Learning and Randomness

Why Not Be A Bayesian?

Fudenberg/Kreps example

	H	T
H	0,0	1,1
T	1,1	0,0

You know your own payoffs and are playing against an unknown opponent

Suppose your "model" of the opponent is i.i.d. play

What Does A Bayesian Do?

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 including “prior” sample
- not well defined if there are ties, but for generic payoff/prior there will be no ties

What Do Identical Bayesian Do?

suppose prior is $\sqrt{2}/2, 0$ ($\sqrt{2}/2$ is about 0.7) observations of H and T

irrational guarantees no ties

period 1: both play T

new sample: $\sqrt{2}/2, 1$

period 2: both play H

new sample: $1 + \sqrt{2}/2, 1$

period 3: both play T

…

new sample $t/2 + \sqrt{2}/2, (t+1)/2$ rounding down

even period: play H, odd period play T

What Do Identical Bayesians Get?

- zero in every period – as bad as possible
- worse then the minmax that can be guaranteed by randomizing 50-50
- worse than that – any deterministic procedure Bayesian or not yields the same result when both players are identical

Fictitious Play In The Long Run

- notice that fictitious play only keeps track of frequencies: cannot be expected to do better in the long run then if those frequencies (but not the order of the sample) was known in advance

- Universal (or Hannan) Consistency

let u_t^i be actual utility at time *t,* let ϕ_t^{-i} be frequency of opponents' play

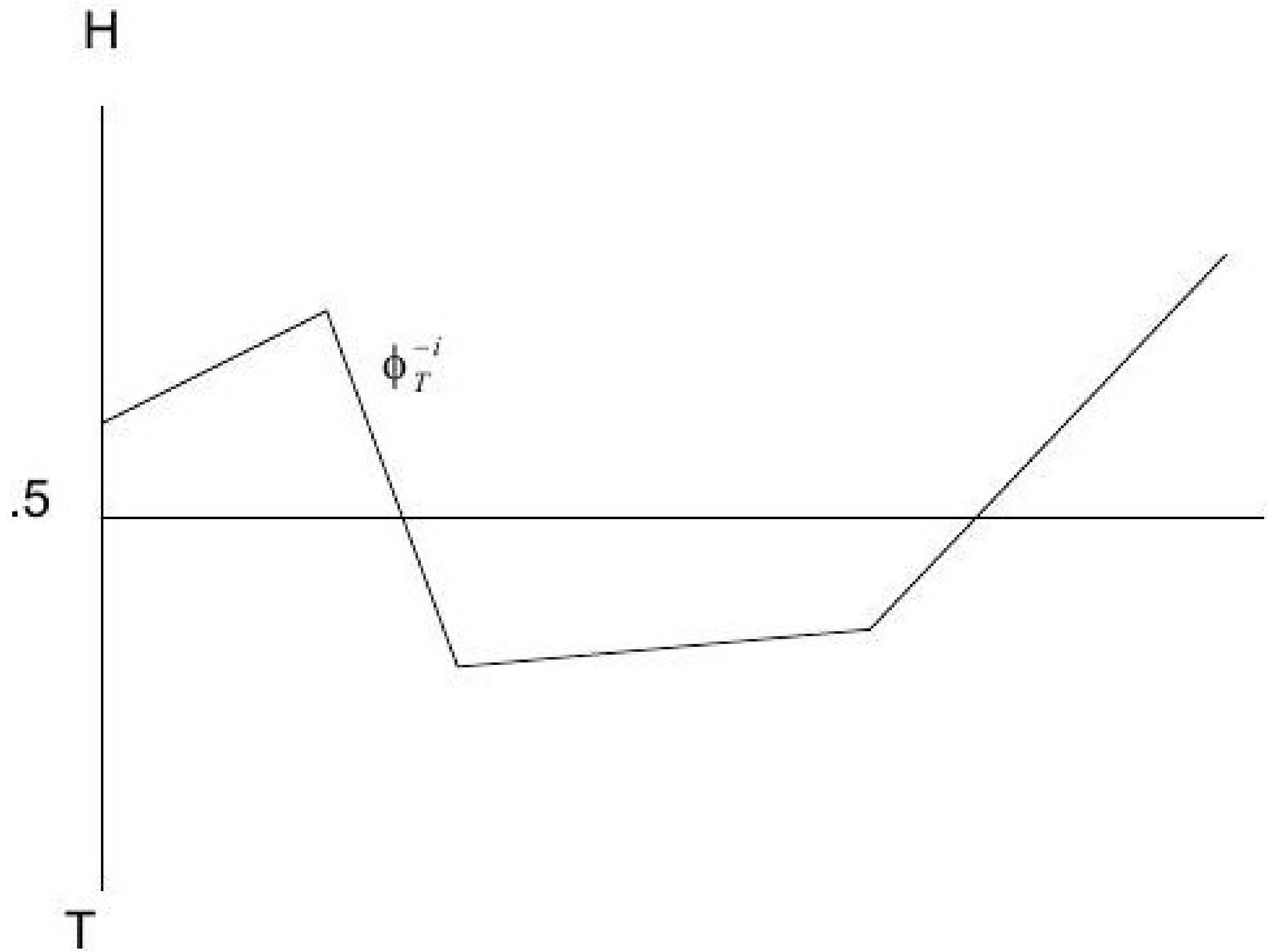
universal consistency: for *all* (note that this does not say "for almost all") sequences of opponent play

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

remark on terminology: $u^i(s^i,\phi_T^{-i}) - (1/T)\sum_{t=1}^{T}u_t^i$ is the Hannan regret for strategy s^i

Non-Universal Consistency

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


H
ϕ_T^{-i}
.5
T

Randomize?

Why not randomize when near indifferent?

Smooth fictitious play: instead of maximizing $u^i(s^i, \phi^i_{t-1})$ maximize $u^i(s^i, \phi^i_{t-1}) + \lambda v^i(\sigma^i)$

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

example: the *entropy* $v^i(\sigma^i) = -\sum_{s^i} \sigma^i \log \sigma^i(s^i)$

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

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

Existence of Universally Consistent Learning Rules

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

Calibration

Notice that pattern recognition is ruled out

Instead, use conditional probabilities; specifically

$\phi_T^{-i}(\widetilde{s}^i)$ sample just when you played that strategy

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

called calibration

Interpretation of Calibration

weather forecasting example: calibrated beliefs, versus calibrated actions

- Foster and Vohra – existence of universally calibrated algorithms
- Fudenberg and Levine – by bootstrapping universally consistent algorithms
- key consequence of universal calibration: global convergence to set of correlated equilibria

How Do You Do It?

$\hat{\sigma}^i(\phi)$ smooth fictitious play or something else universally consistent

suppose you play $\widetilde{\sigma}^i$; with probability $\widetilde{\sigma}^i(s^i)$ you play s^i

if you choose s^i then you "should" play $\hat{\sigma}^i(\phi_{t-1}^{-i}(s^i))$

so overall, you "should" play $\sum_{s^i}\widetilde{\sigma}^i(s^i)\widehat{\sigma}^i(\phi_{t-1}^{-i}(s^i))$

but what you should play depends on what you do!

a fixed point problem: $\widetilde{\sigma}^i = \sum_{s^i}\widetilde{\sigma}^i(s^i)\widehat{\sigma}^i(\phi_{t-1}^{-i}(s^i))$

easy to solve, and indeed the solution is indeed calibrated

Categorization Schemes

classify observations into subsamples

countable collection of categories Ψ

classification rule $\psi^i(h^i_{t-1}, s^i_t) \in \Psi$

$\phi_t^{-i}(\psi)$ empirical distribution of opponent's play conditional on ψ

effective categories: minimal finite subset of Ψ constaining all observations through time t

m_t denotes the number of effective categories

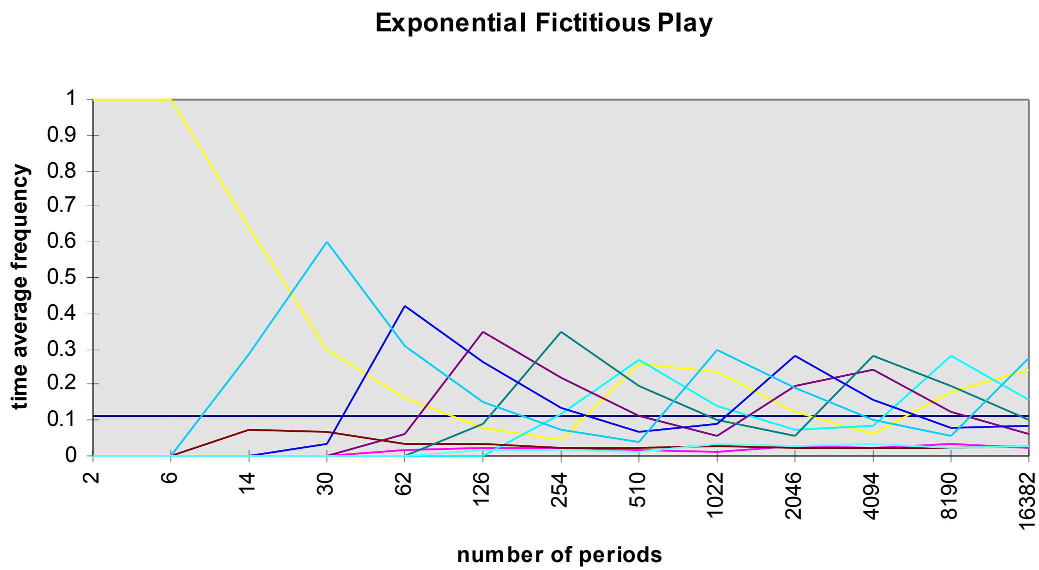
need $m_t/t \to 0$

method of sieves

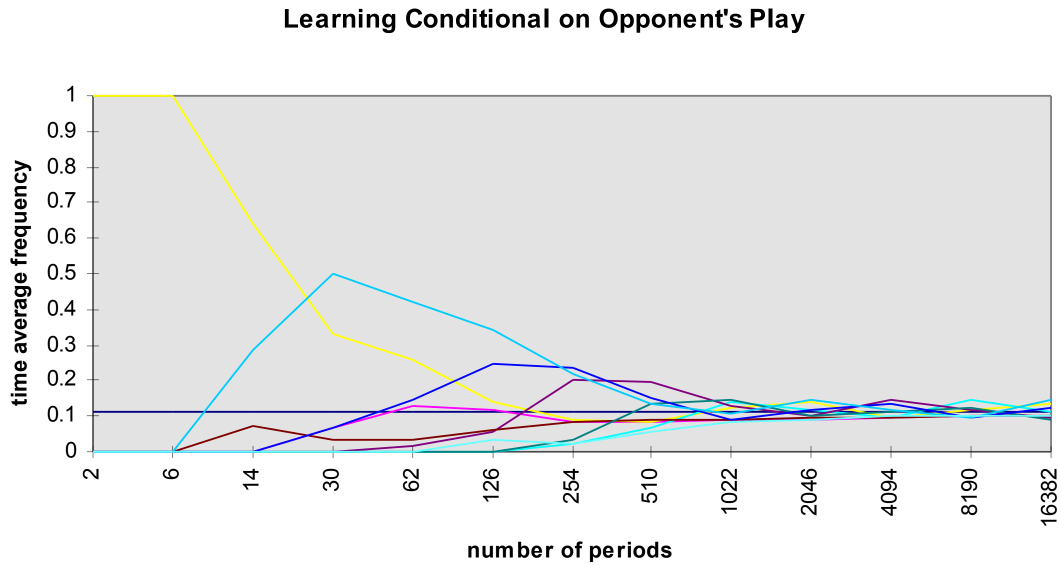
Shapley Example

	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

Limits of Calibration: Jordan Example

- three player matching pennies, where 1 wants to match 2 wants to match 3 wants not to match 1
- consider HHH ->HHT->HTT->TTT->TTH->THH->
- also consider that equal probability over these six outcomes is a correlated equilibrium
- take player 1: told to play H, then he faces 2H,T so it is strictly best to play H
- not that this makes much sense

Questions

- being Bayesian generically?
- synchronicity and asynchronicity of play and 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?)

Learning With Recency

- empirically people place more weight on current observations: Cheung and Friedman (1997) , Argawal et al (2008), Erev and Haruvy (2013)
- two models - weighted observation, limited memory Cheung and Friedman (1997), Sutton and Barto (1998), Camerer and Ho (1999), Benaim, Hofbauer and Hopkins (2009), Young (1993)

The Learning Model

periods $t = 1, 2, \ldots$ t

finite actions $a \in A$

finite outcomes $y \in Y$, mixtures γ

utility $u(a, y)$, mixtures α

strategies σ depend on histories $h_t = (a_1, y_1, \ldots, a_t, y_t)$ with initial null history h_0 the null history

conditional probability of a_t, y_t is $\sigma(h_{t-1})[a_t]\rho(h_{t-1})[y_t]$

Belief Based Strategies

A Markov belief based strategy

a prior belief $\phi_0 \in \Delta(Y)$

a Markov learning kernel $P(\phi|\phi_{t-1}, y_t)$

a response map $\alpha(\phi_{t-1})$ (for example a selection from the best-response)

$f_\tau(y|h_t) = f(y|y_\tau)$, the indicator function for whether the period-τ outcome is y

Recursive Weighting

a weight $0 < \mu < 1$

deterministic kernel $\phi(y|h_t) = \mu f_t(y|h_t) + (1-\mu)\phi(y|h_{t-1})$

Weighted Sampling

a weight $\lambda > 1$

beliefs

$$\phi(y|h_t) = \frac{\sum_{\tau=1}^{t} \lambda^\tau f_\tau(y|h_t) + \sum_{\tau=-\infty}^{0} \lambda^\tau \phi_0(y)}{\sum_{\tau=-\infty}^{t} \lambda^\tau}$$

equivalent to recursive formulation with

$$\mu = \frac{\lambda^t}{\sum_{\tau=-\infty}^{t} \lambda^\tau} = 1 - \lambda^{-1}$$

Limited Memory

memory has size M

a k, p, M procedure where $0 < p \leq 1, 1 \leq k \leq M$ proceeds as follows:

1. Choose randomly a subset of M of size k

2. Discard each observation in the subset independently and randomly with probability p

3. Replace all the discarded observations with the observation from the current period.

The simplest version has $k = 1, p = 1$ - choose one observation at random from memory and discard it. In this case when the signal y is i.i.d., the ergodic distribution is multinomial

Relation to Weighted Sampling

k, p procedure allows us to separate memory size M from λ while allowing the construction of procedures with arbitrary values of λ.

the probability an observation is thrown out of the sample is pk/M so the corresponding value of λ is M/pk

Recursive Weighting versus Limited Memory

initialize two systems so that the distribution of observations in the limited memory is the same as the prior ϕ_0

fix any sequence of observations y_t

consider the deterministic sequence ϕ_t from recursive weighting and the random process $\tilde{\phi}_t$ from limited memory

Theorem: *For any fixed* $\mu \in (0, 1)$, *as* $M \to \infty$ *then* $E[|\tilde{\phi}_t - \phi_t|] \to 0$ *uniformly in* t *and the sequence of observations* $(y_1, y_2, \ldots)$*.*

Approximate Universal Consistency of Slightly Weighted Sampling

let ϕ_t denote beliefs of the weighted sampling scheme

let γ_t denote the weighted beliefs through and including observations at time t excluding the prior

fix a scale parameter $\overline{U} > 0$, let $0 < \zeta \leq 1$ be a "smoothing" parameter

let ν be a "smoothing" function that maps the interior of the simplex to the reals, is bounded by $\overline{U}$, is smooth, strictly differentiably concave and satisfies the boundary condition that as γ approaches the boundary of the simplex the norm of the derivative becomes infinite. (For example, entropy.)

for any probability distribution γ define $v(\alpha, \gamma) = u(\alpha, \gamma) + \zeta\nu(\alpha)$

Properties of Smoothed Utility

the smoothed best response is $\hat{\alpha}(\gamma) = \arg\max_\alpha v(\alpha, \gamma)$.

$v(\hat{\alpha}(\gamma), \gamma)$ is Lipschitz with constant of the form $B\overline{U}/\zeta$ where B depends only on ν.

as $\zeta \to 0$ the smooth best response approaches (pointwise) the best response

Weighted Universal Consistency

$u_t = \sum_{\tau=1}^{t} \lambda^\tau u(\alpha(h_\tau), f_\tau)$ total weighted expected utility received through period t where f_t is the distribution that places weight one on y_t

$U(\gamma) = \max_\alpha u(\alpha, \gamma)$

$\Lambda_t = \sum_{\tau=1}^{t} \lambda^\tau$

$c_t = \Lambda_t U(\phi_t) - u_t,\ c_0 = 0,$

weighted universal consistency is $\limsup_{t\to\infty} c_t/\Lambda_t \le \epsilon$

Smooth Recursive Learning Universally Consistent

suppose the agent at each date sets $\alpha(h_t) = \hat{\alpha}(\phi_t)$.

Theorem: *For any ν there exists a constant $B > 0$ such that for all utility functions $|u(a, y)| \leq \overline{U}$ the recursive memory model with parameters μ, ζ satisfies $c_t/\Lambda_t \leq 7\overline{U}|1/(\mu\Lambda_t) + \zeta + B\mu/\zeta|$.*

A Game

define $y^i = a^{-i}$

choose a "monotone" ν^i such that $u(a^i, \gamma^i) \geq u(\tilde{a}^i, \gamma^i)$ implies $\hat{\alpha}^i(\gamma^i)[a^i] \geq \hat{\alpha}^i(\gamma^i)[\tilde{a}^i]$ (for example the entropy function)

The Weighted Procedure

Fix ϵ and set $\epsilon_1 = \epsilon/4$ and $\epsilon_2 = \epsilon/(4\overline{U})$ (and also smaller than 1/2).

Choose ζ sufficiently small that two properties hold

1. $7\overline{U}\zeta \leq \epsilon_1$.

2. if a^i is any $\psi\overline{U}$-strict best response then $\hat{\alpha}^i(\gamma^i)[a^i] \geq 1 - \epsilon_2$.

Next choose μ such that $7\overline{U}B\mu/\zeta \leq \epsilon_1$.

This procedure by the earlier theorem is $2\epsilon_1$ universally consistent

The Limited Memory Procedure

choose $k^i = M^i$, that is, we potentially discard all observations

choose M^i large enough that $\overline{U} E[|\tilde{\phi}_t^i - \phi_t^i|] \leq \epsilon_1$

then the procedure replacing ϕ_t^i with $\tilde{\phi}_t^i$ is $3\epsilon_1$ universally consistent

The Sticky Procedure

$\overline{\alpha}^i(h_t^i)$

1. (the stuck state)

if all the observations in the memory are identical and $\hat{\alpha}^i(\gamma^i)[a^i] \geq 1 - \epsilon_2$ then $\overline{\alpha}^i(\gamma^i)[a^i] = 1$

2. otherwise, $\overline{\alpha}^i(\gamma^i)[a^i] = \hat{\alpha}^i(\gamma^i)[a^i]$.

This procedure is $3\epsilon_1 + \overline{U}\epsilon_2 = \epsilon$ universally consistent.

A Convergence Theorem

a simultaneous move game with observable actions and payoffs bounded by $\overline{U}$

Theorem: *For any $\epsilon, \psi, \overline{U}$ there exist recursive-memory learning procedures that are ϵ-universally consistent with respect to the payoff bound $\overline{U}$ for each player such if the game has a $\psi\overline{U}$-strict Nash equilibrium then with probability one the learning procedures converge to some strict Nash equilibrium*

Noise on the Equilibrium Path

key fact is absence of noise on the equilibrium path not strictness

universally consistent procedures that converge to epsilon-Nash with observable mixed strategies (including Nature)

cannot have probability one convergence with noise on equilibrium path and universal consistency

would have to stop being responsive despite noise (for example, Hart-Mas-Colell)

hence nasty opponent could get you stuck then do something bad forever

at best show as did Foster and Young high probability of Nash in ergodic distribution