Lecture Notes for 14.126

These notes were designed to accompany a classroom presentation and may be incomplete without it.

Kandori
"Social Norms and Community Enforcement"

Ellison "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching"

Closely related papers, Kandori came first, posed some very interesting questions.
Glenn's paper follows up with much better answers, and in the process introduced a small trick that has been used in several other papers.

<u>Idea</u>: agents play repeated games, but against different people each day, and the community is large enough that agents don't directly observe what everyone else is doing. Standard arguments about cooperative play don't apply. Can cooperative play be sustained as an equilibrium in such an environment? If so, how?

Kandori

N player 1's, N player 2's
Each period, everyone matched to play a stage game.
Uniform random matching.

payoff matrix

	C	D	
C	$1,1$	$-\lambda, 1+g$	$\lambda, g > 0$
D	$1+g, -\lambda$	$0,0$	

Suppose agents only observe history of play in their own matches. So strategies can't single out deviators for punishment.

Instead, look for a "contagion equilibrium":
Agents start out playing C . Stick with C unless/until they see anyone play D, then switch to D forever. (society is collapsing, sauve qui peut, etc.)

<u>Theorem 1</u> (Kandori) for fixed g and N, the "contagion strategy" is a sequential equilibrium if δ and λ are large enough.

Intuition:

On-path deviations don't pay if delta large, since deviating leads to eventually all D

Off-path deviations are trickier to handle, this is where the assumptions on M and g come in.

To see the problem, consider an agent who has met a deviator- or deviated himself-so that the contagion has begun.

By playing C instead of D, the agent can slow down the spread of contagion. And this might be desirable- that is, the "contagion" strategy may be too strong in stopping cheating, as it may also stop people from carrying out the punishment phase.

Payoffs at such a history depend on the fraction of the population the agents thinks is currently playing D, since by C gives a loss against *D*.

Because agents only see the outcomes of their own matches, they know how many others are deviating. But if λ is very large, even having one other person playing D is too big a risk, and players want to play D.

This is why the theorem works, but its hypothesis is pretty strong. One of the things Ellison does is show that the assumption of large λ isn't needed.

Proposition 1 (Ellison): For any $g, \lambda > 0$ there is a $\underline{\delta} < 1$ such that for all $\delta > \underline{\delta}$ there is a sequential equilibrium where all agents play C on the equilibrium path.

The proof uses public random device: a iid sequence $\{q_t\}$ that is uniform on [0,1]; paper later shows how to do the same thing w/o it.

Strategies:
Begin in Phase I

In phase I: play C

If outcome at date *t* is (C,C) stay in I

If not, then goto II if $q_{t+1} \leq q(\delta)$ else stay in 1.

Phase II: play D

Go to phase I if $q_{t+1} > q(\delta)$

Kandori's strategies correspond to $q = 1$.
Here we know that for large δ no on wants to deviate on the path, but they might want to deviate off path

q= 0 is the other extreme. Here IC satisfied in phase II but not in I. So the goal is to show that both constraints are satisfied at some intermediate value.

Intuition: Regardless of a player's own history, playing C instead of D is worse if the current opponent is in phase II (and so playing D).

If the opponent is in Phase I (and so playing C) then the comparison between C and D depends on the player's own state: D is relatively better if the player is in phase II, since then he knows that cooperation is breaking down anyway, so there is less impact from one more deviation.

To formalize this, let k be number of agents playing according to II at t.

Let $f(k, \delta, q)$ be the per- period payoff from playing D when this agent and k-1 others are in phase II.

Key lemma: f is convex in k.

Now consider the condition for no profitable deviations in II:

a) meet someone else who is also in phase II. Then deviating to C can't help.

b) meet someone in phase I. Then deviating to C would cost g today, but there would be 1 less infected agent tomorrow.

The agent doesn’t know the number k who are infected- so the IC has an expectation over *k*- but it is sufficient (not necessary) to show that playing D is better than C when meeting an agent in phase I, pointwise over the number k of agents in phase II:

$$(1-\delta)g \geq \delta q(\delta)(f(j,\delta,q(\delta)) - f(j+1,\delta,q(\delta)):$$

call this (*)

(playing D gives a gain of g today, with probability $q(\delta)$ tomorrow is still a punishment state, and the number of infected agents has gone up by 1.)

The condition for no deviations in phase 1 is

$(1-\delta)g \le \delta q(\delta)(1-f(2,\delta,q(\delta)))$. (%)

(Along the equilibrium path, agents assign probability 1 to everyone being in phase I. If the agent plays D, then with probability $q(\delta)$ there will be at two agents in phase II tomorrow; the agent's strategy says to play D in phase II.)

Since f is convex, if we pick q to make (%) hold with equality, (*) holds for all j>1.

The theorem says "there exists $\underline{\delta}$;" the proof shows how to compute it. The biggest that q can be is 1, so $\underline{\delta}$ is the value of delta where with q= 1 the players are just indifferent about cooperating on the equilibrium path.

Numerical calculations :

M= 100 g =1 need $\delta = .85$; g = 10 need .$\delta = .985$

(λ doesn't enter in to it)

The contagion spreads exponentially. So doubling the frequency of meeting-which is the same as taking the square root of the discount factor- lets the population size go up as the square.

The theorem is also only for the prisoner's dilemma. Ellison explains how to extend to other games with a strictly dominant strategy.

The reason having a dominant strategy matters is that then the maximum one-period gain to cheating on the equilibrium path is the same as the short tem loss players must incur to slow the spread of the contagion.

In other games, we might use reversion to a static Nash equilibrium as the "punishment phase," but this strategy won't be the same strategy as most tempting deviation from the equilibrium path, which adds an extra complication

Robustness: Kandori points out that the strategies he constructs aren't "robust to noise": if there is some small chance that C looks like D, and player use the above strategies, then players will end up in all D, and if patient their LR payoff is about 0.

Kandori says that strategies are "globally stable" if their continuation values return to the equilibrium payoffs along any path:

$$\lim_{t\to\infty} E(v_i(t) \mid h_\tau) = v_i^* \text{ for all i, h}$$

where v_i^* is the equilibrium payoff.

He uses this to motivate attention to "local information processing" where players have "labels" that are updated as a result of how they play. Interesting interpretations of these structures in terms of things like credit bureaus and other institutions.

But Ellison raises some prior points about global stability and the equilibria of the no-labels model with noise.

- If there is noise, the players care about expected present values, and an equilibrium can be globally stable yet have crummy expected present value if it is very slow at returning to cooperation.
- Conversely, an equilibrium might not be globally stable, and yet have a high expected present value.

Ellison shows that there are approximately efficient equilibria in the model with noise.

Idea: Kandori’s construction used “grim” strategies, so any tremble lead to everyone playing D forever. But this is "more punishment" than is needed for incentives; Ellison's strategies use finite length punishments with a public randomizing device to "reset the clock."

Formally, add noise by assuming that players are constrained to play D with probability $\varepsilon > 0$ after every history.

Proposition 2: There exists $\underline{\delta} < 1$, and strategies $s^*(\delta)$ for $\delta \in [\underline{\delta}, 1)$ s.t.:

1) $s^*(\delta)$ is a sequential equilibrium of the unperturbed game. Along the path of $s^*(\delta)$, all players play C .

2) there is $\overline{\varepsilon} > 0$ s.t. for all $\varepsilon < \overline{\varepsilon}$, the strategy $s^{**}(\delta, \varepsilon) = (\varepsilon D, (1-\varepsilon) s^*(\delta))$ is a sequential equilibrium of the game with ε noise.

3) $\lim_{\varepsilon \leftrightarrow 0} \lim_{\delta \to 1} u(s^{**}(\delta, \varepsilon)) = 1$.

(recall that 1 is the cooperative payoff)

So for a fixed ε, the equilibrium payoff needn't be efficient (unlike in repeated game) but the inefficiency vanishes as $\varepsilon \to 0$.

<u>Idea</u>: 1) Start in the unperturbed (that is, no-noise) game, set $q(\delta)$ as before to make players indifferent on the path, then pick a $q^*(\delta)$ that is just a bit bigger than $q(\delta)$, so that IC holds strictly in phase I but phase II constraints still hold. Thus we still have an equilibrium of the unperturbed game.

2) Then use two sorts of continuity argument to extend to the perturbed game

a) small epsilons make epsilon changes in the continuation payoffs, show that IC still hold given the beliefs in the equilibrium of the unperturbed game.

b) Those beliefs aren't compatible with sequential equilibrium in the perturbed game, since even in phase I, a player can't think there is probability 1 that others are in Phase I. This introduces anther small epsilon that we have to show dosnt matter.

The continuity arguments establish claim 2: the strategies constructed are a sequential equilibrium of the game with noise.

Claim 3 follows because the punishment phases for δ near 1 have a length that is bounded independent of δ; the length of these phases is approximately $1/(1-q(1))$, since every time the public randomization comes up 'heads" we reset the strat back to the coop phase.

The continuity argument in step 2 relies on ε being smaller than some $\bar{\varepsilon}$. This $\bar{\varepsilon}$ shrinks (and gets very small!) as the population size grows.

Ellison speculates that in a local interaction model it might be possible to construct approximately efficient equilibria where $\underline{\delta}$ and $\bar{\varepsilon}$ can be chosen independent of the population size.

Ellison's construction relies a lot on public randomization, and this might make you think that the public randomizing device was doing too much of the work.

So the question: are there approximately efficient equilibria that don't use public randomizations?

It turns out that there are. The key role of the public randomizations in the proofs was to soften the "grim" punishment enough to get people to play D off path while still playing C in phase I.

The same thing can be done by spreading punishments over time.

<u>Lemma</u>: Let $G(\delta)$ be any repeated game of complete information. Suppose there is a non- emty interval $[\delta_0, \delta_1]$ such that $G(\delta)$ has a sequential equilibrium $s^*(\delta)$ with outcome a for all $\delta \in [\delta_0, \delta_1]$.
Then there exists a $\underline{\delta}$ such that for each $\delta \in [\underline{\delta}, 1]$, there is a strategy profile $s^{**}(\delta)$ which is a sequential equilibrium of $G(\delta)$ with outcome a.

Idea: Start with an equilibrium profile s^* for $\delta = 4/9$. Now suppose $\delta = 2/3$, and construct strategies s^{**} that treat odd and even periods as if they were in different games- play in odd periods depends only on play in past odd periods.

Then the effctive discount factor in each "subgame" is 4/9, so s^{**} is an equilibrium for $\delta = 2/3$.

Proof: take $\underline{\delta} = \delta_o / \delta_1$.

Claim: For any $\delta \in (\underline{\delta}, 1)$ there is an integer $N(\delta)$ for which $\delta^{N(\delta)} \in (\delta_0, \delta_1)$.

Proof of claim:

For any $\delta > \underline{\delta}$, let N be the largest integer such that $\delta^{N(\delta)} \geq \delta_0$; this will be a finite integer greater than 0.

By definition, $\delta^{N(\delta)+1} < \delta_0$.
Suppose that $\delta^{N(\delta)} \geq \delta_1$.

Then $\delta = \dfrac{\delta^{N(\delta)+1}}{\delta^{N(\delta)}} < \dfrac{\delta_0}{\delta_1} = \underline{\delta}$, a contradiction.

Now have players treat game $G(\delta)$ as if it were $N(\delta)$ separate games. That is, if N = 7, play on Mondays depends only on what happened in past Mondays, etc. This proves the lemma.

Now lets apply the lemma to the propostion about approximately efficient equilibria: that is the proposition that for all sufficiently large δ there are sequential equilibria $s^{**}(\delta,\varepsilon)$ such that $\lim_{\varepsilon\to 0} \lim_{\delta\to 1} u(s^{*}(\delta,\varepsilon)) = 1$.

The strategies used there do not use public randomization when $q(\delta)=1$, which is the case when the discount factor equals the $\underline{\delta}$ of the proposition.

To apply the lemma, we need to show that we can use $q=1$ for a range of δ the neighborhood of this $\underline{\delta}$.

The constructed strategies had strict inequalities in all the IC constraints so this is immediate.

So far we've been analzing the equilibria of various games.

Next Sergei will start talking about
models of learning/evolution/adjustment

that can explain why and when we might expect play to actually look like an equilibrium..

Remember: rationality, even common knowledge of rationality, doesn’t in general imply play must look like a NE. There are special cases where a form of iterated dominance gives a unique solution, but many if not most games of interest have multiple equilibria.

And in a game with multiple equilibria, there is no reason to expect play to look like <u>any</u> of the equilibria without some explanation for how players might all end up playing the <u>same</u> equilibrium. (*can’t “mix and match” strategies from various equilibria.)*

The answer we'll look at is that equilibrium, when it occurs, is the long-run result of some process of adaption that we can variously think of as learning, evolution, imitation, etc.

There are many versions of this idea.

One is that individual agents are hard-coded to play in certain ways, and that individuals never change their play at all- instead, the population fractions using each strategy evolve. This is the approach in evolutionary biology; it leads to the "replicator dynamic" and related models,

Alternatively, we can look at models where individual agents do adjust their play, based on their experience or experience of others that they talk to.

In both cases, what we are interested in is the long-run behavior of the resulting dynamic process.

There are many sorts of non-equilibrium adjusment processes, each of which have their own merits and disadvantages, and it can sometimes be hard to assign them all to genres based on a fixed classification system since there are always cases that seem to fall between the lines..

Common themes:

1) Non-equilibrium adjustment.

2) Most formal results will be on LR play:

3) The processes don't always converge.

4) Adjustment processes may suggest equilibrium refinements

5) Players play repeatedly w/o playing a "repeated game."

Explanations:

a) lock-in and impatience: this rationalizes Counot's process and also the myopic BR models.

b) large populations:
either anonymous random matching and only see outcome in own match, or random match and announce population aggregates, or announce aggregate and "play the vs. average " (same as above in EV)

c) hard-coded for stage-game strategies

Each of these defenses has a different domain of applicability. For b), a fuzzy but potentially important idea is "extrapolation across similar games."