PRISONER'S DILEMMA

ARVID LUNNEMARK.

SUPERVISED BY: MICHAEL SIPSER

1. Introduction

Introduce the prisoner's dilemma, the iterated prisoner's dilemma and the goal of finding out how cooperation can appear. Introduce Axelrod along with some other previous results.

Should some of the definitions be in the introduction?

2. Setup

Definition 1. The *prisoner's dilemma* is a symmetric two-player game with two actions, cooperate (C) and defect (D), where, if player 1 selects action a and player 2 selects action b, player 1 gets reward

$$r(a,b) = \begin{cases} R & \text{if } a = C, b = C \\ T & \text{if } a = D, b = C \\ S & \text{if } a = C, b = D \\ P & \text{if } a = D, b = D \end{cases}$$

We require $T > R > P > S$ and $2R > T + S$.

A common choice in simulations of the iterated prisoner's dilemma is $T = 5$, $R = 3$, $P = 1$ and $S = 0$.

We want to study the *iterated* prisoner's dilemma, for which we can define strategies that determine their next move based on the history of previous moves. As discussed previously, we want to restrict ourselves to strategies with finite memory.

Definition 2. A *strategy* s is a Moore machine (finite automaton with outputs) over the input and output alphabet $\{C, D\}$.

Notation-wise, we will use c to denote states in s, $G_s(c)$ to denote the output at state c, and $T_s(c, a)$ to denote the state that c transitions to upon receiving input a. For simplicity, we will also define the $\neg$ operator such that $\neg C = D$ and $\neg D = C$, and $c_{\text{start}}(s)$ to be the start state of s.

We will consider strategies in the presence of noise. To model that, we will assume that a strategy has a probability $1 - p$ of following the correct transition and a probability p of following the incorrect transition, at every step. Note that this models noise in *perception*. One could also imagine modeling noise in *action taken*, but it is easy to see that the two are equivalent up to a change of the values of R, S, T, P.

Definition 3. Suppose that strategy s_1 plays against strategy s_2. This defines an s_1-s_2 *Markov chain* where each state x is the vector (c_1, c_2) where c_1 is a state in s_1 and c_2 is a state in s_2. The transition probabilities are defined in the obvious way, using the error probability p.

We use the notation $G_{s_1,s_2}(c_1, c_2)$ to refer to the vector $(G_{s_1}(c_1), G_{s_2}(c_2))$, and we use S_{s_1,s_2} to refer to the set of all states in the Markov chain.

Definition 4. Let X_t be the random variable designating which state the s_1-s_2 Markov chain is in at time t. The payoff of strategy s_1 when played against strategy s_2 is

$$v_{s_1}(s_2) = E\left[\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} r(G_{s_1,s_2}(X_t)) \;\middle|\; X_0 = (c_{\text{start}}(s_1), c_{\text{start}}(s_2))\right]$$

That is, when s_1 plays against s_2, we define its payoff to be the average payoff over all possible infinite sequences of moves. Note that the expectation is taken over the infinite sequence $(X_0, X_1, \ldots)$. The limit inside is thus simply a normal time-average limit of bounded real numbers, which clearly exists.

We will now introduce the notion of a time average distribution which will lead us to a second way of defining the payoff $v_{s_1}(s_2)$.

is this really necessary?? should i pick one or the other?? should i move one of them to the next section? which one is easier to understand? do they trivially say the same thing?

Definition 5. The *time average distribution* of the s_1-s_2 Markov chain given the start state (a, b), denoted $\pi^{(a,b)}$, is the distribution such that

$$\pi^{(a,b)}_{c_1,c_2} = E\left[\text{fraction of time in state } (c_1, c_2) \mid \text{initial state is } (a, b)\right]$$

where the fraction of time is taken over the infinite sequence $(X_0, X_1, \ldots)$.

Lemma 1. The payoff when s_1 plays against s_2 is

$$v_{s_1}(s_2) = \sum_{(c_1,c_2)\in S_{s_1,s_2}} \pi^{(c_{\text{start}}(s_1), c_{\text{start}}(s_2))}_{c_1,c_2} \cdot r(c_1, c_2).$$

We may also make $r(c_1, c_2)$ into a vector, denoted by r, and write this as

$$v_{s_1}(s_2) = \pi^{(c_{\text{start}}(s_1), c_{\text{start}}(s_2))} \cdot r.$$

Proof of lemma 1. The key idea is that a time average sum where each element is one of finitely many values can be written as a frequency-weighted finite sum instead. Let $I_{c_1,c_2,t}$ be the indicator variable that is 1 if $G_{s_1,s_2}(X_t) = (c_1, c_2)$ and 0 otherwise. Then, we can write

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} r(G_{s_1,s_2}(X_t)) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \sum_{(c_1,c_2)\in S_{s_1,s_2}} r(c_1, c_2) \cdot I_{c_1,c_2,t}$$

We may now exchange the order of summation and move the finite sum out of the limit, to get

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T}\sum_{(c_1,c_2)\in S_{s_1,s_2}} r(c_1,c_2)\cdot I_{c_1,c_2,t} = \sum_{(c_1,c_2)\in S_{s_1,s_2}} r(c_1,c_2)\cdot \lim_{T\to\infty}\sum_{t=0}^{T}\frac{I_{c_1,c_2,t}}{T}$$

We can now use definition 4 and linearity of expectation to find that

$$v_{s_1}(s_2) = \sum_{(c_1,c_2)\in S_{s_1,s_2}} r(c_1,c_2)E\left[\lim_{T\to\infty}\sum_{t=0}^{T}\frac{I_{c_1,c_2,t}}{T} \mid X_0 = (c_{\text{start}}(s_1), c_{\text{start}}(s_2))\right]$$

Finally, we note that this is exactly the statement of lemma 1, which proves our lemma. □

Appendix A contains more details on time average distributions. In particular, if a unique stationary distribution exists, it is equal to the time-average distribution, which enables us to quickly find the time-average distribution in many cases.

We're now ready to look at how strategies interact.

Definition 6. A *population* of strategies $P = (S, f)$ is a set S of strategies and a function $f : S \to (0, 1]$ such that $\sum_{s\in S} f(s) = 1$, representing the frequency of each strategy in the population.

Definition 7. The *fitness* of a strategy s in a population $P = (S, f)$ is

$$F(s) = \sum_{s'\in S} f(s')v_s(s').$$

One can think of this as saying that we have infinitely many members of the population, and that they all interact with everyone else. This justifies the usage of expectation when definining $v_{s_1}(s_2)$.

We can now use the fitness of a strategy to compare it with other strategies in the same population. If a strategy s_1 has a higher fitness than another strategy s_2, that means that the frequency of s_1 will increase on the expense of the frequency of s_2, in the next step of the evolutionary process. This is getting us close to how we want to define stable strategies; our next move is looking not only at a single evolutionary step, but the entire evolutionary process.

Definition 8. A strategy s_1 is *ϵ-invadable* if there exists a strategy s_2 such that in all populations P with $S = \{s_1, s_2\}$ and $f(s_2) \geq \epsilon$, we have

$$F(s_2) > F(s_1)$$

That is, if s_1 is ϵ-invadable, there exists a strategy s_2 that can start as only a tiny fraction ϵ of the total population, and consistently have higher fitness than s_1, eventually causing overtaking s_1 completely. We are now finally ready to state our main definition.

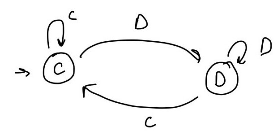

FIGURE 1. TFT.

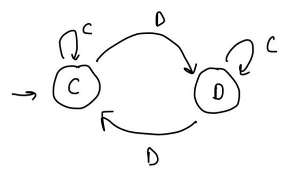

FIGURE 2. Pavlov.

Definition 9. A strategy s_1 is *evolutionarily stable* if there exists parameters p_0 and α, both in $(0,1)$, such that for all $p < p_0$, and all $\epsilon < \alpha$, s_1 is not ϵ-invadable.

That is, a strategy s_1 is evolutionarily stable if it can withstand invasion attempts from any strategy that starts off in low numbers, as the probability of noise tends to 0.

3. Results

We can now state our results! Together, the following two theorems prove that in the setup described here, mutual cooperation arises as the only stable choice.

Theorem 1. Suppose that a strategy s_1 is evolutionarily stable. Then $\lim_{p\to 0} v_{s_1}(s_1) = R$.

Theorem 2. The Pavlov strategy is evolutionarily stable.

use tikz for these figures!

Remark. Tit-for-tat, displayed in section 3, is not evolutionarily stable. It has the stationary distribution $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ in its own Markov chain, which has a payoff that is significantly smaller than R. This should be intuitive: if tit-for-tat makes one mistake, it goes into a defection cycle that it doesn't break out of until it makes a second mistake.

4. Proofs

4.1. **Helpful Lemmas.**

Lemma 2. The limit

$$\lim_{p\to 0} v_{s_1}(s_2)$$

exists, for any strategies s_1 and s_2.

Proof. Use the fact that this is the dot product of a stationary distribution with a fixed coefficient vector, and then say something about how any stationary distribution with the correct sum needs to be a polynomial in p maybe? $\square$

prove this; seems obvious

Lemma 3. For any strategy s,

$$v_s(s) \leq R$$

Proof. For notational simplicity, we will let s_1 and s_2 be two copies of strategy s. Then, $v_s(s) = v_{s_1}(s_2) = v_{s_2}(s_1)$. By definition, we have

$$v_{s_1}(s_2) = \sum \pi_{c_1,c_2} \cdot r(c_1, c_2)$$

and

$$v_{s_2}(s_1) = \sum \pi_{c_2,c_1} \cdot r(c_2, c_1).$$

Note that π_{c_1,c_2} and π_{c_2,c_1} refer to the same state, so we thus have

$$v_{s_1}(s_2) + v_{s_2}(s_1) = \sum \pi_{c_1,c_2} \cdot (r(c_1, c_2) + r(c_2, c_1))$$

which implies that

$$v_s(s) = \sum \left(\pi_{c_1,c_2} \cdot \frac{r(c_1, c_2) + r(c_2, c_1)}{2} \right).$$

Now, note that $r(c_1, c_2) + r(c_2, c_1) \in \{R + R, S + T, T + S, P + P\}$. Since $P < R$ and $T + S < 2R$, we thus find that

$$v_s(s) \leq \sum \pi_{c_1,c_2} \cdot R = R \sum \pi_{c_1,c_2} = R,$$

as desired. $\square$

4.2. **Evolutionary Stability Implies Utilitarianism.** With these lemmas, we are now ready to prove our first theorem.

Proof of theorem 1. Suppose that the strategy s_1 is such that it is *not* true that

$$\lim_{p \to 0} v_{s_1}(s_1) = R$$

By lemma 2 and lemma 3, this assumption implies that the limit is strictly less than R. Define $\gamma = v_{s_1}(s_1)$. We thus know that

$$\gamma < R.$$

We want to prove that s_1 is not evolutionarily stable.

To do that, we want to prove that for all $p_0, \alpha \in (0, 1)$, there exists $p < p_0$ and $\epsilon < \alpha$, such that s_1 is ϵ-invadable. We choose $\epsilon = \alpha/2$, and present a strategy s_2 that can invade s_1 for sufficiently small p.

We create the strategy s_2 as follows. First, copy the entire s_1 machine into s_2. Suppose that the state corresponding to the start state of s_1 is c_s. Recall that the output at c_s is $G(c_s)$, and that the state s goes to upon perceiving the opponent move $G(c_s)$ is $T(c_s, G(c_s))$. Now, create two new states: c_0 and c_1. Define the transitions as

$$\begin{aligned} T(c_0, G(c_s)) &= T(c_s, G(c_s)) \\ T(c_0, \neg G(c_s)) &= c_1 \\ T(c_1, \cdot) &= c_1 \end{aligned}$$

and the outputs as

$$\begin{aligned} G(c_0) &= \neg G(c_2) \\ G(c_1) &= C. \end{aligned}$$

add a figure!! the construction is simple but this description is deceivingly hard

Let the start state of s_2 be c_0.

Claim 1. Given the above construction of s_2, the following inequalities hold:

$$\begin{aligned} v_{s_1}(s_1) &\leq (1-p)^2\gamma + 2(1-p)pR + p^2R \\ v_{s_1}(s_2) &\leq (1-p)\gamma + pT \\ v_{s_2}(s_1) &\geq (1-p)\gamma + pS \\ v_{s_2}(s_2) &\geq (1-p)^2R + 2(1-p)p(\tfrac{S+T}{2}) + p^2\gamma. \end{aligned}$$

Before proving this claim, we will use it to finish our proof of theorem 1.

Now, we simply compute $F(s_2) - F(s_1)$, which we want to show is greater than 0.

$$\begin{aligned} F(s_2) - F(s_1) = \\ &= (1-\epsilon)\cdot v_{s_2}(s_1) + \epsilon \cdot v_{s_2}(s_2) - (1-\epsilon)\cdot v_{s_1}(s_1) - \epsilon \cdot v_{s_1}(s_2) \\ &= (1-\epsilon)(\gamma + p(\dots)) + \epsilon(R + p(\dots)) - (1-\epsilon)(\gamma + p(\dots)) - \epsilon(\gamma + p(\dots)) \\ &= \epsilon(R-\gamma) + p(\dots) \end{aligned}$$

We know that $R - \gamma > 0$ by our initial assumption. Clearly, since $(\dots)$ is some polynomial in p, given an ϵ we can find a sufficiently small p such that the full expression is positive. This proves that s_2 can invade s_1, and thus, that s_1 is not ϵ-invadable for this value of p. In conclusion, then s_1 is not evolutionarily stable, which concludes the proof of theorem 1.

□

Proof of claim 1. We can prove this using either of the two definitions. □

4.3. **Evolutionarily Stable Strategies Exist.**

Proof of theorem 2. TBC.

□

5. Discussion of Model

5.1. **Potential Other Models.** Right now we have only modeled noise in perception. One could think of another possible kind of noise: a "failure of the mind," which perhaps could be modeled instead by a probability p of being transported to any random state, instead. This would create ergodicity which is nice.

6. Appendix: Non-stationary limiting distributions

We might have periodicity, but for our purposes, we might as well extend the definition and look at periodic distributions as stationary too. The following two lemmas help with that.

Lemma 4. Given a starting distribution v and a Markov matrix M, for every $\epsilon > 0$, there will exist a k such that $|vP^{nk} - vP^{mk}| < \epsilon$ for all n and $m > 0$.

This proves that a Markov chain will always reach a periodic state.

Lemma 5. Suppose distributions form a chain $p_1 \to p_2 \to \cdots \to p_n \to p_1$. Then $\pi = \frac{p_1 + \ldots p_n}{n}$ is stationary.

This proves that we're able to talk about stationary distributions even when they don't really actually exist.