University of Illinois **Fall 2017**

ECE 586BH: Problem Set 3

Fictitious play dynamics, minimum regret learning from forecasters

Due: Tuesday, October 17, at beginning of class
Reading: Course notes, Chapter 3

1. **[Which symmetric games are potential games?]**
A two player normal form game $G = (\{1,2\}, S_1, S_2, u_1, u_2)$ is symmetric if $S_1 = S_2$ and $u_1(i,j) = u_2(j,i)$ for any $i, j \in S_1$. The following definition makes sense because u_2 is determined by u_1 for such a game. Define a function u_1 on a set $S_1 \times S_1$ to have property P if it is the payoff function for player 1 in a symmetric two person potential game.

   (a) If there is a function g such that $u_1(i,j) = g(i)$ for all $i, j \in S_1$, does u_1 have property P ? If so, give the potential function. If not, give a counter example.

   (b) If $u_1(i,j) = H(i,j)$ such that H is symmetric (i.e. $H(i,j) = H(j,i)$ for all $i, j \in S_1$) does u_1 have property P? If so, give the potential function. If not, give a counter example.

   (c) For S_1 fixed and a function $u_1 : S_1 \times S_1 \to \mathbb{R}$, does property P hold if and only if there exist $g : S_1 \to \mathbb{R}$ and $H : S_1 \times S_1 \to \mathbb{R}$ such that H is symmetric and $u_1(i,j) = g(i) + H(i,j)$ for all $i, j \in S_1$? Justify your answer. Hint: How would you construct and verify a potential function for the symmetric game, given u_1?

2. **[Doubling trick]**
Proposition 3.18 of the notes on prediction with expert advice provides the upper bound $\max_i R_{i,n} \leq \sqrt{2n \ln N}$ for the exponentially weighted forecaster with parameter $\eta = \sqrt{\frac{2 \ln N}{n}}$. In some situations there is not a single time horizon n known in advance, so it would be nice to have a forecaster and similar maximum regret bound that holds uniformly over time. That can be accomplished by letting η get smaller with time. A simple approach to this problem is to partition time into periods $I_0 = \{1\}$, $I_1 = \{2,3\}$, ..., $I_k = \{2^k, \ldots, 2^{k+1} - 1\}, \ldots$ doubling in length. Instead of carrying over information from one interval to the next, suppose we simply apply the algorithm used Proposition 3.18 separately for each interval.

   (a) Verify that the proof of Proposition 3.18 shows that a bit more is true, namely, $\max_i R_{i,n'} \leq \sqrt{2n \ln N}$ for any n' with $1 \leq n' \leq n$.

   (b) For the forecaster based on the doubling trick, describe the choice of η in each interval and find the smallest constant c so that $\max_i R_{i,n} \leq c\sqrt{2n \ln N}$ for all $n \geq 1$. (Hint: The cumulative regret at $t \in I_k$ is the sum for intervals I_0 through I_k.

3. **[Existence of Hannan consistent estimators]**

   (a) Show that if $(D_t : t \geq 0)$ is a martingale difference sequence such that $|D_t| \leq 1$ for all t, then the corresponding martingale defined by $X_n = X_1 + \cdots + X_n$ satisfies $\limsup_{n\to\infty} \frac{X_n}{2\sqrt{n \ln n}} \leq 1$ almost surely (i.e. with probability one). (Hint: Use the Azuma-Hoeffding inequality and the Borel Cantelli lemma – both in the notes.)

   (b) Combine part (a) and previous problem to show a Hannan consistent estimator exists.

4. **[One Hannan consistent player]**
Consider a finite two-player zero-sum game such that $\ell : S_1 \times S_2 \to \mathbb{R}$ serves as the loss function of player 1 and reward function of player 2, as usual. Suppose the game is played repeatedly to produce a sequence of outcome tuples $(I_t, J_t)_{t\geq 1}$.

   (a) State what it means in this instance for player 1 to use a Hannan consistent strategy.

   (b) If player 1 uses a Hannan consistent strategy, and if p^* is a limit point of the empirical distribution of plays of player one, $(\hat{p}_n^I)_{n\geq 1}$, then must p^* be a minmax optimal strategy for player 1 for a single play of the game? If yes, give a proof. If no, give a counter example.

5. **[Blackwell approachability for a simple game with vector valued payoff]**
Consider a two player game with strategy spaces

$$S_1 = \left\{ \binom{0}{1}, \binom{0}{-1}, \binom{1}{0}, \binom{-1}{0} \right\} \text{ and } S_2 = \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right\},$$

such that if player 1 selects $x \in S_1$ and player 2 selects $y \in S_2$, the payoff vector for player 1 is $x + y$.

   (a) An arbitrary half space H in the plane can be expressed as $H = \{v : v \cdot u \leq c\}$, where u is a unit length vector in $\mathbb{R}^2$ and $c \in \mathbb{R}$. Under what condition on u and c is this half space approachable (for player 1)?

   (b) Let $B(r)$ denote the disk in $\mathbb{R}^2$ of radius r. Find the minimum value of r so that $B(r)$ is approachable.

6. **[Pursuit evasion game]**
Consider a two-player, zero sum game with $S_1 = S_2 = \{1, \ldots 6\}$ and $\ell(i,j) = \mathbf{1}_{\{|i-j|\geq 2\}}$. Player 1 seeks to select i to minimize $\ell(i,j)$, i.e., to have $|i-j| \leq 1$, whereas player 2 seeks to select j to maximize $\ell(i,j)$, i.e. to have $|i-j| \geq 2$. Basically, player 1 is trying to get close to player 2 while player 2 wants to evade player 1.

   (a) Show that for any Nash equilibrium, or equivalently, saddlepoint, (in mixed strategies), each player wins with equal probability.

   (b) Find the minmax strategy for player 1 (it is unique) and all the maxmin strategies for player 2. Simplify your answer as much as possible.

7. **[Simulations for pursuit evasion game]**
Run Python simulations of repeated play of the game in the previous problem. See the Jupyter notebook at
http://nbviewer.jupyter.org/urls/courses.engr.illinois.edu/ece586GT/fa2017/ece586GTps3.ipynb?flush_cache=true. You are free to make use of the notebook for your assignment, for example by modifying some part of it. Investigate three scenarios, with both players using:

Fictitious play (FP) At each time $t+1$, each player plays a strategy that is a pure strategy best response to the empirical average of the previous t plays of the other player. You can break ties any way you'd like.

Regularized fictitious play (RFP) Weighted predictors with exponential weights $w_{i,t} = \mathrm{e}^{-\eta_t L_{i,t}}$ with $\eta_t = \frac{\eta_1}{t}$. (Perhaps taking $\eta_1 = 10$ or $\eta_1 = 100$ works well.)

Optimized regularized fictitious play (ORFP) Weighted predictors with exponential weights $w_{i,t} = \mathrm{e}^{-\eta_t L_{i,t}}$ with $\eta_t = \sqrt{8(\ln(6)/t}$.