Learning in the Worst Case

David K. Levine

September 1, 2005

Global Convergence?

- “grail” of learning research: global convergence theorem for convincing learning processes
- easy to construct examples of learning processes that don’t converge
- non-convergence looks like cob-web; people repeat the same mistakes over and over; not terrifically plausible
- we seem to see much “equilibriumness” around us (traffic example)

- possible and difficult to construct learning processes with global convergence properties (more or less must be stochastic) to Nash equilibrium; but the processes don’t make much sense (fishing for Nash equilibrium)

- I’ll try to convince you that “all sensible” learning procedures lead in the long-run to correlated equilibrium
- I’ll start by motivating learning processes from an individual perspective (i.e. processes that “work”)
- I’m only going to talk about pure forecasting (no causality)

Worst-case or Universal analysis vs. Bayesian analysis

- opponents may be smarter than you
- their process of optimization may result in play not in the support of your prior
- probability 1 with respect to your own beliefs is not meaningful in the setting of a game
- example: everyone believing that they face a stationary process (a common statistical assumption) implies that no one will actually behave in a stationary way
- these deficiencies in the robustness of Bayes learning are why there is no satisfactory global convergence theorem for learning procedures

“Classical” Case of Fictitious Play

- keep track of frequencies of opponents’ play
- begin with an initial or prior sample
- play a best-response to historical frequencies
- not well defined if there are ties, but for generic payoff/prior there will be no ties
- optimal procedure against i.i.d. opponents
- how well does fictitious play do if the i.i.d. assumption is wrong?

How well can fictitious play do in the long-run?

- notice that fictitious play only keeps track of frequencies: can fictitious play do as well in the long-run as if those frequencies (but not the order of the sample) was known in advance?
- alternatively: suppose that a player is constrained to play the same action in every period, so that he does not care about the order of observations

Universal Consistency

let u_t^i be actual utility at time *t*

let ϕ_t^{-i} be frequency of opponents' play (joint distribution over S^{-i})

suppose that for *all* (note that this does not say "for almost all") sequences of opponent play

$$\liminf_{T\to\infty}(1/T)\sum_{t=1}^{T}u_t^i - \max_{s^i} u^i(s^i,\phi_T^{-i}) \geq 0$$

then the learning procedure is universally consistent

Is fictitious play universally consistent? Fudenberg and Kreps example

0,0	1,1
1,1	0,0

this coordination game is played by two identical players

suppose they use *identical deterministic* learning procedures

then they play UL or DR and get 0 in every period

this is not individually rational, let alone universally consistent

Theorem [Monderer, Samet, Sela; Fudenberg, Levine]: fictitious play is consistent provided the frequency with which the player switches strategies goes to zero

Smooth Fictitious Play

instead of maximizing $u^i(s^i, \phi^i_{t-1})$ maximize

$$u^i(\sigma^i, \phi^i_{t-1}) + \lambda v^i(\sigma^i)$$

where v^i is smooth, concave and has derivatives that are unbounded at the boundary of the unit simplex

example: the *entropy*

$$v^i(\sigma^i) = -\sum_{s^i} \sigma^i(s^i) \log \sigma^i(s^i)$$

as $\lambda \to 0$ this results in an approximate optimum to the original problem

however the solution to $u^i(\sigma^i, \phi^i_{t-1}) + \lambda v^i(\sigma^i)$ is smooth and interior (always puts positive weight on all pure strategies)

Theorem [Blackwell, Hannan, Fudenberg and Levine and others]: smooth fictitious play is ε universally consistent with $\varepsilon \to 0$ as $\lambda \to 0$

Conditional Probability Models: Experts

allow time dependent games

$$\liminf_{T\to\infty}(1/T)\sum_{t=1}^{T}u_t^i-(1/T)\sum_{t=1}^{T}\max_{s^i}u_t^i(s^i,s_t^{-i})\geq 0$$

same theorem holds, without change in proof

a “model” makes conditional probability forecasts

an “expert” makes recommendations about how to play

$$s_t^i = e^i(h_{t-1}^i)$$

set $v_t^i(e^i,s_t^{-i})=u^i(e^i(h_{t-1}^i),s_t^{-i})$

conclusion: can do as well as if you knew who the best expert was in advance

Conditional Probability Models: Direct

classify observations into subsamples

countable collection of categories Ψ

classification rule $\psi^i : H \times S \to \Psi$

$\psi^i(h_{t-1}^i, s_t^i)$

$\phi_t^{-i}(\psi)$ empirical distribution of opponent's play conditional on the category ψ; $n_t(\psi)$ is number of time category has occured

effective categories: minimal finite subset $\Psi_t \in \Psi$ with all observations through time *t*

m_t denotes the number of effective categories

Assumption 1: $\lim_{t \to \infty} m_t / t = 0$

This is essentially the method of sieves

Universal Conditional Consistency

total utility actually received in the subsample ψ is $u_t^i(\psi)$

$$c_t^i(\psi) = \begin{cases} n_t(\psi) \max_{s^i} u^i(s^i, \phi_t^{-i}) - u_t^i(\psi) & n_t(\psi) > 0 \\ 0 & n_t(\psi) = 0 \end{cases}$$

universal conditional consistency

$$\lim \sup (1/T) \sum_{\psi \in \Psi_t} c_T^i(\psi) \le 0$$

Non Calibrated Case

categorization rule depends only on history, not on own plans

1) given h_{t-1}^{i}, $\psi(h_{t-1}^{i})$ chooses the category

2) play a smooth fictitious play against the sample in the chosen category $\phi_{t-1}^{-i}(\psi)$

3) add the new observation s_{t}^{-i} to the category $\psi(h_{t-1}^{i})$

Works like smooth fictitious play within each category, so universally conditionally consistent

Calibrated Case

try to use a rule $\psi(h_{t-1}^i, s_t^i)$

focus on special case $\psi(s_t^i), \Psi = S$

each category ψ has a corresponding smooth fictitious play $\sigma^i(\phi_{t-1}^{-i}(\psi))$

suppose we choose category ψ with probability $\lambda(\psi)$, then overall play is

$$pr(s^i) = \sum_{\psi} \lambda(\psi)\sigma^i(\phi_{t-1}^{-i}(\psi))[s^i]$$

but categories correspond to own strategies: fixed point property:
$\lambda(s^i) = pr(s^i)$

$$\lambda(s^i) = \sum_{\psi} \lambda(\psi)\sigma^i(\phi_{t-1}^{-i}(\psi))[s^i]$$

unique fixed point, solvable by linear algebra

Interpretation of Calibration

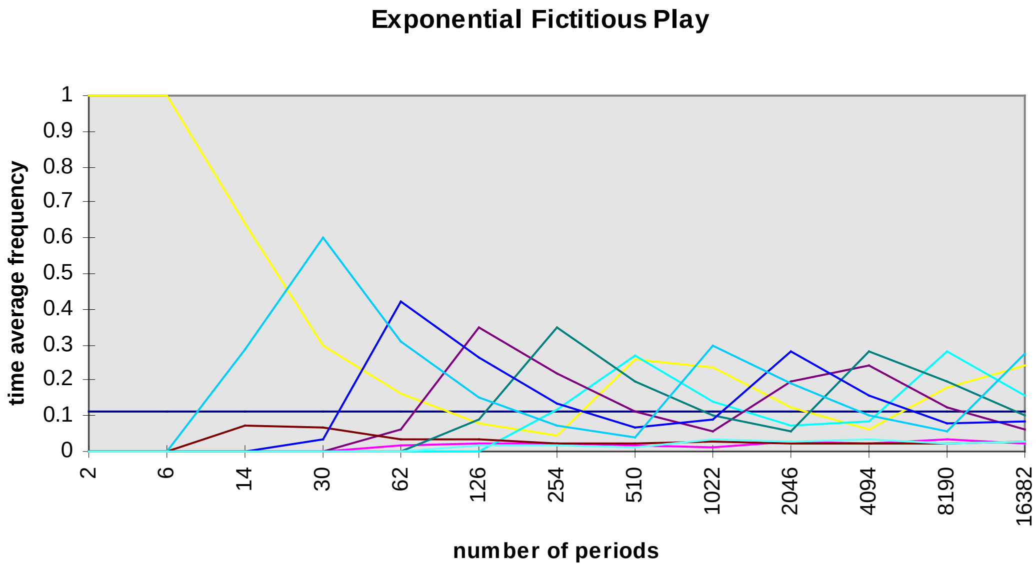
weather forecasting example: calibrated beliefs, versus calibrated actions

consequence of universal calibration: global convergence to the set of correlated equilibria

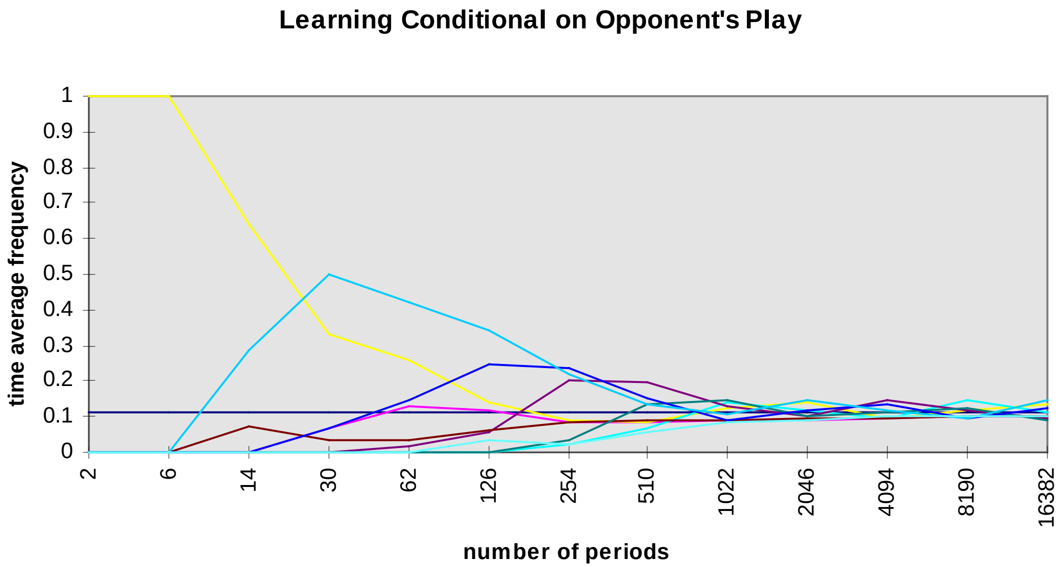
Shapley Example

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

smooth fictitious play (time in logs)



condition on opponents last period play (time in logs)


Learning Conditional on Opponent's Play
time average frequency
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
2
6
14
30
62
126
254
510
1022
2046
4094
8190
16382
number of periods

Discounted Learning

A learning procedure $\hat{\rho}$ is ε*-as good as* a procedure ρ if for all sequences of discount factors $\{\beta_t\}$ and all histories h_t^i

$$\sum_{t=1}^{\infty} \beta_t u(\rho(h_{t-1}), s_t^{-i}) \leq \sum_{t=1}^{\infty} \beta_t u(\hat{\rho}(h_{t-1}), s_t^{-i}) + \varepsilon$$

Proposition 2: For any learning procedure ρ and any ε there exists a categorical smooth fictitious play $\hat{\rho}$ that is ε-as good as ρ

exploits the fact that the time average result must be true for at every time

Questions

- synchronicity and asynchronicity of play and the consequences for convergence.
- what constitute good categorization schemes (pattern recognition)
- how can data be pooled across “similar” categories?
- dynamic programming/ state variables
- inference of causality
- procedures in large strategy spaces (genetic algorithms, for example)
- empirical analysis of these learning rules vs. others such as stimulus-response.
- use of payoff irrelevant information, such as observations about the experience of other players.
- averaging versus distributed lag