LEARNING AND EQUILIBRIUM

Simons Institute Economics and Computation Boot Camp
UC Berkeley
August 2015

DREW FUDENBERG

Today: Static Games.

Each player takes a single action, actions simultaneous.

Tomorrow: Extensive Form Games.

Strategies as "complete contingent plans."

Background and Overview

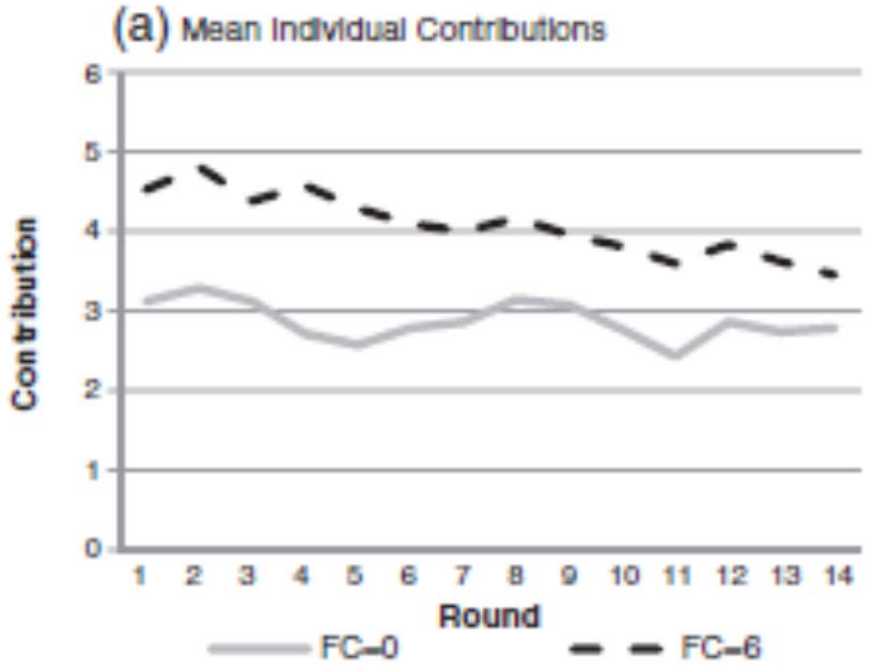
- Rationality (even common knowledge of rationality) is neither necessary nor sufficient for Nash equilibrium. ("NE")

- Not sufficient: In games with multiple NE, no reason for play to look like any of the equilibria unless there is a reason all players expect the same equilibrium.

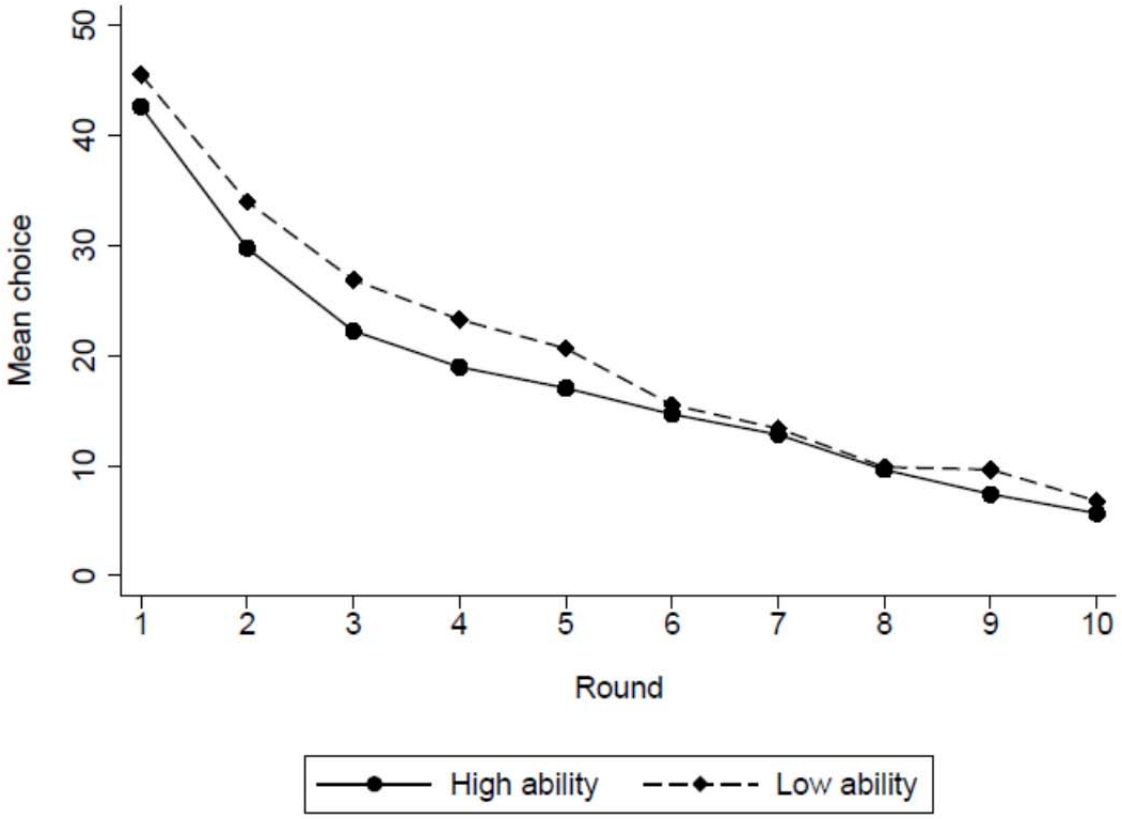
- Not necessary theoretically (replicator dynamic can converge to NE) or empirically (convergence to approximation of NE seen in colonies of bacteria.)

- Learning-Theoretic Explanation: equilibrium arises as long run outcome of a non- equilibrium adaptive process**.**

- Experimental play does converge to Nash equilibrium in a reasonable time frame in some games of interest to economist, including Cournot duopoly, "voluntary contribution" games, the "beauty contest" game, and the "double auctions" used to explain equilibrium prices.



(Vesterlund et al [2011] *J Pub Econ* on two different voluntary contribution games where the Nash equilibrium is for both players to contribute 3.)



Gill and Prowse [2012] on a "beauty contest" game with NE =0

To understand how and when equilibrium arises, look at long-run behavior of non-equilibrium dynamic processes.

- Many sorts of adjustment processes, including biological evolution, have been said to involve "learning" in a broad sense. And it can be hard to draw a bright line between learning and other sorts of adjustment. *Examples: what if everyone knows the current distribution of strategies and plays a best response to it? What if each agent gets a noisy signal of the current distribution?*

- Today : narrow sense of “learning” :
  - Explicitly specify how individual agents use observations to change their behavior
  - Self-interested play by the agent- not agents hoping to (or designed to) find an equilibrium!
  - “Uncoupled” (each agent’s rule independent of payoff functions and rules of the others)
  - Use performance of the rules as a check on their plausibility, not as a goal in itself.

Common themes in the learning-in-games literature:

Non-equilibrium adjustment.

- Pointless to explain equilibrium in a game by assuming equilibrium in some larger adjustment game.
- So **must allow for** players who "are making a mistake" in the sense that their behavior isn't a best response to the behavior of the other.
- The issue is **not** whether a model can generate suboptimal play but rather whether players would notice that some other adjustment rule would be better.

- Focused on LR play
- Convergence to equilibrium not guaranteed.
- Learning can suggest equilibrium refinements.
- Play the given game repeatedly with anonymous random matching. (Note: they can repeatedly play a repeated game as in repeated-game experiments.)
- Learning depends on what the players observe…
- And on whether their actions change what is observed: "passive learning" or "active."

Criteria for Learning Rules

- Hard to empirically determine exact learning rules used, so pick rules using a combination of simplicity, plausibility, and accuracy.

- Different criteria in the literature

  - worst case/minmax considerations (e.g. no regret)

  - Bayesian, Savage-rational utility maximization

  - exogenously specified, "boundedly rational" or "behavioral" (may not differ from Bayesian)

Fictitious Play

Introduced by Brown [1951] as a way to compute equilibrium in two-player games. (hence "fictitious" play.)

Now used as a simple stylized descriptive model of learning. Easy to motivate and analyze, too simple to match experimental data; foundation for more complex models.

Motivation: Suppose that an agent is going to repeatedly play a fixed strategic-form game. The agent knows the structure of the game (the strategy spaces and payoff functions) but not how the other side is going to play. (*can be adapted to less prior knowledge.)*

All that the agent observes is the outcome of play in their own matches. Doesn't observe what happens in other matches, or opponents' past play.

- Agent believes she is playing against a randomly drawn opponent from a large population, doesn't try to influence opponent's play: *strategic myopia*.

- What the agent observes is independent of own action, no incentive for "experimentation": *passive learning*

- Agents act as if they are Bayesian expected utility maximizers, facing a stationary, but unknown, distribution of opponents strategies.

- Stationarity i splausible first hypothesis in many situations.
- Might expect players to stick with it when it is approximately right but to reject stationarity given sufficient evidence to the contrary- as when there is a strong time trend or a high-frequency cycle.
- Stationarity impies all observations equally informative, and that beliefs are *asymptotically empirical*, so that they converge to the empirical distribution as the sample grows.

- In some settings people seem to give less weight to older data (displaying "recency bias") as if facing a hidden Markov process. More on this at end of talk.

In FP the belief updating has a special simple form:

- Player *i* has an initial weight function that gives a weight to each opponent's strategy.
- Add 1 to the weight of each opponent strategy each time it is played.
- Probability that player *i* assigns to -*i* playing s^{-i} is the weight on s^{-i} divided by the sum of the weights.

Fictitious play is any behavior rule that at each history specifies a static best response to these beliefs.

FP has a Bayesian interepretation:

- Player believes opponents' play is sequence of i.i.d. multinomial random variables with a fixed but unknown distribution.

- Prior beliefs: Dirichlet distribution, where the parameters correspond to FP's initial weights. (details at end of these notes.)

- Play maximized expected payoff given beliefs.

- Methodological Point: the distinction between "rational" and "boundedly rational" isn't very firm…

- Conceptual point: beliefs about distribution of opponent's play is the expected value of beliefs over opponent's mixed strategies. The updating depends on more than this expected value: Compare "I'm sure my opponent plays (1/2 H, 1/2T) to "with probability .5 my opponent will play H every period, with probability .5 she will always play T."

- Functional form of FP convenient but not needed for most results.

- Key features are that the player treats the environment as stationary so that beliefs are *asymptotically empirical,* and that players are *asymptotically myopic:* at least in the long run they play best responses to their beliefs.

- Note: in FP (or in any Bayesian scheme) the beliefs differ from the empirical distribution because of the influence of the prior. Over time the data swamps the prior, and the assessments converge to the marginal empirical distributions.

- FP beliefs suppose opponent's strategy is constant; this is wrong if the process cycles or has a time trend.

- If cycles persist, player might eventually notice them.

- But at least his beliefs will not be falsified in the first few periods! (in contrast to Cournot adjustment).

Alternative learning model: (*more popular in psychology)* Reinforcement learning (REL). (cut for lack of time… could add at end..)

The Interpretation of Cycles in Belief-Based Learning

(Fudenberg Kreps *GEB* [1993])

	A	B
A	0,0	1,1
B	1,1	0,0

FP, 1 agent per side, each player's initial weight $(1, \sqrt{2})$

- First period: both think other will play B, both play A.
- Next period weights are $(2, \sqrt{2})$) and both play B; the outcome is the alternating sequence ((B,B),(A,A),(B,B), etc.)

- Empirical frequencies of each player's choices converge to (½, ½), which is the Nash equilibrium. So FP "works" for the purpose of computing equilibrium.

- Not a good model of learning: Both players receive payoff 0 in every period even though each can guarantee a payoff of ½.

Reason: the empirical joint distribution on pairs of actions does not equal the product of the two marginal distributions, so that the empirical joint distribution corresponds to correlated as opposed to independent play.

- Claim: Players would notice this cycle and do something else, maybe form more sophisticated beliefs.

- So in general we won't want to identify a cycle with its average.

- And we may want to worry about how sensitive the players are to correlations in the data.

Important facts about FP:

- If actions converge they converge to a pure strategy NE.

- If time averages of empirical marginals converge the joint distribution is a NE. (*proof sketch: if player 2's marginal converges to* σ_2*, 1's beliefs converge to* σ_2*. If the mixed action* σ_1 *corresponding to the limit of 1's empirical marginal is not a best response to* σ_2*, some* s_1 *is strictly better than any other strategy in the support of* σ_1 *for all large enough times. At such times 1 must play only* s_1)

- The above holds in any belief-based learning model that is asymptotically empirical and "asymptotically myopic" (eventually players choose actions that are myopic best responses to their beliefs.)

- FP "behaves well" when there are "infrequent" changes of play:[1] In this case it is "ε-consistent" (do as well as maximizing vs time average of opponents' play) and it also converges to the best response dynamic. (Monderer, Samet, and Sela *JET* [1997]). Note that the example where FP looked odd had a two cycle.

---

[1] Meaning that for every ε there exists a T s.t. for all $t > T$ the fraction of the periods τ where play at τ differs from that in $\tau - 1$ is at most ε .

Heuristic: Time rescaling and asymptotics of FP

Let empirical distribution of i's play be d_t^i.

Ignoring priors (won't matter for large t) then

$$d_t^i = \frac{t-1}{t} d_{t-1}^i + \frac{1}{t} BR^i(d_{t-1}^{-i}).$$

(make a pure selection when indifferent.)

Now change the units in which time is measured:

$$\tau = \log t \leftrightarrow t = \exp(\tau), \quad \tilde{d}_\tau^i = d_{\exp(\tau)}^i$$

Suppose infrequent switches, so for large enough τ play remains more or less constant between τ and $\tau + \Delta$. (*this rules out any fixed-period 2-cycle, any cycle must be slowing down..)*

Then for large *t* and small Δ can approximate the difference equation by

$$\dot{\tilde{d}}_{\tau}^{i} = BR^{i}(\tilde{d}_{\tau}^{-i}) - d_{\tau}^{i}.$$

When this approximation is valid, can study long run behavior of FP by studying the continous time best response dynamic.

Steady states= NE, and get stability and convergence conditions.

Stochastic (or "Smooth") Fictitious Play

(Fudenberg-Kreps *GEB* [1993])

Like FP but with a smooth (continuous) "stochastic best response function" that assigns a mixed strategy response to each belief in place of the exact best response on FP.

Advantages:

- If beliefs converge behavior does too; not the case with standard fictitious play

- Allows convergence to mixed-strategy equilibria in fictitious play-like models: Actual play in FP can't converge to a mixed equilibrium.

- Avoids the discontinuity in standard fictitious play, where a small change in the data can lead to an abrupt change in behavior.

- Discontinuous responses
  - may not be descriptively realistic
  - can lead to “frequent switches” and so poor worst-case performance,
  - create technical complications with the ODE approximation, which becomes a differential inclusion.

- There is a (non-Bayesian) sense in which stochastic rules perform better than deterministic ones- they can be "ε-consistent."

- "Harsanyi-purification" foundation based on private payoff shocks: Let η^i be the random shock to i's payoff, then player *i*'s stochastic best response to mixed strategy profile σ^{-i} for i's opponents is

$$\overline{BR}^i(\sigma^{-i})(s^i) =$$

$$\text{Prob}[\eta^i \text{ s.t. } s^i \text{ is a best response to } \sigma^{-i}]$$

- Stochastic responses also arsie from indecision or ambiguity-aversion on the part of the agents.

- Intersection of the smooth best response curves may be far from any Nash equilibria of the original game, but converge to them Nash equilibria of the unperturbed game as random components converge to 0 in probability (Hofbauer and Sandholm *Ema* [2002]).

Observation: If v^i is a smooth, strictly differentiable, concave function on the interior of Σ^i whose gradient becomes infinite at the boundary, then argmax $u^i(\sigma) + v^i(\sigma^i)$ is a smooth best response function, and the argmax assigns positive probability to each of i's pure strategies.

Call such v^i "admissible perturbations"

If *v* is bounded and $\sup_{i,\sigma^i,\sigma^{i'}} | v^i(\sigma) - v^i(\sigma^{i'}) |$ is small, fixed points of BR are become close to the Nash equilibria.

Canonical example: $v^i = \beta \sum \sigma_i \ln \sigma_i$ is entropy;
generates logit best responses

(Fudenberg, Iijima and Strzalecki *Ema* [2016] characterizes the revealed- preference implications of a subclass of these perturbations, and show they correspond to ambiguity-aversion by an agent who is afraid of making the wrong choice.)

In smooth/stochastic FP, players form beliefs as in FP, play smooth best responses.

Stochastic Approximation:

- **D**etermine long-run behavior of *discrete-time stochastic* systems with 1/t step size by analyzing related *deterministic continuous time* systems.

- Applied to stochastic FP we can use this to relate the long-run outcome to Nash equilibria of the perturbed game- i.e. the intersection of the smooth best response functions.

- Can be applied to any system (like REL) whose evolution can be determibed from the empirical distribution of outcomes…

Some definitions for reference…

The ω-limt set of a sample path $\{\theta_t\}$ is the set of long-run outcomes:
y is in the ω-limit if there is an increasing sequence of times $\{t_k\}$ such that $\theta_{t_k} \to y$ as $k \to \infty$.

A *flow* on X is a continuous function $\Phi : X \times R \to X$ such that $\Phi_0(x) = x$ and $\Phi_s(\Phi_t(x)) = \Phi_{t+s}(x)$.

(*X a subset of finite dimensional Euclidean space.)*
(the solution of a differential equation is a semi-flow: time is non-negative.)

Extend to image of sets: $\Phi_t(A) = \{\Phi_t(x) : x \in A\}$

Invariant set : A s.t. $\Phi_t(A) \subseteq A$ for all t .

Attractor: A is non-empty, compact, invariant, and "attracts" a nghbd W:
$dist(\Phi_t x, \mathrm{A}) \to_t 0$ uniformly for $x \in W$

Consider the discrete-time stochastic system

$\theta_{t+1} - \theta_t = (F(\theta_t) + \eta_{t+1} + b_{t+1}) / (t+1)$, where

- F is C^2
- the η_t are mean-zero noise terms: $E[\eta_{t+1} \mid \theta_t, \ldots \theta_1] = 0$; bounded variance (and as needed bounds on additional moments)
- b_t converges to 0 a.s.

Ideas:

- 1/t step size- so absent the stochastic terms we expect a continuous-time limit when things are "well behaved."
- Limit system deterministic because weight on shocks is 1/t, so use variant of LOLN.
- b_t are nuisance terms that vanish asymptotically
- Limit is a continuous time system after time rescaling.

Important: the η_t don't need to be independent or even exchangeable.

First step: show that almost surely the sample path lies in some invariant set of the continuous-time process.

Proposition: (Benaim and Hirsch *GEB* [1999]) With probability one, the ω-limit set of any realization of the discrete- time process is an invariant set of the continuous-time process; this set is compact, connected, and contains no proper subsets that are attractors for the continuous-time process. (so it is connected and "internally chain recurrent.")

(Benaim- Hirsch have $b_t = 0$, easy extension in Fudenberg-Takahashi *GEB* [2011].)

Benaim-Hirsch : If the stochastic system eventually converges to a point or a cycle, the point or cycle should be a closed orbit of the continuous-time dynamics.

Moreover, the noise will eventually "kick" the system away from any unstable states, at least if there is "enough noise."

Applying Stochastic Approximation to Stochastic FP

Following Benaim-Hirsch assume only one agent per player role, even though this undercuts the idea that agents don't treat this as a repeated game; Fudenberg-Takahashi extend to large populations.

- State space is the empirical distribution of play
- F is $\overline{BR}(\theta) - \theta$.
- Noise terms are the differences between the expected value of $\overline{BR}(\theta_t)$ and its realized value; so (ignoring the prior) the noise terms have a conditional expectation of zero, but are not in general i.i.d. or even exchangeable.

Illustration: 2 players, 2 actions each.

Let $\theta = (\theta_1, \theta_2)$ where θ_i is the empirical fraction of the time that i takes his first action. Theorem says we can determine long run behavior by study of

$$\dot{\theta} = \overline{BR}(\theta) - \theta = \begin{matrix} \overline{BR}_1(\theta_2) - \theta_1 \\ \overline{BR}_2(\theta_1) - \theta_2 \end{matrix}.$$

(here players respond to empirical distribution and not their beliefs- but the two are close when players have a lot of observations, so can correct with a nuisance term b_t)

More generally, the *mean field* for smooth fictitious play is the *smooth best response dynamic* $\dot{\theta} = \overline{BR}(\theta) - \theta$.

Theorem (Benaim and Hirsch): Smooth fictitious play converges to the Nash distribution in any game where the (unique) Nash distribution is a global attractor for the continuous-time dynamics.

- In 2x2 games with unique mixed equilibrium (like "matching pennies"") the process must converge to somewhere near the mixed equilibrium when the payoff perturbations are small.

- What about 2x2 games with 2 strict equilibria and one mixed?

  Benaim Hirsch show that SFP can't cycle in 2x2 games as it is "volume contracting." So it must converge to a steady state. Which?

Proposition (Benaim and Hirsch) If every strategy profile has positive probability at every state, and θ^* is an asymptotically stable equilibrium of the continuous time process, then $P[\theta_t \to \theta^*] > 0$.

The mixed equilibrium is unstable, and when payoff perturbations are small the only stable equilibria are near the NE.

Conclusion: In battle of the sexes, the two pure equilibria have positive probability and the mixed equilibrium has probability 0.

What about other games?

Hofbauer-Sandholm *Ema* [2002] show no cycles in "potential games:"

Potential game: we can transform payoffs w/o changing best responses so that the game is a team problem: $u_i(s) = u_j(s)$ for all *i,j,s.*

In other games continuous-time smooth best response dynamic does cycle. This corresponds to the underlying discrete-time system cycling slower and slower. (*remember the time rescaling !)*

Cycles can be stable, even in presence of noise and even when agents move at different rates.

SFP with Heterogeneous Agents

- Non-strategic behavior in SFP doesn't make sense with one agent per role.

- Fudenberg-Takahasi (GEB 2011) extend to case of a large populations of agents each of whom only sees outcomes of own matches.

- Most interesting model: one population, "asynchronous clocks": a pair of agents is selected at random each period, so ex-post some agents may play more often than others.

- Derive a continuous time deterministic system with heterogeneous beliefs using stochastic approximation: here we track the beliefs of each individual agent.

- Argue that *if the system is sufficiently "well mixed"* it converges to homogeneous (identical) beliefs.

- And from there to an attractor of the lower-dimensional homogeneous-belief system.

- So convergence, local stability etc. on the smaller space implies same on the larger one. In particular if a cycle is a global attractor for the homogeneous system the cycle is also an attractor with heterogeneous agents.

- Well mixed: the interaction probabilities faced by different agents "not too different."

  Formally, let

  - p_{ij} : per-period probability the pair (*i, j*) plays ($p_{ii} = 0$.)
  - $q_{ij} = p_{ij} \Big/ \sum_k p_{ik}$: conditional probability i plays j.
  - $\Delta = \max_{1 \le i \le j \le M} \sum_{k=1}^{M} | \, q_{ik} - q_{jk} \, | \, /2$:

    (minimized by uniform random matching, disjoint network has $\Delta = 1$)

- K: Lipschitz constant for the smooth best response map.

.

- Beliefs and play converge to homogeneous limit when $K\Delta < 1$.

Rules out most interesting network structures. We don't know what happens on more general networks- maybe different regions can play different strategies? Maybe there are “waves”?

Smooth Fictitious Play is Universally Consistent

Say that a learning rule is *consistent along a history* if it yields at least the payoff of optimizing against the time-average of the history.

Deterministic fictitious play is approximately consistent for histories that satisfy the "infrequent switching condition."

When infrequent switching is not satisfied, players can do worse than they could have guaranteed by randomizing in every period, as in the example where players switched every period and always mis-coordinated.

This suggests two properties a learning rule might have:

1) S*afety*: the player's realized average utility is almost surely at least his minmax payoff regardless of the opponents' play. This property failed in the example of frequent switching.

Or the stronger condition of

2) *universal consistency*: regardless of opponents' play, player almost surely gets at least as much utility as could have gotten if had known the frequency but not the order of observations in advance. ("no regret")

(Since the utility of a best response to the actual frequency distribution must be at least the minmax payoff, universal consistency implies safety.)

Definition A rule ρ^i (a map from histories to mixed actions) is ε*-universally consistent* (ε -UC) if for any opponents rule ρ^{-i}

$$\limsup_{T\to\infty} \max_{\sigma^i} u^i(\sigma^i, d_T^{-i}) - \frac{1}{T}\sum_t u^i(\rho_t(h_{t-1})) \leq \varepsilon$$

almost surely w.r.t. the distribution over outcomes induced by (ρ^i, ρ^{-i}).

(Here d_T^{-i} is the empirical distribution of opponent's play at *T*.)

Doesn't require that players detect cycles, just to get as good a payoff as if they knew the average frequencies. (more demanding consistency notions check performance on more measures, e.g. can also check play in even and odd periods..)

Any Bayesian *expects* to be both safe and consistent.
(*economists typically assume agents are Bayesian*)

Universal consistency asks for consistency against all alternatives. (even those with prior probability 0).

This requires randomization (consider matching pennies!)

Smooth ficitious play is universally consistent (Fudenberg-Levine *JEDC* [1995]).

Hart-MasCollel *JET* [2001] characterize a larger family of no-regret learning rules.
Roughgarden [2009] extends POA analysis to general no-regret dynamics.

Recency bias:

- Large psychology literature on this, see e.g. Erev and Haruvy *Handbook oF Experimental Economics vol. 2.*

- Benaim Hofbauer Hopkins JET [2009] analyze smooth fictitious play when agents have vanishingly little recency bias. (For each level of recency their model has an ergodic distribution, they characterize the limit of the distributions as the receny effect vanishes.)

- Fudenberg-Levine PNAS [2014] extend universal consistency to recency; the result is only interesting when recency bias is small.

- Fudenberg-Peysakhovich EC [2014] document (large) recency bias in a "lemons" problem: even after 20 observations people reacted "a lot" to the next one.

- Recency can be greatly diminished by providing summary statistics of past play. Need to better understand what sort of feedback encourages and discourages recency.

- Relatedly: many subjects make systematic computation errors, especially in computing conditional probabilities; this can let non-Nash play persist.

- The prevalence of these mistakes depends in part on the sort of feedback and computation aids (e.g. calculators) provided. More research needed here too.

- It would be nice to be able to say something about the implications of non-trivial recency for learning in games. But hard to develop formal results when there is lots of receny bias as the system doesn't settle down; may need to use simulations. (*and economists need to work out how to make better use of simulations)*

That's all for now ☺

Tomorrow: Learning in Extensive-Form Games

Reinforcement Learning (REL)

Two steps to defing a REL process:

a) what is reinforced? Actions, strategies, rules?

b) how?

Typically in REL the way 1 updates depends only on his realized payoff; he doesn't think about "regret", which is the payoff he would have received if he had played differently.

Simple(st?) version :
Cumulative Proportional Reinforcement (CPR)

- Normalize all utilities positive, and give initial weights to each action *k*.
- Update the score of the action that was played by its realized payoff.
- Do not update other scores.
- Probability of action *k* is weight on *k* divided by sum of weights.

- Response to “my opponent played R” can depend on player’s own action.

- Rational players who know the structure of the game shouldn’t condition on own action.

- Some lab subjects seem to do so- but the extent to which this happens is controversial. (Hard to estimate individual-specific learning models and fitting aggregate play with a model of a single agent leads to heterogeneity bias (Wilcox *Ema* [2006]).

- REL agents don't respond at all to “hypothetical reinforcement”- what they could have gotten by playing something else- and it seems that they do.

- Subjects play differently if told the play of others and not just own payoffs- goes against standard REL models.

- Camerer-Ho *Ema* [1999] nests REL and fictitious play as special cases by adding a parameter; best fit (for a “representative agent”) is “in the middle.”

- FO and REL also fit the data reasonably well in cases where an agent using it would do a reasonably good job of optimizing, and these are cases where play is not changing very quickly over the course of the experiment.

- But in games where play has a strong trend (like "beauty contest game" : guess 2/3 the average) none of the models CH consider do well- because people do eventually notice the time trend.

Extra: ***Dirichlet Priors and Multinomial Sampling***

Taken from DeGroot [1970] *Optimal Statistical Decisions*.

1) *The Multinomial Distribution:* Consider a sequence of *n* i.i.d. trials, where each period one of *k* outcomes occurs, with p_z denoting the probability of outcome *z*. Denote the outcome of the *n* trials by the vector κ, where κ_z is the number of the outcomes of type *z*. Then the distribution of the outcome is

$$f(\kappa) = \frac{n!}{\kappa_1 ! \cdot \cdot \kappa_k !} {p_1}^{\kappa_1} \cdots {p_k}^{\kappa_k}$$

for κ such that $\sum_{z=1}^{k} \kappa_z = n$.

2) *The Dirichlet Distribution:* Let Γ denote the gamma function. A random vector *p* has the Dirichlet distribution with parameter vector α if its density is given by

$$f(p) = \frac{\Gamma(\alpha_1 + ... + \alpha_k)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)} {p_1}^{\alpha_1 - 1} \cdots {p_k}^{a_k - 1}$$

for all *p*>0 such that $\sum_{z=1}^{k} p_z = 1.$

(the gamma function generalizes the factorial to the reals. Its role here is just as the "integrating constant": for *f* to be a density it has to integrate to 1.)

Fact: If *p* has the Dirichlet distribution, then $\int p_z f(p)dp = \alpha_z \,/\, \sum_{w=1}^{k} \alpha_w$.

So weights α correspond to relative probability of each outcome.

Two densities with the same expected values correspond to different ways the agent will update beliefs.

Fact: The Dirichlet distributions are a conjugate family for multinomial Sampling: if data is κ and prior is prior is Dirichlet with parameter α then posterior is Dirichlet with parameter $\alpha + \kappa$.

To see why consider posterior over p:

$$f(p \mid \kappa) = \frac{f(\kappa \mid p) f(p)}{\int f(\kappa \mid p) f(p) dp} \propto \prod_{z=1}^{k} p_z^{\alpha_z - 1} \prod_{z=1}^{k} p_z^{\kappa_z} = \prod_{z=1}^{k} p_z^{\alpha_z + \kappa_z - 1},$$

If player *i*'s date-t beliefs about player -*i's* mixed strategy have a Dirichlet distribution, player *i's* assessment of the probability that -*i* plays s^{-i} in period *t* is

$$\gamma_t^i(s^{-i}) = \int_{\Sigma^{-i}} \sigma^{-i}(s^{-i})\mu_t^i[\sigma^{-i}]d\sigma^{-i} = \alpha_z \,/\, \sum_{w=1}^{k} \alpha_w ,$$

This is the expected value of the component of σ^{-i} corresponding to s^{-i}.

So after observing sample κ, player *i*'s assessment of probability that the next observation is strategy *z* is

$$\frac{\alpha_z'}{\sum_{w=1}^{k} \alpha_w'} = \frac{\alpha_z + \kappa_z}{\sum_{w=1}^{k} (\alpha_w + \kappa_w)},$$ which is the formula for fictitious play.