Jennifer Tillett

Bounded Rationality in the Iterated Prisoner's Dilemma

Supplement to address errors in LaTeX render

In *Definitions*:

Let π_i^G be the payoff for player i in game G. Fix a strategy σ_2 in the set of Player 2's strategies $\sum_2^G$. Let G^∞ be the limit of the means game and G^δ be the discounted game. For G^∞: A strategy σ_1 is *optimal* if for every strategy $\sigma_1' \in \sum_1^G$

$$\pi_1^{G^\infty}(\sigma_1, \sigma_2) - \pi_1^{G^\infty}(\sigma_1', \sigma_2) \geq 0. \tag{1}$$

For G^δ: A strategy σ_1 is *optimal* if for every strategy $\sigma_1' \in \sum_1^G$

$$\liminf_{\delta \to 1^-}(\pi_1^{G^\delta}(\sigma_1, \sigma_2) - \pi_1^{G^\delta}(\sigma_1', \sigma_2)) \geq 0. \tag{2}$$

A strategy is ϵ-*optimal* when 0 is replaced with -ϵ in the above equations. A strategy σ_1 is *dominant* if for every strategy $\sigma 2$ in $\sum_2^G$, σ_1 is optimal.

In *Machine Learning*:

- The learning rate λ decreases over time such that $\sum_{\lambda=0}^t \lambda = \infty$ and $\sum_{\lambda=0}^t \lambda^2 < \infty$.
- Each agent samples each of its actions infinitely often.
- The probability $P_t^i(a)$ of agent i choosing action a is nonzero.
- Each agent's exploration is exploitive. In other words, $\lim_{t\to\infty} P_t^i(X_t) = 0$, where X_t is a random variable denoting the event that some nonoptimal action was taken based on i's estimated values at time t.