Algorithmic Game Theory, Summer 2017 Lecture 4 (4 pages)

Online Learning and Coarse Correlated Equilibria

Instructor: Thomas Kesselheim

In view of the fact that Nash equilibria are hard to compute, it is at least questionable that players will reach them. Fortunately, there are other, weaker equilibrium concepts that generalize Nash equilibria but are easy to compute. In this lecture, we will get to know one of these concepts. It is particularly appealing because it is the limit point of natural learning dynamics.

1 Minimizing External Regret

Consider the following problem. There is a single player playing T rounds against an adversary, trying to minimize his cost. In each round, the player chooses a probability distribution over N strategies (also termed actions here). After the player has commited to a probability distribution, or mixed strategy as we will say, the adversary picks a cost vector fixing the cost for each of the N strategies.

In round $t = 1, \dots, T$, the following happens:

- The player picks a probability distribution $p^t = (p_1^t, \dots, p_N^t)$ over his strategies.
- The adversary picks a cost vector $\ell^t = (\ell_1^t, \dots, \ell_N^t)$, where $\ell_i^t \in [0, 1]$ for all i.
- A strategy a^t is chosen according to the probability distribution p^t. The player incurs this strategy's cost and gets to know the entire cost vector.

What is the right benchmark for an algorithm in this setting? The *best action sequence in hindsight* achieves a cost of $\sum_{t=1}^{T} \min_{i\in[N]} \ell_i^t$. However, getting close to this number is generally hopeless as the following example shows.

Example 4.1. *Suppose $N = 2$ and consider an adversary that chooses $\ell^t = (1, 0)$ if $p_1^t \geq 1/2$ and $\ell^t = (0, 1)$ otherwise. Then the expected cost of the player is at least $T/2$, while the best action sequence in hindsight has cost* 0*.*

Instead, we will swap the sum and the minimum, and compare to $L_{\min}^T = \min_{i\in[N]} \sum_{t=1}^{T} \ell_i^t$. That is, instead of comparing to the best action sequence in hindsight, we compare to the *best fixed action in hindsight.*

The expected cost of some algorithm $\mathcal{A}$ that uses probability distributions $p^1, \dots, p^T$ against cost vectors $\ell^1, \dots, \ell_T$ is given as $L_{\mathcal{A}}^T = \sum_{t=1}^{T} \sum_{i=1}^{N} p_i^t \ell_i^t$. The difference of this cost and the cost of the best single strategy in hindsight is called *external regret.*

Definition 4.2. *The* external regret *of algorithm $\mathcal{A}$ is defined as $R_{\mathcal{A}}^T = L_{\mathcal{A}}^T - L_{min}^T$.*

Definition 4.3. *An algorithm is called* no-external-regret algorithm *if for any adversary and all T we have $R_{\mathcal{A}}^T = o(T)$.*

This means that the *average* cost per round of a no-external-regret algorithm approaches the one of the best fixed strategy in hindsight or even beats it.

1.1 The Multiplicative-Weights Algorithm

By the definition it is not even clear that there are no-external-regret algorithms. Fortunately, there are. In this section, we will get to know the *multiplicative-weights algorithm* (also known as randomized weighted majority or hedge).

The algorithm maintains weights w_i^t, which are proportional to the probability that strategy i will be used in round t. After each round, the weights are updated by a multiplicative factor, which depends on the cost in the current round.

Let $\eta \in (0, \frac{1}{2}]$; we will choose η later.

- Initially, set $w_i^1 = 1$, for every $i \in [N]$.
- At every time t,
  - Let $W^t = \sum_{i=1}^N w_i^t$;
  - Choose strategy i with probability $p_i^t = w_i^t / W^t$;
  - Set $w_i^{t+1} = w_i^t \cdot (1-\eta)^{\ell_i^t}$.

Let's build up some intuition for what this algorithm does. First suppose $\ell_i^t \in \{0, 1\}$. Strategies with cost 0 maintain their weight, while the weight of strategies with cost 1 is multiplied by $(1-\eta)$. So the weight decays exponentially quickly in the number of 1's. Next consider the impact of η. Setting η to zero means that we pick a strategy uniformly at random and continue to do so, on the other hand the higher η the more we punish strategies which incurred a high cost. So we can think of η as controlling the tradeoff between exploration (small η) and exploitation (large η).

Theorem 4.4 (Littlestone and Warmuth, 1994)**.** *The multiplicative-weights algorithm, for any sequence of cost vectors from* $[0, 1]$*, guarantees*

$$L_{MW}^T \leq (1+\eta) L_{min}^T + \frac{\ln N}{\eta} \ .$$

Setting $\eta = \sqrt{\frac{\ln N}{T}}$ yields

$$L_{\text{MW}}^T \leq L_{\text{min}}^T + 2\sqrt{T \ln N} \ .$$

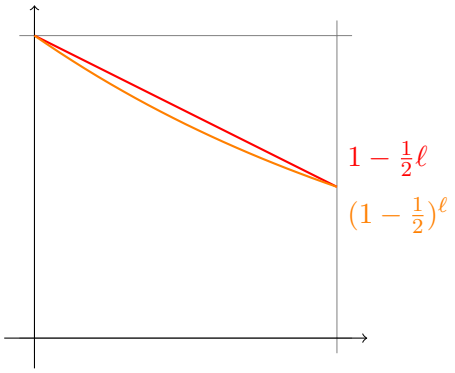
Corollary 4.5. *The multiplicative-weights algorithm with* $\eta = \sqrt{\frac{\ln N}{T}}$ *has external regret at most* $2\sqrt{T \ln N} = o(T)$ *and hence is a no-external-regret algorithm.*

Notice that this matches the above lower bound.

Proof. Let us analyze how the sum of weights W^t decreases over time. It holds

$$W^{t+1} \ = \ \sum_{i=1}^N w_i^{t+1} \ = \ \sum_{i=1}^N w_i^t (1-\eta)^{\ell_i^t} \ .$$

Observe that $(1-\eta)^\ell \ = \ (1-\ell\eta)$, for both $\ell = 0$ and $\ell = 1$. Furthermore, $(1-\eta)^\ell$ is a convex function in ℓ. For $\ell \in [0, 1]$ this implies $(1-\eta)^\ell \ \leq \ (1-\ell\eta)$.



This gives us

$$W^{t+1} \;\leq\; \sum_{i=1}^{N} w_i^t(1-\ell_i^t\eta) \;=\; W^t - \eta\sum_{i=1}^{N} w_i^t\ell_i^t \;.$$

Let ℓ^t denote the expected cost of MW in step t. It holds $\ell^t = \sum_{i=1}^{N} \ell_i^t w_i^t / W^t$. Substituting this into the bound for W^{t+1} gives

$$W^{t+1} \;\leq\; W^t - \eta\ell^t W^t \;=\; W^t(1-\eta\ell^t) \;.$$

As a consequence,

$$W^{T+1} \;\leq\; W^1\prod_{t=1}^{T}(1-\eta\ell^t) \;=\; N\prod_{t=1}^{T}(1-\eta\ell^t) \;.$$

The sum of weights after step T can be upper bounded in terms of the expected costs of MW. On the other hand, the sum of weights after step T can be lower bounded in terms of the costs of the best strategy as follows:

$$W^{T+1} \;\geq\; \max_{1\leq i\leq N}(w_i^{T+1}) \;=\; \max_{1\leq i\leq N}\left(w_i^1\prod_{t=1}^{T}(1-\eta)^{\ell_i^t}\right) \;=\; \max_{1\leq i\leq N}\left((1-\eta)^{\sum_{t=1}^{T}\ell_i^t}\right) \;=\; (1-\eta)^{L_{\min}^T} \;.$$

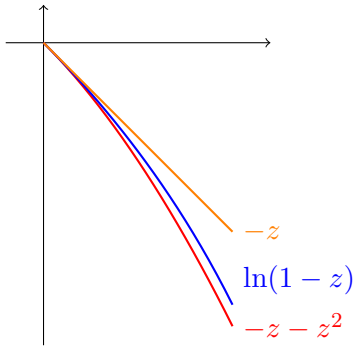
Combining the bounds and taking the logarithm on both sides gives us

$$L_{\min}^T \ln(1-\eta) \;\leq\; (\ln N) + \sum_{t=1}^{T}\ln(1-\eta\ell^t) \;.$$

In order to simplify, we will now use the following estimation

$$-z - z^2 \;\leq\; \ln(1-z) \;\leq\; -z \;,$$

which holds for every $z \in [0, \frac{1}{2}]$.



This gives us

$$\begin{aligned} L_{\min}^T(-\eta-\eta^2) \;&\leq\; (\ln N) + \sum_{t=1}^{T}(-\eta\ell^t) \\ &=\; (\ln N) - \eta L_{\mathrm{MW}}^T \;. \end{aligned}$$

Finally, solving for L_{MW}^T gives

$$L_{\mathrm{MW}}^T \;\leq\; (1+\eta)L_{\min}^T + \frac{\ln N}{\eta} \;. \quad \square$$

2 No-Regret Dynamics

If we repeatedly play a game, it is quite reasonable that a players play according to a no-regret algorithm. Let us now consider the outcomes that occur if every player does so. Note that in all these considerations the players do not necessarily use the multiplicative-weights algorithm or any other fixed algorithm.

That is, we fix a cost-minimization game. Without loss of generality, we assume that all cost values are from $[0, 1]$. We derive a sequence of mixed strategy profiles $\sigma^1, \ldots, \sigma^T$ as follows. In each round t, each player i decides on a probability distribution over his strategies, which we now interpret as a mixed strategy σ_i^t. Each player i observes a vector $c_i(\cdot, \sigma_{-i}^t)$, indicating the expected costs of alternative strategies. This vector is used as the cost vector ℓ^t and in the decision regarding σ_i^{t+1}. As a no-regret algorithm achieves vanishing regret per round for any adversary, this is particularly true for the costs induced by the game play of the other players.

Do such dynamics converge to Nash equilibria? Not necessarily. However, "on average" the players play according to an approximate coarse correlated equilibrium.

Definition 4.6. *An* ϵ-approximate coarse correlated equilibrium *(or* ϵ-coarse correlated equilibrium*) of a cost-minimization game is a probability distribution* p *on the set of strategy profiles* $S = \Pi_{i \in \mathcal{N}} S_i$ *such that for all* $i \in \mathcal{N}$ *and all* $s_i' \in S_i$ *we have*

$$\sum_{s \in S} p(s) \cdot c_i(s) \leq \left(\sum_{s \in S} p(s) \cdot c_i(s_i', s_{-i}) \right) + \epsilon \ .$$

The case of $\epsilon = 0$ *is called* coarse correlated equilibrium*.*

Note that any mixed Nash equilibrium is also a coarse correlated equilibrium. The difference is that in a mixed Nash equilibrium the random choices of different players are always independent. In a coarse correlated equilibrium, they might be correlated. The traditional understanding is that there is some kind of coordination device that tells players what to do. Deviating from this advice is not beneficial. This sounds actually very strange — who would coordinate such a game? In our case of simultaneous learning, however, the coordination is rather implicit because the players coordinate themselves by their history of play.

Proposition 4.7. *Let* $\sigma^1, \ldots, \sigma^T$ *be generated by no-regret dynamics such that each player's regret is at most* ϵT*. Let* p *be the probability distribution that first selects a single* $t \in [T]$ *uniformly and then chooses for every* $i \in \mathcal{N}$ *one* s_i *according to* σ_i^t*. Then* p *is an* ϵ*-coarse correlated equilibrium.*

One also says that the average history of play converges to a coarse correlated equilibrium.

Recommended Literature

- Chapter 4 in the AGT book.
- Tim Roughgarden's lecture notes `http://theory.stanford.edu/~tim/f13/l/l17.pdf` and lecture video `https://youtu.be/ssAEgJKRe9o`
- N. Littlestone, M. Warmuth. The Weighted Majority Algorithm. Information and Computation 108(2):212–261, 1994.