Evolution of Signaling: Part III - Learning

Models and Simulations in Philosophy
December 16th, 2013

This Month: The Platonic Puzzle concerning Meaning

The puzzle consists of three different questions:

- Definition: What makes a signal meaningful?
- Evolution: How did meaningful signals evolve?
- Stability: Why are signals stable?

Last Two Classes:

- Lewis' answers: signaling systems, higher-order knowledge, and coordination equlibria.
- Millikan's and Skyrms' criticisms/improvements to Lewis' solutions to the three questions.
- Replicator dynamics and the evolution of signaling [Skyrms, 2010].

Today: More on Evolution: ABMs and learning to signal

Outline

1. Review and Preview
2. Multi-Armed Bandit Problems
3. Learning Algorithms
   - Low Rationality
   - Roth-Erev
     - Bush-Mosteller
   - High Rationality
     - Simulated Annealing
     - Best-Response
4. Learning and Signaling Games
5. Signaling on Networks
6. References

One-Armed Bandits

An old-fashioned term for a slot machine is a one-armed bandit.

Consequently, the phrase multi-armed bandit problem is used to describe a mathematical model, in which

- There is a slot machine with several arms,
- The expected payoff of pulling different arms is different, but
- You don't know the expected payoffs.
- Luckily, you get to play the slot-machine at least several times.
- Obviously, your goal is to maximize your payoff.

There is an obvious trade-off in multi-armed bandit problems:

- Because you don't know that payoffs ahead of time, you must explore different arms.
- But as you play more, you want to exploit your knowledge and play the arm that seems to be best.

Question: How should you choose which arm to pull?

Model Simplifications

Simplifications:

- Each arm either gives (a) zero payoff or (b) a payoff of 1 dollar.
- So you just want to play the arm that has the highest probability of paying 1 dollar.

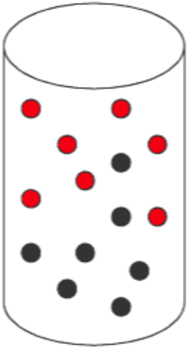
Learning Algorithms

Here are two common and simple learning algorithms:

1. Roth-Erev Reinforcement learning
2. Bush-Moseller Reinforcement Learning

Let's take them one at a time.

Roth-Erev Leaning

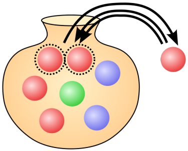


Imagine you have a big urn full of balls.

For each arm of the slot machine, there is an associated color.

Initially, there is one ball of each color.

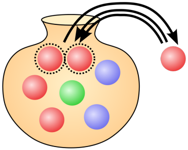
Roth-Erev Leaning



On each stage, you pick a ball at random from the urn, and do the corresponding action.

- No payoff: Put the ball back.
- Yes payoff: Put the ball back and **add** another ball of the same color. This is called reinforcement.

Roth-Erev Leaning



Idea: Actions with higher probabilities of success are reinforced more often.

So the associated colored balls become more prominent in the urn.

So those actions are in turn picked more often.

Roth-Erev Leaning

This is called Roth-Erev Reinforcement Learning.

A second type is called Bush-Mosteller.

To explain the general model, it's helpful to digress a bit about averaging probability distributions.

Vanilla Averages:

- Suppose Farmer Jim and Farmer Jane are equally reliable at estimating the weight of a cow.

Vanilla Averages:

- Suppose Farmer Jim and Farmer Jane are equally reliable at estimating the weight of a cow.
- Jim guesses the weight of a particular cow is 1173 pounds; Jane guesses 1256 pounds.

Vanilla Averages:

- Suppose Farmer Jim and Farmer Jane are equally reliable at estimating the weight of a cow.
- Jim guesses the weight of a particular cow is 1173 pounds; Jane guesses 1256 pounds.
- An intuitively reasonable way to improve our estimate is by averaging their guesses.

Now suppose Farmer Jane is more reliable than Farmer Jim at estimating the weight of a cow. Jim, however, is also pretty good.

WEIGHTED AVERAGES

Weighted Averages:

- An intuitively reasonable way to improve our estimate is by taking a weighed average:

$$\text{Guess} = (\alpha \cdot \text{Jane's guess}) + (1 - \alpha) \cdot (\cdot \text{Jim's guess})$$

  where $0 \leq \alpha \leq 1$.

- The closer α is to one, the more strongly our guess is determined by Jane's guess.

Weighted Averages:

- An intuitively reasonable way to improve our estimate is by taking a weighed average:

$$\text{Guess} = (\alpha \cdot \text{Jane's guess}) + (1 - \alpha) \cdot (\cdot \text{Jim's guess})$$

  where $0 \le \alpha \le 1$.
- The closer α is to one, the more strongly our guess is determined by Jane's guess.
- When $\alpha = \frac{1}{2}$, then our guess is an average, and we treat Jim and Jane as equally reliable.

Suppose now that instead of estimating the weight of the cow, Jim and Jane are estimating several probabilities.

E.g., There is a six-sided die, and Jim and Jane must estimated $P(1), P(2), P(3)$, and so on.

	Jim	Jane	Average
1	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{7}{24}$
2	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$
3	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{5}{24}$
4	$\frac{1}{9}$	0	$\frac{1}{18}$
5	$\frac{1}{9}$	0	$\frac{1}{18}$
6	$\frac{1}{9}$	0	$\frac{1}{18}$
Sum	1	1	1

Moral: An average of probability distributions is a probability distribution.

Weighted Averages of Probability Distributions

Fact: A weighted average of probability distributions is also a probability distribution.

That is, suppose p and q are probability distributions. Define r so that for all events E:

$$r(E) = \alpha p(E) + (1 - \alpha) q(E).$$

It is easy to check that r is a probability distribution.

Bush-Mosteller Learning:

- Start with an initial probability distribution p over arms in the slot-machine.
- On stage 1, an arm is chosen at random: the probability that arm a is pulled is $p(a)$.
  - If no success: Let $p' = p$
  - If success: Let q_a be the probability distribution that assigns probability one to playing arm a. Define a new probability distribution:

    $$p' = \alpha \cdot q_a + (1 - \alpha) \cdot p$$

    In this case, the action a is reinforced if $\alpha > 0$.
- On the next round, each arm a is pulled with probability $p'(a)$.

Idea 1: Actions with the highest chance of success will be reinforced most often.

Consequently, they will be played most often.

Recall, when an action is reinforced:

$$p' = \alpha \cdot q_a + (1 - \alpha) \cdot p$$

Idea 2: Higher $\alpha \Rightarrow$ Greater weight placed on recent observations. The past is less important.

Both Roth-Erev and Bush-Mosteller are considered "low rationality" rules. Why?

- Both can be simulated by mechanical processes (e.g. drawing balls from urns) that require no real insight.
- Neither rule requires maximizing utility.
- There is research supporting their empirical plausibility in animals with limited cognitive abilities.

In contrast "high" rationality rules require agents to maximize expected utility or do something similar in light of their evidence.

Exploration and Exploitation Again

- According to the traditional story, rational agents ought to maximized expected utility.
- Neither Roth-Erev nor Bush-Mosteller do so: they choose suboptimal actions with non-trivial probabilities.
- On the other hand, if one always picks a seemingly optimal action, then one may not test enough of the actions/arms.

This motivates the following strategy, which is called simulated annealing.

On stage n:

- With probability p_n, pick the action that has been most successful in the past.
- With probability $1 - p_n$, pick another action (uniformly) at random.

- If $p_n \rightarrow_{n \rightarrow \infty} 1$, then one plays a seemingly optimal action with increasing probability.
- If $1 - p_n$ does not decrease too slowly, then one is guaranteed to explore all actions infinitely often.

Best-Response

A second “high rationality” rule is applicable only in games: best-response.

This rule does what it sounds like it does.

Group Work: Explain how these rules can be followed in signaling games.

References I

Skyrms, B. (2010). *Signals: Evolution, learning, & information*. Oxford University Press.