Recency, Consistent Learning, and Nash Equilibrium

Learning with Recency Bias☆

Drew Fudenberg[1], David K. Levine[2]

Abstract

We examine the long-run implication of two models of learning with recency bias: recursive weights and limited memory. We show that both models generate similar beliefs, and that both have a weighted universal consistency property. Using the limited memory model we are able to produce learning procedures that are both weighted universally consistent and converge with probability one to strict Nash equilibrium, the first example of which we are aware of learning procedures that have this convergence property and also have desirable properties for the individual agents who use them.

JEL Classification Numbers: 001,002

Keywords: Learning, Recency, Fictitious Play, Game Theory, Universal Consistency

1. Introduction

There is substantial evidence from both the laboratory and the field that people display "recency bias," meaning that they react more heavily to recent observations and experiences than they do to older ones. [3]The magnitude of recency bias varies with the setting and the form of feedback; in particular, some forms of summary statistics can make this bias very small. Recency has been incorporated into both belief-based and reinforcement-based models of learning, by adding a parameter that controls the speed of informational discounting (see, for example, Cheung and Friedman (1997), Sutton and Barto (1998), Camerer and Ho (1999), Benaim, Hofbauer and Hopkins (2009)), or by supposing that individuals retain only a finite sample in their memory (Young (1993)).

Here we investigate models of learning with recency. We consider both informational discounting and finite memory models, and show that the beliefs in the two models are roughly the same if the finite memory is large enough. Moreover in this case of little recency, both models satisfy a weighted universal consistency property of achieving about as much utility as would be possible as if the "informationally discounted" sample path were known in advance. This is a variation

☆This version: First Version: October 28, 2013. We are grateful to NSF Grants SES-08-51315 and 1258665 for financial support.

*Corresponding author: David K. Levine

Email address: david@dklevine.com (David K. Levine)

[1] Deparment of Economic, Harvard University

[2] Department of Economics, European University Institute and Washington University in St. Louis

[3] Cheung and Friedman (1997) is probably the first empirical study of play in a game to estimate a "recency" parameter. See, for example, Argawal et al (2008) for evidence of recency in the field, and Erev and Haruvy (2013) for a recent survey of recency bias in single-agent decision experiments.

on a widely used non-Bayesian criterion for the success of a learning process based on the notion of worst-case performance. The weighted universal consistency property generalizes the notion of universal or Hannan consistency discussed, for example, in Fudenberg and Levine (1995). This requires that regardless of the sample path in the long run the player does as well as if he knew the time average of the sample ahead of time. For procedures that keep track only of averages (or weighted averages) this is a sensible criterion for "success." We also show that there are universally consistent finite-memory learning procedures that converge with probability one to strict Nash equilibrium. This is the first result of this type of which we are aware: earlier results such as Foster and Young (2006) and Hart and Mas-Colell (2013) give convergence results, but do not show that the procedures have universal consistency properties. These results complement those of Benaim, Hofbauer and Hopkins (2009), who related the long-run behavior of stochastic fictitious play with very little informational discounting to the smooth best response dynamic that describes the asymptotic behavior of smooth fictitious play without recency.

2. The Model

We consider a one person decision problem. Each period $t = 1, 2, \ldots$the player chooses an action a from a finite set of actions A, then observes an outcome $y \in Y$, a finite set. The utility from action a when the outcome is y is denoted by $u(a, y)$. The space of probability distributions over a (finite) set S is denoted by $\Delta(S)$. Mixed actions are denoted by $\alpha \in \Delta(A)$, and mixed outcomes by $\gamma \in \Delta(Y)$. We will write $u(\alpha, \gamma)$ for the expected utility to mixed actions and mixed outcomes. A strategy for the player can depend only on the information available to him when he moves, namely the past values of his own play and the outcome. A history of play for the player is denoted by $h_t = (a_1, y_1, \ldots, a_t, y_t)$, with h_0 the null history and the space of all histories of play denoted by H. A (behavior) strategy for the player is a map $\sigma : H \to \Delta(A)$, while an outcome function is a map $\rho : H \to \Delta(Y)$. Each strategy-outcome function pair induces a stochastic process over action/outcome pairs, where given the history h_{t-1} the conditional probability of a_t, y_t, is $\sigma(h_{t-1})[a_t]\rho(h_{t-1})[y_t]$. In other words, the player and nature must base their play only on the history of actions and outcomes. In some interpretations, the outcomes may be chosen by other players rather than by nature.

3. Notions of Recency

3.1. Belief Based Strategies

A Markov belief based strategy consists of a prior belief $\phi_0 \in \Delta(Y)$, a Markov kernel $P(\phi|\phi_{t-1}, y_t)$ that specifies how beliefs are updated, and a map $\alpha(\phi_{t-1})$ from beliefs at time $t - 1$ to a mixed action at time t. One such map is the best-response map;[4] we will also consider various smooth approximations to the best-response map. For the moment we focus on modeling the evolution of

[4] Although this is not single-valued, we can make an arbitrary choice in case of indifference.

beliefs. In doing so, it will be convenient to define $f(y|y_t)$ to be equal to 1 if $y = y_t$ and 0 otherwise, and $f_\tau(y|h_t) = f(y|y_\tau)$, the indicator function for whether the period-τ outcome was equal to y.

3.2. Recursive Weighting

We are given a weight $0 < \mu < 1$ an initial condition ϕ_0 and the deterministic kernel $\phi(y|h_t) = \mu f_t(y|h_t) + (1 - \mu)\phi(y|h_{t-1})$ where $\phi(y|h_0) \equiv \phi_0(y)$. We now show how this Markov belief process is equivalent to several others.

3.3. Weighted Sampling

We define the weighted sample with respect to weight $\lambda > 1$ by giving the tth observation weight λ^t ; this is equivalent to always weighting the most recent (time t) observation by 1 and discounting observations at times $t - \tau$ by $\omega = 1/\lambda < 1$.

To make this stationary and incorporate the effect of the prior, we use weights stretching back to $-\infty$. Specifically, we define

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

In the latter form, it is clear that the relationship is recursive, and if we define

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

we can write this in the recursive weighted form at $\phi(y|h_t) = \mu f_t(y|h_t) + (1 - \mu)\phi(y|h_{t-1})$. Note that $\lambda \to \infty$ corresponds to the case $\mu \to 1$ where only the most recent observation matters, while Bayesian updating for a fixed unknown distribution of osbervations corresponds to a non-stationary model in which $\lambda = 1 \text{and} \mu_t \to 0$.

3.4. Base Rate Neglect

In the case where the agent believes that ρ is generated by iid sampling, let $\pi_t \in \Delta(\Delta(Y))$ be beliefs over $\Delta(y)$. Benjamin, Bodoh-Creed and Rabin (2013) propose a model of base rate neglect with the updating rule

$$\pi_t(\phi) = \frac{\phi(y_t)[\pi_{t-1}(\phi)]^\nu}{\int \phi(y_t)[\pi_{t-1}(\phi)]^\nu d\phi}.$$

Notice that if $\nu = 1$ this is ordinary Bayesian updating. If the prior is Dirichlet, the posterior mean is simply the maximum of this function with respect to ϕ. For $\nu \neq 1$ the posterior mean is difficult to compute, so we continue to measure central tendency by taking the maximum of the function, that is, the posterior mode rather than mean. To consider the maximum of this function with respect to ϕ, we can ignore the denominator and maximize the logarithm $\log \phi(y_t) + \nu \log[\pi_{t-1}(\phi)] = \sum_{\tau=1}^{t} \nu^{t-\tau} \log \phi(y_\tau) + \nu^t \log \pi_0(\phi)$. If we assume that the prior is such that $\log \pi_0(\phi) = \sum_{\tau=-\infty}^{0} \nu^{-\tau} \log \phi(y_\tau)$ for some weighted fictitious prior sample $y_\tau|_{\tau=-\infty}^{0}$ then we may write the log-likelihood with prior as $\log \phi(y_t) + \nu \log[\pi_{t-1}(\phi)] = \sum_{\tau=-\infty}^{t} \nu^{t-\tau} \log \phi(y_\tau)$ and the maximum likelihood is simply given by the weighted sample averages

$$\hat{\phi}(y|h_t) = \frac{\sum_{\tau=-\infty}^{t} \nu^{-\tau} f_\tau(y|h_t)}{\sum_{\tau=-\infty}^{t} \nu^{-\tau}}$$

so that if we take $\lambda = \nu^{-1}$ and define $\phi_0(y) = \hat{\phi}(y|h_0)$this is exactly the same point belief as generated by recency generated by weighted sampling.

3.5. Limited Memory

So far we have supposed in effect that there is unlimited memory for past observations, or at least that any value of ϕ can be recorded in the memory. We now instead suppose that the memory has size M where M refers to the number of observations that can be stored[5] Our goals are to show that when M is large that the agent receives about the same utility as he would with infinite memory, and also to examine the stationary distribution of play when the data is generated iid.

Then 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 of this procedure has $k = 1, p = 1$, that is, 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.

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

Except in our existing case where it is multinomial, the stationary distribution of this procedure while not intrinsically complicated (as in the case with deterministic weighting) does not seem especially easy to compute.

In the k, p procedure the probability an observation is thrown out of the sample is pk/M. The corresponding value of λ should be M/pk.

4. Recursive Weighting versus Limited Memory

The recursive memory weighting model has a deterministic transition kernel and an infinite state space. The limited memory model has random transitions and a finite state space. The latter has some advantages in analyzing properties such as universal consistency. In the case where ρ generates iid values of y the stationary distribution of the recursive weighting model can be extremely complicated and need not have a density even when y is binary. The stationary distribution of the limited memory model always exists and in some cases is quite simple: in the

[5] Note that this different than the limited-history processes considered by Young (1993), where the memory always contains the M most recent observations)

case in which $k = 1, p = 1$ each observation is drawn from exactly the same distribution ϕ^*, so the belief each period is simply a multinomial with M observations drawn from ϕ^*.[6]

We now want to relate the two models. Suppose that $\mu = pk/M$, so that the expected weight that the limited-memory model gives to the most recent observation is the same as in the recursive model, and initialize the two systems so that the distribution of observations in the limited memory is the same as the prior $\phi_0 \in \Psi_k$. Fix any sequence of observations y_t and consider the deterministic sequence ϕ_t from recursive weighting and consider the random process $\tilde{\phi}_t$ from limited memory. Then

Theorem 1. *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 $(y_1, y_2, \ldots)$.*

Proof. Fix the sequence f_t, define

$$z_t = \frac{\tilde{\phi}_t - (1-\mu)\tilde{\phi}_{t-1}}{\mu} - f_t$$

and observe that $\tilde{\phi}_t - \phi_t = \sum_{\tau=1}^{t} \lambda^t z_t / \Lambda_t$. Hence to prove the theorem it is sufficient to prove that $E[|z_t|] \to 0$.

Now let ℓ be the number of observations in updating to period t that are discarded and let $\tilde{\phi}^{\ell}_{t-1}$ be the frequencies in the remaining sample. Note that we can think of this as drawing $M - \ell$ observations from , $\tilde{\phi}_{t-1}$ without replacement, and we arbitrarily define $\tilde{\phi}^{M}_{t-1} = \tilde{\phi}_{t-1}$. Since $\mu < 1$ it is less than some $\overline{\mu} < 1$. Simple algebra shows that [7]

$$E[|z_t||\tilde{\phi}_{t-1}, \ell] \;\; \leq \;\; 2\left|\frac{\ell}{\mu M} - 1\right| + \frac{1-\mu}{\mu}\left(Pr(\ell > (\mu/\sqrt{\overline{\mu}})M) + \max_{\ell \leq (\mu/\sqrt{\overline{\mu}})M} E\left[\left|\tilde{\phi}_{t-1} - \tilde{\phi}^{\ell}_{t-1}\right| \left|\tilde{\phi}_{t-1}, \ell\right]\right) \quad (4.1)$$

where the last line follows from the facts that both $\tilde{\phi}^{\ell}_{t-1}$ f_t ,and the difference between $\tilde{\phi}^{\ell}_{t-1}$,and $\tilde{\phi}_{t-1}$ are all between 0 and 1.

Next we observe that $E|X| \leq 2\sqrt{E|X|^2}$ hence it is enough to prove that each of the expectations on the RHS of 4.1 has square deviation that goes to zero. Examining first $E\left|\frac{\ell}{\mu M} - 1\right|^2$,recall that $\mu = pk/M$ and $E\ell = pk$, so we need only compute the variance of $\frac{\ell}{\mu M}$. The variance of ℓ is $p(1-p)k = \mu M(1-p)$, so the variance of $\frac{\ell}{\mu M}$ is

$$\frac{\mu M(1-p)}{\mu^2 M^2} \leq \frac{1}{\mu M}$$

.

Turning to the second term $E\left[\left|\tilde{\phi}_{t-1} - \tilde{\phi}^{\ell}_{t-1}\right|^2 \left|\tilde{\phi}_{t-1}, \ell\right]\right.$ observe that $E[\tilde{\phi}^{\ell}_{t-1}|\tilde{\phi}_{t-1}, \ell] = \tilde{\phi}_{t-1}$ and that the variance is bounded above by sampling with replacement, which is at worst $1/(M - (\mu/\sqrt{\overline{\mu}})M) \leq 1/(M(1-\sqrt{\overline{\mu}}))$. Hence $\max_{\ell \leq L} E\left[\left|\tilde{\phi}_{t-1} - \tilde{\phi}^{\ell}_{t-1}\right| \left|\tilde{\phi}_{t-1}, \ell\right]\right. \leq 1/(M(1-\sqrt{\overline{\mu}})) \to 0$. □

[6] The same is true in the limited-history model of Young (1993), where the agent always forgets the oldest observation in the memory. However, that model does not fit our framework as it requires the state to keep track of the order in which the observations were acquired.

[7] The details of this computation can be found in the Web Appendix.

5. Approximate Universal Consistency of Slightly Weighted Sampling

We continue to let ϕ_t denote the beliefs of the weighted sampling scheme, and let γ_t denote the weighted beliefs through and including observations at time t excluding the prior .

Fix a scale parameter $\overline{U} > 0$.

For any probability distribution γ suppose for some $0 < \zeta \leq 1$ we define $v(\alpha, \gamma) = u(\alpha, \gamma) + \zeta\nu(\alpha)$ where ν maps the interior of the simplex to the reals, is bounded by $\overline{U}$, smooth, strictly differentiably concave and satisfies the boundary condition that at γ approaches the boundary of the simplex the norm of the derivative becomes infinite. The function $v(\hat{\alpha}(\gamma), \gamma)$ is Lipschitz (Fudenberg and Levine (1999)) and from the implicit function theorem the Lipschitz constant has the form $B\overline{U}/\zeta$ where B depends only on ν. This perturbation of the utility function serves to induce mixing, and allows the approximation of the best response function by the smoothed best response $\hat{\alpha}(\gamma) = \arg\max_\alpha v(\alpha, \gamma)$. Suppose suppose the agent at each date sets $\alpha(h_t) = \hat{\alpha}(\phi_t)$.

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

Past work (Fudenberg and Levine (1995)) has shown that when $\lambda = 1$ (no recency at all), the rule $\alpha(h_t) = \hat{\alpha}(\phi_t)$ is $\varepsilon-$universally consistent, meaning that regardless of the probability law of the y_t, $\limsup_t U(\phi_t) - u_t/t \leq \epsilon$ where $\epsilon > 0$ can be made arbitrarily small by taking the weight ζ on the non-linear term ν to be small.

To extend this to allow for recency (that is $\lambda > 1$) define $c_t = \Lambda_t U(\phi_t) - u_t$, $c_0 = 0$, where ϕ_t now varies with λ We will show that when $\zeta$and μ are small enough, $\limsup_{t\to\infty} c_t/\Lambda_t \leq \epsilon$ for all utility functions that satisfy the payoff bound $\overline{U}$, and we will conclude that the learning procedure is *ϵ-universally consistent with respect to $\overline{U}$.*

To see why this terminology makes sense, note that in the case $\lambda = 1$ this is the approximate universal consistency condition described above, as c_t/Λ_t reduces to $U(\phi_t) - u_t/t$, The condition $\lambda > 1$ is stronger than $\lambda = 1$ in that it places more weight on the next (unknown) observation than on past observations. Hence, conceptually the bigger is λ the stronger is the notion of universal consistency.

Theorem 2. *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 $c_t/\Lambda_t \leq 7\overline{U}|1/(\mu\Lambda_t) + \zeta + B\mu/\zeta|$.*

The proof, which adapts the method used in Fudenberg and Levine (1999) in the case $\lambda = 1$ can be found in the Web Appendix.

6. Convergence to Strict Nash Equilibrium

We study simultaneous move games with observable actions $w(a_1, a_2, \ldots, a_n)$. Say that a pure action profile is a *δ-strict Nash equilibrium* if each player loses at least δ from deviating to any pure action. Then we can show

Theorem 3. *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 that if $|w| \leq \overline{U}$ and the game*

w has a $\psi\overline{U}$-strict Nash equilibrium then with probability one the learning procedures converge to some strict Nash equilibrium.

Proof. Set $\epsilon_1 = \epsilon/4$ and $\epsilon_2 = \epsilon/(4\overline{U})$ (and also smaller than $1/2$). Define $y^i = a^{-i}$. For each player choose a 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]. Next choose ζ sufficiently small that two properties hold. First, $7\overline{U}\zeta \leq \epsilon_1$. Second, note that as $\zeta \to 0$ then $\hat{\alpha}$ approaches the best response, so in particular the probability of a strict best response goes to 1. Hence we can also choose ζ small enough that if a^i is any $\psi\overline{U}$-strict best response then $\hat{\alpha}^i(\gamma^i)[a^i] \geq 1 - \epsilon_2$. Then choose μ such that $7\overline{U}B\mu/\zeta \leq \epsilon_1$. This is $2\epsilon_1$ universally consistent by Theorem 2. By Theorem 1 we can choose M^i large enough that $\overline{U}E[|\tilde{\phi}_t^i - \phi_t^i|] \leq \epsilon_1$, and suppose that $k^i = M^i$, that is, we potentially discard all observations. Then the procedure replacing ϕ_t^i with $\tilde{\phi}_t^i$ is $3\epsilon_1$ universally consistent, since the payoffs remain within that distance. Now define a procedure $\overline{\alpha}^i(h_t^i)$ such that if all the observations in the memory are identical and $\hat{\alpha}^i(\gamma^i)[a^i] \geq 1 - \epsilon_2$ then $\overline{\alpha}^i(\gamma^i)[a^i] = 1$ (we call this the "stuck" state), otherwise, $\overline{\alpha}^i(\gamma^i)[a^i] = \hat{\alpha}^i(\gamma^i)[a^i]$. This procedure is no worse than $3\epsilon_1 + \overline{U}\epsilon_2 = \epsilon$ universally consistent.

Suppose that α^i, α^{-i} is $\psi\overline{U}$ strict. and that γ^i contains only these observations for $i = 1, 2, \ldots, n$. Then we see from construction that such a state is absorbing.

Next suppose that all players are in the stuck state. Observe that all must play a strict best response, since the probability of a non-strict best response is less than $1/2$ and so less than ϵ_2 in the original procedure. If the best responses are identical to the samples, we are are absorbed in a strict equilibrium. If not then the sample must change for all but one player, and in particular the next period at least one player is not in the stuck state.

Now suppose that at least one player is not in the "stuck" state. Then that player plays an action different than his last period action with a positive probability bounded below with a bound that depends only on ζ, ν both of which are fixed, and with positive probability remains unstuck. Hence the next period there is a positive probability that all players are unstuck. When all players are unstuck there is a positive probability that all play the $\psi\overline{U}$ strict equilibrium action, and there is a positive probability that all observations in all their samples are replaced with this action, resulting in the absorbing state. □

Notice that we do not assert that all weighted universally consistent learning procedures have a convergence property. This is unlikely to be the case, since **?** study smooth fictitious play with exponentially decreasing weights and show that the limit of the ergodic distribution as decay vanishes is contained in a Birkhof center of the flow of the corresponding mean field. They use this to determine that in some games the process converges to (close to an) equilibrium while in others it cycles.

What about mixed equilibria, or since mixed equilibrium will be difficult to hit with a finite number of states, mixed approximate equilibrium? Hart and Mas-Colell (2013) gives uncoupled learning algorithms that converge with probability one to a mixed equilibrium. However Hart and Mas-Colell (2013)'s learning procedures cannot be universally consistent because once in equilibrium play never changes regardless of the data.. By contrast Foster and Young (2006)'s procedure continues learning and does not converge with probability one to a Nash equilibrium, but the set of Nash equilibria does have probability near one in the ergodic distribution. We do not know if their procedure is universally consistent. At this point the issue of what sorts of extensions of Theorem 3 apply to games with only mixed equilibria remains open; we hope to explore it in future work.

7. Conclusion

We examine two models of learning with recency, recursive weights and limited memory. These are similar in the sense that have the same mean beliefs, and with high probability beliefs that are very close. We show that recursive weights with suitably smoothed best responses is weighted universally consistent, and argue that this is a sensible criterion. It follows that limited memory has the same property provided the grid is fine enough. This is useful because it can produce limited memory algorithms that are weighted universally consistent and also converge with probability one to strict Nash equilibrium. This is the first example of which we are aware of a learning processes that has global convergence to Nash equilibrium and is also shown to satisfy any sort of criteria of adequate responsiveness to individual incentives.

References

Agarwal, S., Driscoll, J. C., Gabaix, X., and Laibson, D. (2008): “Learning in the credit card market,” National Bureau of Economic Research.

Benaïm, M., J. Hofbauer, and E. Hopkins (2009): “Learning in Games with Unstable Equilibria,” *Journal of Economic Theory* 144: 1694-1709.

Benjamin, D., A. Bodoh-Creed, and M. Rabin (2013): “The Dynamics of Base Rate Neglect,” in preparation.

Camerer and Ho (1999) Camerer, C., and T. Ho (1999): “Experience-weighted Attraction Learning in Normal Form Games,” *Econometrica*, 67: 827-874.

Cheung, Y. and D. Friedman (1997): “Individual Learning in Normal Form Games: Some Laboratory Results,” *Games and Economic Behavior* 19: 46-76.

Erev, I. and Haruvy, E. (2013). Learning and the economics of small decisions. Forthcoming in *The Handbook of Experimental Economics,* Vol. 2 Roth, A.E. & Kagel, J. Eds.

Foster, D. P. and H. P. Young (2006): “Regret Testing: Learning to Play Nash Equilibrium without Knowing You have and Opponent,” *Theoretical Economics*, 1: 341-367.

Fudenberg, D. and D.K. Levine (1995): “Consistency and Cautious Fictitious Play,” *Journal of Economic Dynamics and Control* 19:1065-1089.

Fudenberg, D. and D. K. Levine (1999): “Conditional Universal Consistency,” *Games and Economic Behavior*, 29:104-130.

Fudenberg, D. and A. Peysakhovich (2013) “Recency, Records, and Recaps:The Effect of Feedback on Behavior in a Simple Decision Problem,” mimeo.

Hannan, J. (1957): “Approximation to Bayes Risk in Repeated Plays," in C*ontributions to the Theory of Games, 3*, ed. by M. Dresher, A. Tucker, and P. Wolfe, pp. 97{139.

Hart, S. and A. Mas-Colell (2013): *Simple Adaptive Strategies, From Regret-Matching to Uncoupled Dynamics*, World Scientific Publishing, Singapore.

Sutton and Barto (1998) Sutton, R. , and A. Barto (1998): Reinforcement learning: An introduction .Cambridge Univ Press, Cambridge UK.

Young, P. (1993): “The Evolution of Conventions,” *Econometrica* 61: 57-84

Web Appendix

Theorem. *[Theorem 1 in text] 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 $(y_1, y_2, \ldots)$.*

Proof. Fix the sequence f_t, define

$$z_t = \frac{\tilde{\phi}_t - (1-\mu)\tilde{\phi}_{t-1}}{\mu} - f_t$$

and observe that $\tilde{\phi}_t - \phi_t = \sum_{\tau=1}^{t} \lambda^t z_t / \Lambda_t$. Hence to prove the theorem it is sufficient to prove that $E[|z_t|] \to 0$.

Now let ℓ be the number of observations in updating to period t that are discarded and let $\tilde{\phi}^{\ell}_{t-1}$ be the frequencies in the remaining sample. Note that we can think of this as drawing $M - \ell$ observations from , $\tilde{\phi}_{t-1}$ without replacement, and we arbitrarily define $\tilde{\phi}^{M}_{t-1} = \tilde{\phi}_{t-1}$. Since $\mu < 1$ it is less than some $\overline{\mu} < 1$. By definition

$$\begin{aligned}
E[|z_t||\tilde{\phi}_{t-1}, \ell] &= E\left[\left|\frac{\tilde{\phi}_t - (1-\mu)\tilde{\phi}_{t-1}}{\mu} - f_t\right| \middle| \tilde{\phi}_{t-1}, \ell\right] \\
&= E\left[\left|\frac{\frac{\ell}{M} f_t + \frac{M-\ell}{M}\tilde{\phi}^{\ell}_{t-1} - (1-\mu)\tilde{\phi}_{t-1}}{\mu} - f_t\right| \middle| \tilde{\phi}_{t-1}, \ell\right] \\
&= E\left[\left|\frac{\frac{M-\ell}{M} - (1-\mu)}{\mu}\tilde{\phi}^{\ell}_{t-1} - \frac{(1-\mu)[\tilde{\phi}_{t-1} - \tilde{\phi}^{\ell}_{t-1}]}{\mu} + \left(\frac{\frac{\ell}{M}}{\mu} - 1\right) f_t\right| \middle| \tilde{\phi}_{t-1}, \ell\right] \\
&\le E\left[\left|\frac{\frac{M-\ell}{M} - (1-\mu)}{\mu}\tilde{\phi}^{\ell}_{t-1}\right| + \left|\frac{(1-\mu)[\tilde{\phi}_{t-1} - \tilde{\phi}^{\ell}_{t-1}]}{\mu}\right| + \left|\left(\frac{\frac{\ell}{M}}{\mu} - 1\right) f_t\right| \middle| \tilde{\phi}_{t-1}, \ell\right] \\
&\le 2\left|\frac{\ell}{\mu M} - 1\right| + \frac{1-\mu}{\mu}\left(Pr(\ell > (\mu/\sqrt{\overline{\mu}})M) + \max_{\ell \le (\mu/\sqrt{\overline{\mu}})M} E\left[\left|\tilde{\phi}_{t-1} - \tilde{\phi}^{\ell}_{t-1}\right| \middle| \tilde{\phi}_{t-1}, \ell\right]\right).
\end{aligned}$$

where the last line follows from the facts that both $\tilde{\phi}^{\ell}_{t-1}$ f_t ,and the difference between $\tilde{\phi}^{\ell}_{t-1}$,and $\tilde{\phi}_{t-1}$ are all between 0 and 1.

Next we observe that

$$E|X| \le 2\sqrt{E|X|^2}$$

hence it is enough to prove that each of the expectations on the RHS has square deviation that goes to zero. Examining first $E\left|\frac{\ell}{\mu M} - 1\right|^2$ recall that $\mu = pk/M$ and $E\ell = pk$, hence we need only compute the variance of $\frac{\ell}{\mu M}$. The variance of ℓ is $p(1-p)k = \mu M(1-p)$, so the variance of $\frac{\ell}{\mu M}$ is

$$\frac{\mu M(1-p)}{\mu^2 M^2} \le \frac{1}{\mu M}$$

.

Turning to the second term $E\left[\left|\tilde{\phi}_{t-1} - \tilde{\phi}^{\ell}_{t-1}\right|^2 \middle| \tilde{\phi}_{t-1}, \ell\right]$ observe that $E[\tilde{\phi}^{\ell}_{t-1}|\tilde{\phi}_{t-1}, \ell] = \tilde{\phi}_{t-1}$ and that the variance is bounded above by sampling with replacement, which is at worst $1/(M - (\mu/\sqrt{\overline{\mu}})M) \le 1/(M(1-\sqrt{\overline{\mu}}))$. Hence $\max_{\ell \le L} E\left[\left|\tilde{\phi}_{t-1} - \tilde{\phi}^{\ell}_{t-1}\right| \middle| \tilde{\phi}_{t-1}, \ell\right] \le 1/(M(1-\sqrt{\overline{\mu}})) \to 0$. □

Theorem. *[Theorem 2 in text] 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 $c_t/\Lambda_t \leq 7\overline{U}|1/(\mu\Lambda_t) + \zeta + B\mu/\zeta|$.*

Proof. Define: $\tilde{V}(\gamma) = \max_\alpha v(\alpha, \gamma)$, $\tilde{v}_t = \sum_{\tau=1}^{t} \lambda^\tau v(\alpha(h_t), f_t) + \sum_{\tau=-\infty}^{0} \lambda^\tau \tilde{V}(\phi_0)$, $\tilde{\Lambda}_t = \sum_{\tau=-\infty}^{t} \lambda^\tau$,$\tilde{c}_t = \tilde{\Lambda}_t V(\phi_t) - \tilde{v}_t$, $\tilde{c}_0 = 0$. Note that $\tilde{\Lambda}_t = \lambda^t/(1 - \lambda^{-1}) = \lambda^t/\mu$ and that

$$|\tilde{v}_t - u_t|/\Lambda_t = |\sum_{\tau=1}^{t} \lambda^\tau [v(\alpha(h_t), f_t) - u(\alpha(h_t), f_t)] + \sum_{\tau=-\infty}^{0} \lambda^\tau \tilde{V}(\phi_0)|/\Lambda_t \leq \zeta\overline{U} + (\tilde{\Lambda}_0/\Lambda_t)\overline{U}.$$

Our first step is to show that $|c_t/\Lambda_t - \tilde{c}_t/\tilde{\Lambda}_t|$ is small, so that we can focus on bounding $\tilde{c}_t/\tilde{\Lambda}_t$.

$$\begin{aligned}
|c_t/\Lambda_t - \tilde{c}_t/\tilde{\Lambda}_t| &= |(c_t - \tilde{c}_t)/\Lambda_t + \tilde{c}_t(\tilde{\Lambda}_t - \Lambda_t)/(\Lambda_t\tilde{\Lambda}_t)| \\
&= |\left(\Lambda_t U(\gamma_t) - u_t - [\tilde{\Lambda}_t V(\phi_t) - \tilde{v}_t]\right)/\Lambda_t + \tilde{c}_t(\tilde{\Lambda}_t - \Lambda_t)/(\Lambda_t\tilde{\Lambda}_t)| \\
&\leq |\tilde{v}_t - u_t|/\Lambda_t + |\left(\Lambda_t U(\gamma_t) - \tilde{\Lambda}_t V(\phi_t)\right)/\Lambda_t + \tilde{c}_t(\tilde{\Lambda}_t - \Lambda_t)/(\Lambda_t\tilde{\Lambda}_t)| \\
&\leq \overline{U}\zeta + (\tilde{\Lambda}_0/\Lambda_t)\overline{U} + |U(\gamma_t) - (\tilde{\Lambda}_t/\Lambda_t)V(\phi_t) + \tilde{c}_t\tilde{\Lambda}_0/(\Lambda_t\tilde{\Lambda}_t)| \\
&= \overline{U}\zeta + (\tilde{\Lambda}_0/\Lambda_t)\overline{U} + |U(\gamma_t) - U(\phi_t) + U(\phi_t) - V(\phi_t) + (\Lambda_t - \tilde{\Lambda}_t)V(\phi_t)/\Lambda_t + \tilde{c}_t\tilde{\Lambda}_0/(\Lambda_t\tilde{\Lambda}_t)| \\
&\leq \overline{U}\zeta + (\tilde{\Lambda}_0/\Lambda_t)\overline{U} + |U(\gamma_t) - U(\phi_t)| + |U(\phi_t) - V(\phi_t)| + | - \tilde{\Lambda}_0 V(\phi_t)/\Lambda_t + \tilde{c}_t\tilde{\Lambda}_0/(\Lambda_t\tilde{\Lambda}_t)| \\
&\leq \overline{U}\zeta + (\tilde{\Lambda}_0/\Lambda_t)\overline{U} + \overline{U}|\gamma_t - \phi_t| + \overline{U}\zeta + | - \tilde{\Lambda}_0 V(\phi_t)/\Lambda_t| + |\tilde{c}_t\tilde{\Lambda}_0/(\Lambda_t\tilde{\Lambda}_t)| \\
&= 2\overline{U}\zeta + (\tilde{\Lambda}_0/\Lambda_t)\overline{U} + \overline{U}|\gamma_t - \frac{\Lambda_t\gamma_t}{\tilde{\Lambda}_t} + \frac{\tilde{\Lambda}_0\phi_0}{\tilde{\Lambda}_t}| + | - \tilde{\Lambda}_0 V(\phi_t)/\Lambda_t| + |\tilde{c}_t\tilde{\Lambda}_0/(\Lambda_t\tilde{\Lambda}_t)| \\
&= 2\overline{U}\zeta + (\tilde{\Lambda}_0/\Lambda_t)\overline{U} + \overline{U}|\frac{\tilde{\Lambda}_t - \Lambda_t}{\tilde{\Lambda}_t}\gamma_t + \frac{\tilde{\Lambda}_0\phi_0}{\tilde{\Lambda}_t}| + | - \tilde{\Lambda}_0 V(\phi_t)/\Lambda_t| + |\tilde{c}_t\tilde{\Lambda}_0/(\Lambda_t\tilde{\Lambda}_t)| \\
&= 2\overline{U}\zeta + (\tilde{\Lambda}_0/\Lambda_t)\overline{U} + \overline{U}|\frac{\tilde{\Lambda}_t - \Lambda_t}{\tilde{\Lambda}_t}\gamma_t + \frac{\tilde{\Lambda}_0\phi_0}{\tilde{\Lambda}_t}| + | - \tilde{\Lambda}_0 V(\phi_t)/\Lambda_t| + |\tilde{c}_t\tilde{\Lambda}_0/(\Lambda_t\tilde{\Lambda}_t)| \\
&\leq 2\overline{U}\zeta + (\tilde{\Lambda}_0/\Lambda_t)\overline{U} + 2\overline{U}\frac{\tilde{\Lambda}_0}{\tilde{\Lambda}_t} + \frac{\tilde{\Lambda}_0}{\Lambda_t}|\max V(\gamma)| + \frac{|\tilde{c}_t}{\tilde{\Lambda}_t}\frac{|\tilde{\Lambda}_0}{\Lambda_t} \\
&\leq 2\overline{U}\zeta + 3\overline{U}\frac{1}{\mu\Lambda_t} + \frac{1}{\mu\Lambda_t}2\overline{U} + 2\overline{U}\frac{1}{\mu\Lambda_t} \leq 7\overline{U}(\zeta + 1/(\mu\Lambda_t))
\end{aligned}$$

Hence the result holds if we can bound $\tilde{c}_t/\tilde{\Lambda}_t$ by $B\overline{U}\mu$ for some B that depends only on ν. To do this, we define the incremental cost $\tilde{g}_t = \tilde{c}_t - \tilde{c}_{t-1}$, so that $\tilde{c}_t = \sum_{\tau=1}^{t} \tilde{g}_t$. Observe that if suppose that $\tilde{g}_t/\lambda^t \leq \epsilon$, then $\tilde{c}_t/\tilde{\Lambda}_t \leq \sum_{\tau=-\infty}^{t} \lambda^t\epsilon/\tilde{\Lambda}_t = \epsilon$ so we only need a bound on $\tilde{g}_t/\lambda^t$. We compute

$$\begin{aligned}
\tilde{g}_t &= \tilde{c}_t - \tilde{c}_{t-1} = \tilde{\Lambda}_t V(\phi_t) - \tilde{v}_t - \tilde{\Lambda}_{t-1}V(\phi_{t-1}) + \tilde{v}_{t-1} \\
&= \tilde{\Lambda}_t V(\phi_t) - \tilde{\Lambda}_{t-1}V(\phi_{t-1}) - \lambda^t v(\hat{\alpha}(\phi_{t-1}), f_t) \\
&= \tilde{\Lambda}_t V(\mu f_t + (1 - \mu)\phi_{t-1}) - \tilde{\Lambda}_{t-1}V(\phi_{t-1}) - \lambda^t v(\hat{\alpha}(\phi_{t-1}), f_t) \\
&\leq \tilde{\Lambda}_t\mu v(\hat{\alpha}(\phi_t), f_t) + \tilde{\Lambda}_t(1 - \mu)v(\hat{\alpha}(\phi_t), \phi_{t-1}) - \tilde{\Lambda}_{t-1}V(\phi_{t-1}) - \lambda^t v(\hat{\alpha}(\phi_{t-1}), f_t) \\
&\leq \lambda^t[v(\hat{\alpha}(\phi_t), f_t) - v(\hat{\alpha}(\phi_{t-1}), f_t)] - \lambda^t v(\hat{\alpha}(\phi_t), \phi_{t-1}) + (\tilde{\Lambda}_{t-1} + \lambda^t)v(\hat{\alpha}(\phi_t), \phi_{t-1}) - \tilde{\Lambda}_{t-1}V(\phi_{t-1}) \\
&\leq \lambda^t[v(\hat{\alpha}(\phi_t), f_t) - v(\hat{\alpha}(\phi_{t-1}), f_t)] - \lambda^t v(\hat{\alpha}(\phi_t), \phi_{t-1}) + \lambda^t v(\hat{\alpha}(\phi_t), \phi_{t-1}) \\
&= \lambda^t[v(\hat{\alpha}(\phi_t), f_t) - v(\hat{\alpha}(\phi_{t-1}), f_t)]
\end{aligned}$$

Examining the final term, observe that $\hat{\alpha}(\gamma)$ and $v(\alpha, \gamma)$ are Lipschitz continuous and that the Lipschitz constant has the form $B\overline{U}$ where B depends only on ν, hence

$$\begin{aligned} \tilde{g}_t &\leq \lambda^t[v(\hat{\alpha}(\phi_t), f_t) - v(\hat{\alpha}(\phi_{t-1}), f_t)] \\ &\leq \lambda^t(B\overline{U}/\zeta) \, \|\phi_t - \phi_{t-1}\| \\ &\leq \lambda^t B\overline{U}\mu/\zeta \end{aligned}$$

which now gives the desired overall bound. □