Multi-agent learning 2017-18, retake

This exam consists of five items. You may use 2.5 hours to complete the exam. No exit in the first half hour. No internet, no notes. Calculators are allowed. Answers must be justified. In particular, numeric answers must be justified by a full computation. Less is more: incorrect answer fragments and/or unnecessary long answers may lead to subtraction. Points are evenly divided over items. Within items points are evenly divided over sub-items (if any).

Clearly circle problem numbers on answer sheets. (Facilitates finding answers. Thank you.)

Good luck!

1. Two players repeat the following game an indefinite number of times. The probability to continue is $0 \leq \delta < 1$.

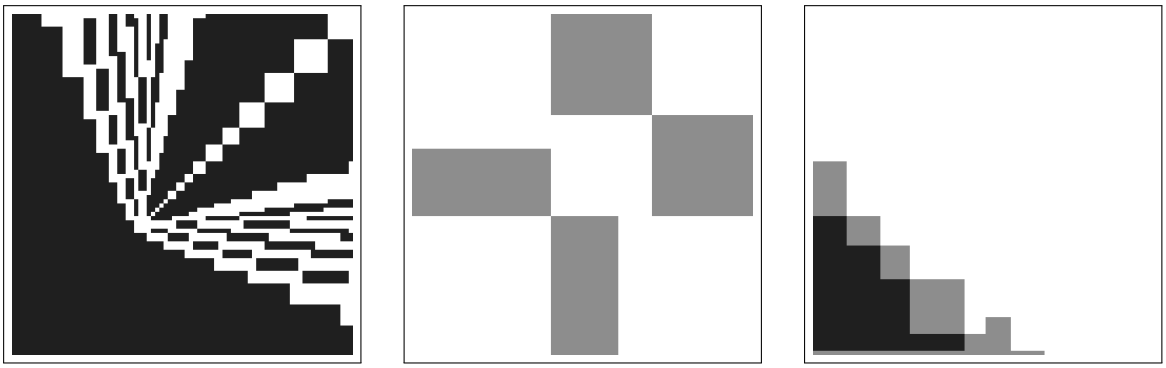
	C	D
C	2,2	0,3
D	3,0	1,1

   The strategy of the row player is to alternate between C and D (starting with C) as long as its opponent alternates between D and C (starting with D). Else the row player falls back to playing D forever. The column player maintains a similar strategy, i.e., to comply with alternating actions and to fall back to D whenever the row player defects. Sample realisation of play:

$$\omega = \begin{matrix} \text{row: } CDCDCDCDCDDDDD \\ \text{col: } DCDCDCDCDDDDDD \end{matrix} \cdots$$

   Compute the values for δ for which the strategies just described form a Nash equilibrium in the repeated game.

2. The following graphs show results of satisficing play. Give an interpretation of axes and colors. Categorise patterns of play. Give all parameters involved. (It is not necessary to give parameter values.)



3. Suppose two agents use Bayesian learning to guide their play in the repeated coordination game. Each considers the set of possible response rules to be the set of all 1-recall response rules. (So this gives 32 mappings from outcomes in the previous turn to actions plus an initial move. For instance, $CC \to D$, $CD \to D$, $DC \to D$, $DD \to D$, initial move D would be the "always defect" strategy.)

   Assume every player starts with a uniform prior over all 32 strategies. (E.g., Pr("always defect") = 1/32.) Use Bayesian updating to calculate the resulting beliefs of a player after observing the following play: CC, CD, DD, DC. Calculate the best response rule for this player to adopt given these new beliefs.

4. Give a formula for smoothed fictitious play. Describe its relation with fictitious play, describe its relation with no-regret, and describe its convergence properties. (You may miss out on one.)

5. Determine the evolutionarily stable strategies of

	L	R
L	−2,−2	1,0
R	0,1	−1,−1

.

End of document.

Answers

1, p. 1: See the slides on repeated games, and/or the corresponding chapter of "Game Theory: A Multi-Leveled Approach" (H. Peters).
A reason for row to defect would be when column plays C. Row would then gain 2 over 0, only to be punished by col from then on. So, if at all, row should defect in an odd round. Suppose row defects in round n. Payoff as from round n:

$$\underbrace{2\delta^{n-1}}_{\text{defect}} + \underbrace{\delta^n + \delta^{n+1} + \ldots}_{\text{col's punishment}} = 2\delta^{n-1} + \delta^n(1 + \delta + \delta^2 + \ldots)$$
$$= 2\delta^{n-1} + \delta^n \frac{1}{1-\delta} \qquad \text{(limit sum of geometric series, basic math).}$$

(The payoff in round n is multiplied by δ^{n-1}, not by δ^n. Remember, δ is the probability to continue, and round $n = 1$ is the first round.) Payoff as from round n if row would not defect:

$$\begin{aligned} 0\delta^{n-1} + 3\delta^n + 0\delta^{n+1} + 3\delta^{n+2} + 0\delta^{n+3} + \ldots &= 3\delta^n + 3\delta^{n+2} + 3\delta^{n+4} + \ldots \\ &= 3\delta^n(1 + \delta^2 + \delta^4 + \ldots) \\ &= 3\delta^n \frac{1}{1-\delta^2}. \end{aligned}$$

The row player had better not defected if (and only if):

$$\begin{aligned} & 2\delta^{n-1} + \delta^n \frac{1}{1-\delta} < 3\delta^n \frac{1}{1-\delta^2} \\ \Leftrightarrow\ & 2 + \delta \frac{1}{1-\delta} < 3\delta \frac{1}{1-\delta^2}, \text{ and } \delta \neq 0 \\ \Leftrightarrow\ & 2(1-\delta^2) + \delta(1+\delta) < 3\delta, \text{ and } \delta \neq 0 \\ \Leftrightarrow\ & 2 - 2\delta^2 + \delta + \delta^2 < 3\delta, \text{ and } \delta \neq 0 \\ \Leftrightarrow\ & 2 - \delta^2 < 2\delta, \text{ and } \delta \neq 0 \\ \Leftrightarrow\ & \delta^2 + 2\delta - 2 > 0, \text{ and } \delta \neq 0. \end{aligned}$$

Zero points with root formula (Dutch: abc-formule):

$$\delta = \frac{-2 \pm \sqrt{4 - 4 \cdot 1 \cdot -2}}{2 \cdot 1} = \frac{-2 \pm 2\sqrt{1-(-2)}}{2} = -1 \pm \sqrt{3}.$$

Because $0 < \delta < 1$, the only relevant root is

$$\delta = -1 + \sqrt{3}.$$

It follows that it is not in the row player's interest to defect if and only if $-1 + \sqrt{3} < \delta \leq 1$. Similarly for the column player: the fact that the column player has reason to defect in other rounds than row is irrelevant because the inequalities hold true in the limit. So it is not in both player's interest to defect if and only if the probability to continue is larger than $-1 + \sqrt{3}$.

2, p. 1: – (Interpretation of axes, 1 point). The graphs show how initial aspirations of two players (x-axis, y-axis) determine the outcome of play.

– (Interpretation of colors, 1 point). The outcome of play may be mutual cooperation (white), mutual defection (black), or other mutual behaviour (gray).

– (Patterns of play, 1 point). Besides stationary behaviour (white, black, sometimes gray) behaviour may be periodic (DD-CD-DD-CD, gray) or even chaotic (gray).

– (Parameters, 1 point). The parameters involved are: initial action player 1, initial action player 2, reward for mutual cooperation, reward for mutual defection, learning rate.

3, p. 1: After observing this history, our player will only consider two opponent strategies possible: opponent: CC → D; CD → D; DC → C; DD → C, initial move C or CC → D; CD → D; DC → D; DD → C, initial move C, each with probability 1/2. All other possible strategies are inconsistent with this play and hence will get zero weight.
A best response rule for our player at this point would be to echo his opponents moves insofar they are known. Further, DC → C and DC → D are equally likely, given the history of play. We end up with the set

$$\{\text{ CC} \to \text{D; CD} \to \text{D; DC} \to \text{p; DD} \to \text{C, initial move C} \mid 0 \leq p \leq 1 \}.$$

Any element of this set would yield the maximum possible payoffs for either of the two possible opponent strategies.

4, p. 1: See the slides on fictitious play and/or the corresponding chapter of "Strategic Learning an its Limits (H. Peyton Young, 2004).

- *Formula*. Let $x_i^1, \ldots, x_i^n$ the actions that are at the disposal of player i. Let y_{-i} player i's counterprofile of empirical frequencies of play. Let $u_k = u_i(x_i^k, y_{-i})$ player i's utility for playing action x_i^k against counterprofile y_{-i}. Let $\gamma > 0$ be the smoothing parameter. Let p_k the probability of playing action x_i^k. Then
$$p_k = \frac{e^{u_k/\gamma}}{\sum_{j=1}^n e^{u_j/\gamma}}.$$
- *Relation with fictitious play*. $\gamma \downarrow 0$ approaches fictitious play.
- *Relation with no-regret*. For every $\epsilon > 0$ and sufficiently small γ, regrets are bounded above by ϵ a.s.
- *Convergence properties*. For every $\epsilon > 0$ and sufficiently small γ, the empirical frequencies of play converge to the set of coarse correlated ϵ-equilibria a.s.

5, p. 1: See the slides on evolutionary game theory, and/or the corresponding chapter of "Game Theory: A Multi-Leveled Approach" (H. Peters).
This game has two pure equilibria $(1,0)$, $(0,1)$ and one mixed equilibrium $p = (1/2, 1/2)$. (See the game theory slides on how to determine mixed equilibria.) The mixed equilibrium is also symmetric. Only symmetric equilibria are candidates for ESSs, so p is the only equilibrium to consider.
Since p is fully mixed, every response q is a best response to p (again, see the game theory slides to see why the latter is true):

$$\text{for all } q : q^T A p \geq p^T A p. \text{ In particular for all } q : q^T A p = p^T A p.$$

So the first condition of an ESS is violated. We'll have to verify the second condition of an ESS:

$$\text{for all } q \neq p : q^T A q < p^T A q.$$

Let $q = (y, 1-y)$, $y \neq 1/4$, be arbitrary. Then

$$p^T A q = \begin{pmatrix} 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} y \\ 1-y \end{pmatrix} = -y,$$
$$q^T A q = \begin{pmatrix} y & 1-y \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} y \\ 1-y \end{pmatrix} = -4y^2 + 3y - 1.$$

It is easy to verify that

$$y \in [0,1] \backslash \{\frac{1}{2}\} \Rightarrow -4y^2 + 3y - 1 < -y$$

which means that the second condition is satisfied. It follows that p is an equilibrium that corresponds to an ESS.

Last modified on 11-7-2018.