Homework #2
CS699 Fall 2017

Due Friday Nov 10, by 04:00pm

General Instructions The following assignment is meant to be challenging. Feel free to discuss with fellow students, though please write up your solutions independently and acknowledge everyone you discussed the homework with on your writeup. We also expect that **you will not attempt to consult outside sources, on the Internet or otherwise,** for solutions to any of these homework problems. Finally, unless otherwise stated please provide a formal mathematical proof for all your claims.

Note To submit the homework, you may either email it to Li, hand it to him in person, or drop it off in the box which we will make available in SAL 246 on Friday. If you need more time, consider using one of your late days.

Problem 1. (10 points)
Recall that, in the online learning setting with n actions and T time steps, the multiplicative weights algorithm we saw in class achieves average external regret $O\left(\sqrt{\frac{\log n}{T}}\right)$. One deficiency of this algorithm, in the form it was presented in class, is that it required knowledge of T in advance (recall that we used T in order to set the parameter ϵ). One might want an online learning algorithm which guarantees vanishing regret for all T simultaneously. Such an algorithm would be useful in scenarios in which the number of decisions T is finite and unknown, as well as scenarios in which the algorithm is designed to be run in perpetuity over an infinite time horizon.

Show, by way of a black-box reduction, how to construct an algorithm with average external regret $O\left(\sqrt{\frac{\log n}{T}}\right)$ for all T simultaneously. You may assume that you have access to an algorithm with the same guarantees as the multiplicative weights algorithm. However, you may only invoke such an algorithm as a black box.

Problem 2. (15 points)
Recall that a two-player zero sum game can be described by a matrix $A \in \mathbb{R}^{m\times n}$, where A_{ij} is the row player's utility when the row player chooses action i and the column player chooses action j. (The column player's utility is $-A_{ij}$). Also recall that the value of the game, defined as the row player's utility at equilibrium, is given by $v^* = \max_{x\in\Delta_m} \min_{y\in\Delta_n} x^T Ay = \min_{y\in\Delta_n} \max_{x\in\Delta_m} x^T Ay$. (The latter equality is the minimax theorem). In class, we showed that if a zero-sum game is played repeatedly using no-external-regret dynamics, then the row player's utility converges to v^*, and the column player's utility converges to $-v^*$.

a. (8 points). As in class, let T be the number of time steps, and let $x_t \in \Delta_m$ and $y_t \in \Delta_n$ denote the row player and column player's mixed strategies at time step $t \in \{1, \ldots, T\}$ when both players are using a no-external-regret algorithm. Show that the average history of joint play, defined as the mixed strategy profile $(\frac{\sum_{t=1}^T x_t}{T}, \frac{\sum_{t=1}^T y_t}{T})$, converges to the set of Nash equilibria of the game as T grows large, in the following sense: the mixed strategy profile $(\frac{\sum_{t=1}^T x_t}{T}, \frac{\sum_{t=1}^T y_t}{T})$ is an ϵ-Nash equilibrium for $\epsilon = \epsilon(T)$ tending to 0 as $T \to \infty$.

b. (7 points). Show that there exists a two-player non-zero-sum game for which no-external-regret dynamics do not converge to a Nash equilibrium. Do this by explicitly describing such a game, and showing that when both players use the the multiplicative weights algorithm, the average history of joint play is far from a Nash equilibrium regardless of the time horizon T.

Problem 3. (10 points)
Given m independent samples $s_1, \ldots, s_m$, drawn from the probability distribution $p : [n] \to [0, 1]$, the *empirical distribution* $\widehat{p}_m : [n] \to [0, 1]$ is defined as follows: for all $i \in [n]$,

$$\widehat{p}_m(i) = \frac{|\{j \in [m] \mid s_j = i\}|}{m}.$$

The *total variation distance* between two distributions $p, q : [n] \to [0, 1]$ is defined as $d_{\text{TV}}(p, q) := \max_{A \subseteq [n]} |p(A) - q(A)| = (1/2) \cdot \sum_{i=1}^n |p(i) - q(i)|$, where $p(A)$ denotes $\sum_{i \in A} p(i)$. We showed in class that, for $m = O(n/\epsilon^2)$, it holds $d_{\text{TV}}(p, \widehat{p}_m) \le \epsilon$ with probability at least 9/10.

Show that, for $m = O((n + \log(1/\delta))/\epsilon^2)$, it holds $d_{\text{TV}}(p, \widehat{p}_m) \le \epsilon$ with probability at least $1 - \delta$.

Problem 4. (10 points)
Consider the setting of a single-buyer, single-item auction. As discussed in class, this is a special case of Myerson's auction and the optimal auction is one which posts a "take it or leave it" price p maximizing $p \cdot (1 - F(p))$, where F is the CDF of the buyer's valuation distribution. We showed that if the distribution of the buyer's valuation is in the interval $[0, 1]$, then after $O(1/\epsilon^2)$ samples, we can compute a posted-price whose revenue is at least $(1-\epsilon)$-times the optimal revenue. We also showed that if the distribution is arbitrary, no finite sample upper bound is possible.

In this problem, we explore the same question for the family of Monotone Hazard Rate (MHR) distributions. Recall that a distribution over $\mathbb{R}_+$ with pdf $f(x)$ and CDF $F(x)$ is called MHR if the function $f(x)/(1 - F(x))$ is non-decreasing. Show an upper bound on the number of samples required to compute a posted-price whose revenue is at least $(1-\epsilon)$-times the optimal revenue when the buyer's valuation is known to be MHR.