Multi-agent learning 2016-17, mid term exam

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. In 1995, Erev and Roth proposed the update formula $\theta^{t+1} = \lambda\theta^t + e^t u^t$, where $0 \leq \lambda \leq 1$ determines a decay of previous propensities. It can be shown that

$$\Delta q^t = \frac{u^t}{\sum_{s \leq t} \lambda^{t-s} u^s}(e^t - q^{t-1}).$$

   Show this expression changes at a rate proportional to $1 - \lambda$, provided $\lambda < 1$ and the payoffs are bounded away from zero.

3. Suppose the following game is played infinitely many times:

	L	R
T	0,4	1,3
B	2,0	3,5

   .

   Suppose the following realisation of play:

round:	1	2	3	4	5	6	7	8	...
row:	T	T	B	T	T	B	T	B	...
column:	L	R	R	R	R	R	L	L	...

   Complete this table with payoffs and regrets per action for the column player. (Three extra rows.)

   Suppose we are in round nine. Give the probability that the column player plays L if it uses regret matching.

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

   .

Answers

1, p. 1: See the slides on repeated games, and/or the corresponding chapter of "Game Theory: A Multi-Leveled Approach" (H. Peters).
The best moment for row to defect would be when it is supposed to play C. When it plays D instead, it enjoys an instant payoff of 1 instead of 0. Suppose row defects in a certain round. Its expected payoff from then on would be

$$\underbrace{1}_{\text{defect}} + \underbrace{1\cdot\delta^1 + 1\cdot\delta^2 + 1\cdot\delta^3 + \dots}_{\text{fallback}} = \frac{1}{1-\delta}.$$

If row did not defect, its expected payoff from then on would be

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

So row has not an incentive to defect if and only if

$$\frac{1}{1-\delta} \le 3\delta\frac{1}{1-\delta^2}.$$

This is the case exactly when $\delta \in [1/2, 1)$. (Solve a second-degree equation with the root formula or, better, write $1-\delta^2 = (1+\delta)(1-\delta)$ and simplify.)
Similarly for col. (The fact that col would deviate in even rather than in odd rounds is irrelevant, for the computation of forfeited profit starts the round in which the deviation took place. And then the situation is symmetrical.)
So there is a Nash equilibrium in the repeated game if and only if $\delta \in [1/2, 1)$. If $\delta = 1/2$ the Nash equilibrium is unstable (a.k.a. weak): no player has an incentive to unilaterally deviate from its strategy, but no player has an incentive to stick to its strategy either. If $\delta = 1$, there is no Nash equilibrium, because in all scenarios the expected payoff is unbounded.

2, p. 1: See the slides on reinforcement learning and/or Ch. 2 of "Strategic Learning an its Limits (H. Peyton Young, 2004).
Let $m = \min\{ u^s \mid s \le t \}$ and $M = \max\{ u^s \mid s \le t \}$. For $\lambda < 1$ we have

$$\left(\sum_{s\le t} \lambda^{t-s}\right) m \;\le\; \sum_{s\le t} \lambda^{t-s} u^s.$$

Letting $t \to \infty$ yields

$$\frac{1}{1-\lambda} m \;\le\; \sum_{s\le t} \lambda^{t-s} u^s.$$

If $m > 0$ (here we use the fact that payoff are bounded away from zero)

$$\frac{1}{\sum_{s\le t} \lambda^{t-s} u^s} \;\le\; \frac{1-\lambda}{m}.$$

Hence

$$\begin{aligned} \frac{u^t}{\sum_{s\le t} \lambda^{t-s} u^s} &\le u^t \frac{1-\lambda}{m} \\ &\le \frac{M}{m}(1-\lambda). \end{aligned}$$

Since $\|(e^t - q^{t-1})\| \leq 1$,

$$\begin{aligned}\|\Delta q^t\| &= \left\| \frac{u^t}{\sum_{s\leq t} \lambda^{t-s} u^s} (e^t - q^{t-1}) \right\| \\ &= \frac{u^t}{\sum_{s\leq t} \lambda^{t-s} u^s} \|(e^t - q^{t-1})\| \\ &\leq \frac{M}{m}(1-\lambda) \cdot 1 \\ &\sim 1-\lambda.\end{aligned}$$

The "$\sim$" stands for "is proportional to", which means "equal, up to some (often uninteresting) factor".

3, p. 1: See the slides on no-regret and/or Ch. 2 of "Strategic Learning an its Limits (H. Peyton Young, 2004).

round nr.:	1	2	3	4	5	6	7	8	…
action of row:	T	T	B	T	T	B	T	B	…
action of column:	L	R	R	R	R	R	L	L	…
payoff:	4	3	5	3	3	5	4	0	…
regret L:	0	1	−5	1	1	−5	0	0	…
regret R:	−1	0	0	0	0	0	−1	5	…

Average regret for L in round eight is $(3-10)/8 = -7/8$. Average regret for R in round eight is $(5-2)/8 = 3/8$.

$$q_x^{t+1} = \frac{[\bar{r}_x^t]_+}{\sum_{x'\in X}[\bar{r}_{x'}^t]_+},$$

$$\text{so } \; q_L^9 = \frac{[\bar{r}_L^8]_+}{\sum_{x\in\{L,R\}}[\bar{r}_x^8]_+} = \frac{[-7/8]_+}{[-7/8]_+ + [3/8]_+} = \frac{0}{0+3/8} = 0.$$

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 Wednesday 8th May, 2019, 14:39 uur.