Learning
Outline and Reading List

Rajiv Sarin

March 4, 2008

I will talk about learning over these two lectures. Most of the models of learning I will be talking about in detail will be concerned about situation in which the agent has very limited information about the environment and experiences or receives very little information with each choice. Learning will be thought of as the behavior change that occurs in response to the information the agent receives. I will talk about both the short run properties, and the long run behavior, associated to particular learning rules, or classes of learning rules.

The environment I will place the agent in will be a (canonical) decision problem. I will not be considering learning in games.[1] Agents will not be assumed to know the probability distribution over outcomes of any action. It may be useful to think of the outcome set as monetary amounts in a compact interval, of which the agent is assumed to know the bounds. Other than that, the agent will only be assumed to know the available set of actions. Time will be discrete, and at any time (stage or period) the agent will take an action and receive a payoff.

I will also try to relate the analyses of learning models with evolutionary models and the expected utility model. Evolutionary models will be taken to those in which large populations of individuals, each identified with the action she chooses, and facing the same decision problem repeatedly. Payoffs in such models are fitness (or offspring). The expected utility model of von

[1] For which you might want to look at Fudenberg and Levine (1998), a recent and wide-ranging enough book on the subject.

Neumann and Mogenstern and Savage will be assumed to be known by all. In it, agents are assumed to know or have beliefs about the distributions over payoffs associated with each action. Payoffs, here, will be restricted to be monetary amounts.

1 References

1. Barron and Erev (2003): "Small feedcak-based decisions and their limited correspondence to description-based decisions," *Journal of Behavioral Decision Making*, 16, 215-233.

2. Beggs, A. (2005): "On the convergence of reinforcement learning," *Journal of Economic Theory*, 122, 1-36.

3. Berry, D. and B. Fristedt (1985): *Bandit Problems*, Chapman and Hall.

4. Börgers, T., A. Morales and R. Sarin (2004): "Expedient and monotone learning rules," *Econometrica*, 72, 383-405.

5. Börgers, T. and R. Sarin (1997): "Learning through reinforcement and replicator dynamics," *Journal of Economic Theory*, 77, 1-14.

6. Börgers, T. and R. Sarin (2000): "Naive reinforcement learning with endogenous aspirations," *International Economic review*, 41, 921-950.

7. Bush, R.R. and F. Mosteller (1951): "A mathematical model for simple learning," *Psychological Review*, 58, 313-323.

8. Bush, R.R. and F. Mosteller (1955): *Stochastic Models for Learning.* Wiley.

9. Camerer. C. and T. Ho (1999): "Experience-weighted attraction learning in normal-form games," *Econometrica*, 67, 827-874.

10. Cross, J. (1973): "A stochastic learning model of economic behavior," *Quarterly Journal of Economics*, 87, 239-266.

11. Cross, J. (1983): *A Theory of Adpative Economic Behavior*, Cambridge University Press.

12. Denrell, J. (2007): "Adaptive learning and risk taking," *Psychological Review*, 114, 177-187.

13. Easley, D. and A. Rustichini (1999): "Choice without beliefs," *Econometrica*, 67, 1157-1184.

14. Edwards, W. (1961): "Behavioral decision theory," *Annual Review of Psychology*, 12, 473-498.

15. Estes, W. (1950): "Toward a statistical theory of learning," *Psychological Review*, 57, 94-107.

16. Fudenberg, D. and D, Levine (1998): *The Theory of learning in Games.* MIT Press.

17. Haruvy, E. and I. Erev (2008): "Learning and the economics of small decisions," manuscript, http://www.utdallas.edu/~eeh017200/papers/LearningChapter.pdf.

18. Luce, R. (1959): *Individual Choice Behavior*, Wiley.

19. March, J. (1996): "Learning to be risk averse," *Psychological Review*, 103, 309-319.

20. Narendra, K. and M.A.L. Thathachar (1989): *Learning Automata: An Introduction.* Prentice Hall.

21. Norman, M.F.(1972): *Markov Processes and Learning Models.* Academic Press.

22. Oyarzun, C. and R. Sarin (2007): "Learning and risk aversion," manuscript.

23. Laslier, J.F. and B. Walliser (2005): "A reinforcement learning process in extensive-form games," *International Journal of Game Theory.*

24. Roth, A. and I. Erev (1995): "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term," *Games and Economic Behavior*, 8, 164-212.

25. Rothschild. M. (1974): "A two-armed bandit theory of market pricing," *Journal of Economic Theory*, 9, 185-202.

26. Sarin, R. and F. Vahid (1999): "Payoff assessments without probabilities: A simple dynamic model of choice," *Games and Economic behavior*, 28, 294-309.

27. Savage, L. (1954): *The Foundations of Statistics*, Wiley.

28. Schmalensee, R. (1975): "Alternate models of bandit selection," *Journal of Economic Theory*, 10, 333-342.

29. Simon, H. (1956): "A comparison of game theory and learning theory," *Psychometrica*, 21, 267-272.

30. Simon, H. (1959): "Theories of decision-making in economics and behavioral science," American Economic Review, XLIX, 253-283.

31. Sutton, R. and A. Barto (1998): *Reinforcement Learning*, MIT Press.

32. Thrall, R., C. Coombs and R. Davis (1954): *Decision Processes*, Wiley.

33. von Neumann, J and O. Morgenstern (1947): *The Theory of Games and Economic Behavior*, Princeton University Press.

2 Historical Introduction

Decision makers often don't know the distribution over outcomes among which they have to choose (as in von Neumann and Morgenstern (1947)). They may not even have (subjective) beliefs about these distributions (as in Savage (1954)). They, hence, may not be suitably described in the manner of choosing as in bandit problems (Rothschild (1974), Berry and Fristedt (1985)) when facing a repeated decision problem. There seems to be room for, and need of, models of choice behavior that make less demanding information or belief assumptions on the agents they are describing.

Almost simulatenously with the developement of (subjective) expected utility theory an alternate approach was taken by Bush and Mosteller (1951, 1955) and Estes (1950). This alternate approach (to decision making) does not assume people know the distributions of payoffs associated with various actions, and nor does it assume decision makers form beliefs about them. Decision makers choose randomly and adjust their behavior in response to their experiences (from the chosen action and payoff).

An early comparison of these two approaches to decision making in provided in many of the papers included in volume by Thrall, Coombs and Davis (1954). A similar exercise is conducted by Simon (1956). The book by Luce (1959) partially compares these two approaches. It also attempts to provide an axiomatic foundation for the learning approach. Bush, Mosteller and Thompson ("A formal structure for multiple-choice situations") in the volume by Thrall et. al. which attemtps to axiomatically develop a learning theory. See also the review article by Edwards (1961).

Cross (1973, 1983) had formally introduced their study in Economics. Simon (1959) had earlier, informally, promoted their usefulness for economics. Schmalensee (1975) had proposed then as an alternative to the Bayesian Bandit approach of Rothschild (1974). This approach, while it has never become the mainstream in Economics, has wide adherence in the machine learning literature (see, e.g., Narendra and Thathachar (1989) and also Sutton and Barto (1998)). Laslier and Walliser (2005) consider how to apply a specific learning model to describe learning in extensive form games.

In the recent experimental literature these learning models have re-appeared as they have been found useful in describing learning behavior in decision problems (e.g., Barron and Erev (2003)) and games (e.g., Camerer and Ho (1999), Roth and Erev (1995)). A related model, that has also been used to explain experimental data on learning in games, that was inspired by this literature was developed in Sarin and Vahid (1999).

3 Framework: Notation and Assumptions

Behavior is described as stochastic. Let A be the finite set of actions available to the decision maker. This set is assumed known by the decision maker. Let $\Delta(A)$ be the set of all probability distributions over A. Behavior is described by $\sigma \in \Delta(A)$. For notational convenience we shall suppose all actions are played with positive probability.

Time is discrete and is indexed by $n \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Initial behavior is $\sigma(0)$ is given. At each point in time, the decision maker chooses an action according to σ and receives a payoff. This information is used to update behavior.[2]

Let F_a be the distribution of payoffs $x \in X$ from action a and let $F = (F_a)_{a \in A}$ be the (payoff) environment the decision maker faces. The decision maker does not know F. No beliefs of the DM about F are assumed. The decision maker, however, is assumed to know some upper and lower bound on the possible payoffs. Hence, we may identify $X = [x_{\min},\ x_{\max}]$. It may be convenient to think of payoffs as monetary magnitudes, though this is not required. The expected payoff from action a is denoted by $\pi_a = \int x dF_a(x)$.

Let $L_{(a',x)}(a)$ denote the probability with which a is chosen in the next period if a' is chosen today and a payoff of x is received. Notice that, for given σ, specifying a learning rule as a function of x involves specifying a (finite, square) matrix of functions,

$$\begin{pmatrix} L_{(a,x)}(a) & L_{(a,x)}(a') & \cdots & \cdots \\ L_{(a',x)}(a) & L_{(a',x)}(a') & \cdots & \cdots \\ \vdots & \vdots & \ddots & \\ \vdots & \vdots & & \ddots \end{pmatrix}.$$

A learning rule tells us how the probability of *each* action $a' \in A$ is updated upon choosing any action a and receiving a payoff x.

The set of learning rules we consider should best be thought of as "reduced-form" learning rules. They only describe how behavior changes in response to experience. They do not model how the experience might change the beliefs the agent hold and through the change in belief result in different behavior.

4 Examples

1. The Cross learning rule (Cross (1973)) is given by

$$\begin{aligned} L_{(a,x)}(a) &= \sigma_a + (1 - \sigma_a) x \\ L_{(a',x)}(a) &= \sigma_a - \sigma_a x \quad \forall a' \neq a, \end{aligned}$$

[2] This is almost the framework assumed in Easley and Rustichini (1999). They, however, assume that the decision maker observes the payoff from all hher actions at each point in time.

where $x \in [0,1]$.

2. The Roth and Erev (1995) learning rule describes the state of an agent by a vector $v \in R_{++}^{|A|}$. The vector v describes the decision makers "attraction" to choose any of her $|A|$ actions. Given v, the agents behavior is given by $\sigma_a = v_a / \Sigma_{a'} v_{a'}$ for all a. If the agent plays a and receives a payoff of x then she adds x to her attraction to play a, leaving all other attractions unchanged. Hence, the Roth and Erev learning rule is given by

$$\begin{aligned} L^v_{(a,x)}(a) &= \frac{v_a + x}{\Sigma_{a''} v_{a''} + x} \\ L^v_{(a',x)}(a) &= \frac{v_a}{\Sigma_{a''} v_{a''} + x} \qquad \forall a' \neq a, \end{aligned}$$

where $x \in [0, x_{\max}]$ and the superscript v on the learning rule defines it at that state of learning.

3. Our next example considers the weighted return model (over gains) studied in March (1996). This learning rule is risk averse but may not be monotonically risk averse. The state of learning is described by a vector of attractions $v \in R_{++}^{|A|}$. Given v, the agents behavior is given by $\sigma_a = v_a / \Sigma_{a'} v_{a'}$ for all a. If action a is chosen and receives a payoff of x then she adds $\beta (x - v_a)$ to her attraction of a, where $\beta \in (0,1)$ is a parameter, leaving all other attractions unchanged. Thus, the learning rule may be written as

$$\begin{aligned} L^v_{(a,x)}(a) &= \frac{v_a + \beta(x - v_a)}{\sum_{a'' \in A} v_{a''} + \beta(x - v_a)} \\ L^v_{(a',x)}(a) &= \frac{v_a}{\sum_{a'' \in A} v_{a''} + \beta(x - v_{a'})} \qquad \forall a' \neq a, \end{aligned}$$

where $x \in [0, x_{\max}]$.

4. March (1996) also considers the fractional adjustment model. This rule is specified over environments in which the decision maker has only two actions. Over gains, this model is specified by

$$\begin{aligned} L_{(a,x)}(a) &= 1 - (1-\alpha)^x (1 - \sigma_a) \\ L_{(a'x)}(a) &= (1-\alpha)^x \sigma_a \qquad \forall a' \neq a, \end{aligned}$$

where $\alpha \in [0,1]$ (* and $x \in$???*). Over losses this model is specified as

$$\begin{aligned} L_{(a,x)}(a) &= (1-\alpha)^{-x}\sigma_a \\ L_{(a',x)}(a) &= 1-(1-\alpha)^{-x}(1-\sigma_a) \qquad \forall a' \neq a. \end{aligned}$$

5. Another learning rule that has received considerable attention is the logistic fictitious play with partial feedback studied by Fudenberg and Levine (1998, section 4.8.4). The agent is described by the $|A| \times 2$ matrix (v, κ) where κ_a denotes the number of times action a has been chosen, $\kappa = (\kappa_a)_{a \in A}$, and $v = (v_a)_{a \in A}$ gives the vector of attractions. The next period attraction of an action that was chosen today is its current attraction plus $(x - v_a)/(\kappa_a + 1)$. The attractions of unchosen actions are not updated. The learning rule is specified as

$$\begin{aligned} L^{v,\kappa}_{(a,x)}(a) &= \frac{e^{v_a+(x-v_a)/(\kappa_a+1)}}{e^{v_a+(x-v_a)/(\kappa_a+1)}+\sum_{a'\neq a} e^{v_{a'}}} \\ L^{v,\kappa}_{(a',x)}(a) &= \frac{e^{v_a}}{e^{v_{a'}+(x-v_{a'})/(\kappa_{a'}+1)}+\sum_{a''\neq a'} e^{v_{a''}}} \qquad \forall a' \neq a \end{aligned}$$

6. Börgers and Sarin (2000) consider the following learning rule.

5 Analysis

I will focus on two of the most widely studied rules in this section. The Cross (1973) learning rule and the Roth and Erev (1995) learning rule. The former has largely appeared in theoretical work, whereas the latter has appeared widely in both the theoretical and empirical literatures. My objective will be to bring you some of the results on these models of learning. Note that, both have the restricitive feature that the agent adds probability to whichever action she chooses. Implicitly, all payoffs of the agents are allowed/assumed to face in these rules are nice. This limits the applicability of these rules to larger classes of environments.