and produce a produce a high quality good, or produce a low quality good at no cost. Short-run consumers may decide to buy the product, or not.

	b	n
H	$1, 1$	$-2, 0$
L	$3, -1$	$0, 0$

Relating errors and payoffs

- player 1 plays $\hat{s}_1$
- player 2 plays s_2, BR to his belief s_1
- player 2's prediction error in his own signal is

$$d(q(z_2|\hat{s}_1, s_2)\|q(z_2|s_1, s_2))$$

Min payoff to P1 from $\hat{s}_1$ if P2 makes an error of at most ε

$$v_{\hat{s}_1}(\varepsilon) = \inf g_1(\hat{s}_1, s_2)$$

$$s_2 \text{ BR to some } s_1,\ d(q(z_2|\hat{s}_1, s_2)\|q(z_2|s_1, s_2)) \leq \varepsilon$$

Assume P1 plays $\hat{s}_1$, P2 plays a BR to his beliefs

Let $p_t = q(z_{2,t}|\hat{s}_1, s_2)$, $q_t = q(z_{2,t}|s_1, s_2)$, $g_{1,t} = g_1(\hat{s}_1, s_{2,t})$

1. $$g_{1,t} \geq v_{\hat{s}_1}(d(p_t \| q_t))$$

2. $$(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} \, \mathrm{E}_P \, d(p_t \| q_t) \leq -(1-\delta)\log \mu$$

Let $w_{\hat{s}_1}$ be the largest convex mapping below $v_{\hat{s}_1}$:

$$(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} \, \mathrm{E}_P \, g_{1,t} \geq w_{\hat{s}_1}(-(1-\delta)\log \mu)$$

Theorem

The worst Nash Equilibrium payoff to player 1 is at least

$$w_{\hat{s}_1}(-(1-\delta)\log \mu)$$

Consequences

Assume that there are several behavioral types, $\hat{s}_1$ with probability $\mu(\hat{s}_1)$

Theorem

The worst Nash Equilibrium payoff to player 1 is at least

$$\sup_{\hat{s}_1} w_{\hat{s}_1}(-(1-\delta)\log \mu(\hat{s}_1))$$

Let $N(\delta)$ be the worst NE payoff to player 1.

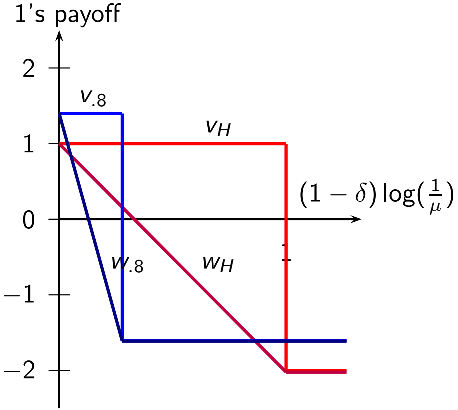
Corollary [Fudenberg Levine 1989, 1992]

If the set of $\hat{s}_1$ such that $\mu(\hat{s}_1) > 0$ is dense in $\Delta(A_1)$, then

$$\liminf_{\delta \to 1} N(\delta) \geq \sup_{\hat{s}_1 \in \Delta(A_1)} v_{\hat{s}_1}(0)$$

Example: Quality choice game

	b	n
H	$1, 1$	$-2, 0$
L	$3, -1$	$0, 0$



$\hat{s}_1 = H$

- $d(H \| \frac{1}{2}) = \log 2 = 1$
- $v_H(d) = 1$ if $d < 1$
- $v_H(d) = -2$ if $d \geq 1$

$\hat{s}_1 = \hat{p}H + (1-\hat{p})L$, $\hat{p} > \frac{1}{2}$

- $d(\hat{p} \| \frac{1}{2}) = 1 - H(\hat{p})$
- $v_{\hat{p}}(d) = 3 - 2\hat{p}$ if $d < 1 - H(\hat{p})$
- $v_{\hat{p}}(d) = -2\hat{p}$ if $d \geq 1 - H(\hat{p})$

Conclusions

Perfect monitoring and pure commitment types: $\exists \delta_0$, for $\delta > \delta_0$

- perfect Markov equilibria all give the same equilibrium path
- the long-run player gets the Stackelberg payoff at every stage
- no matter what is the probability of commitment types, if > 0

Imperfect monitoring general commitment types,

- Fixing a probability of commitment types with "full support"
- When the long-run becomes arbitrarily patient
- All NE give the long-run player at least a payoff arbitrarily close to the Stackelberg payoff