Relation to ratifiability

April 20, 2018

For now, I'll only consider games without anthropics and without different situations. The agent submits a probability distribution once, an action is sampled from it and then the environment behaves in some way depending on that action and probability distribution.

A policy $\pi : \mathbb{N} \times \mathbb{R}^A \rightsquigarrow A = \{a_1, ..., a_n\}$ is a function mapping a time step and a mapping of empirical expected values / Q-values onto probability distributions over actions. Let P_i denote the sequence of probability distributions over A. Note that the P_i are random variables. Let $Q_i \in \mathbb{R}^A$ be the sequence of empirical EVs, again random variables. Let $U : A \times [0,1]^A \to \mathbb{R}$ be the actual EV given an action a and probability distribution p, where U is continuous in p. (Defining U such that it also assigns expected utilities to actions that are assigned zero probability is important for discussing the behavior of U in the limit.) (Unfortunately, this function cannot be as easily defined for problems involving anthropics and so forth.)

We say a sequence of random variables $(X_i)_{i\in\mathbb{N}}$ converges almost surely to x, or $X_i \underset{\text{a.s.}}{\to} x$, if ... (see `https://en.wikipedia.org/wiki/Convergence_of_random_variables#Almost_sure_convergence`). Conversely, we say that the sequence converges to x with positive probability, or $X_i \underset{\text{w.p.p.}}{\to} x$, if $P(\lim_{i\to\infty} X_n = x) > 0$.

Theorem 1. *Let π be a policy s.t. $\pi(i, \cdot)$ is continuous for all $i \in \mathbb{N}$. For each $q \in \mathbb{R}^A$ and each $j \in \{1, ..., n\}$ let*

$$\pi(i, q) \to \pi^\infty(q(a_1), ..., q(a_n)). \tag{1}$$

Assume that π^∞ "other things equal always gives higher probabilities to the actions with higher utility". Furthermore, let $U(a, p)$ be continuous in p for each $a \in A$ and let $P_i \underset{w.p.p.}{\to} \mathbf{p}$. Then there are $b_1, ..., b_n$ such that $b_j = a_j$ if $a_j \in supp(\mathbf{p}) = \{x \in A \mid \mathbf{p}(x) > 0\}$ and $b_j < \min_{a_k \in supp(\mathbf{p})} U(a_k, \mathbf{p})$ such that for all $a_j \in supp(\mathbf{p})$, it is

$$\mathbf{p}(a_j) = \pi^\infty(b_1, ..., b_n)(a_j). \tag{2}$$

Proof. **1.** We first show that for all $a_j \in supp(\mathbf{p})$, if indeed $P_i \to \mathbf{p}$, then

$$Q_i(a_j) \underset{\text{a.s.}}{\to} U(a_j, \mathbf{p}). \tag{3}$$

Hence we define for $a_j \in supp(\mathbf{p})$: $b_j = U(a_j, \mathbf{p})$

2. If $P_i \to \mathbf{p}$, then because π is continuous and prefers better options, there must be an N such that for all $i > N$ and all $a \notin supp(\mathbf{p})$ it must be

$$Q_i(a) < \min_{a_k \in supp(\mathbf{p})} U(a_k, \mathbf{p}). \tag{4}$$

3. As $P_i \to \mathbf{p}$, $Q_i(a_j)$ almost surely converges to some value even for $a_j \notin supp(\mathbf{p})$. Because of step 2, these values are smaller than $\min_{a_k \in supp(\mathbf{p})} U(a_k, \mathbf{p})$. Hence, we will use these limits as b_j.

4. If $Q_i(a_j) \to b_j$ for all $a_j \in A$, then

$$P_i \to \pi^\infty(b_1, ..., b_n). \tag{5}$$

5. From the conditions of the hypothesis and steps 1–4, it follows that with positive probability, it is both $P_i \to \mathbf{p}$ and for all $a_j \in supp(\mathbf{p})$

$$P_i(a_j) \to \pi^\infty(b_1, ..., b_n)(a_j). \tag{6}$$

Hence it must be for all $a_j \in supp(\mathbf{p})$

$$\mathbf{p}(a_j) = \pi^\infty(b_1, ..., b_n)(a_j), \tag{7}$$

where the b_j satisfy the claims made in the hypothesis. □

Corollary 2 (Ratifiability)**.** *Same conditions as for previous theorem, but also assume that π^∞ doesn't explore, i.e. that*

$$\pi^\infty(v_1, ..., v_n)(a_j) > 0 \iff j \in \arg\max_k v_k. \tag{8}$$

Then $U(a, \mathbf{p})$ is constant over $a \in supp(\mathbf{p})$.

Notes for generalization:

- Probably, I could easily generalize this to expected values conditional on some observation.
- If you explore all options infinitely often almost surely, you almost surely converge to a "really ratifiable" solution. If you don't explore all options infinitely often almost surely, there is a positive probability that it doesn't converge to a "really ratifiable" option.
- If it doesn't converge, then there is still ratifiability of some frequency construct, perhaps?
- Include anthropic cases.

Ratifiability of frequencies

Even if the probabilities do not converge at all, the frequencies of actions over many turns usually do. In fact, they often converge to ratifiable ones. E.g., in Death in Damascus, even if the probabilities do not converge, the frequencies converge to the ratifiable 50-50. But this does not seem to be true in general, even if the other prerequisites of the theorem are met. Roughly, the reason is the following: the frequencies arising from applying the learning algorithm are based on the success of actions for the success probabilities, rather than that frequency itself. So, the frequencies can converge to 50-50 based on how the actions behave if the probability is far removed from 50-50, even if at a (hypothetical) probability of 50-50, one of the actions is better than the other.

Some references on ratifiability

Some versions of the tickle defense are based on ratifiability, including