2/17/22:

End of last lecture: showed equivalent definition of correlated equilibria based on switching functions

Def: σ is a correlated equilibrium if for all $i \in [k]$ and for all $\delta: S_i \to S_i$,

$$\mathop{E}_{s \sim \sigma}[C_i(s)] \leq \mathop{E}_{s \sim \sigma}[C_i(s_{-i}, \delta(s_i))]$$

Swap Regret: Back to online learning setting

Saw if OPT is best action sequence in hindsight, can't compete

If OPT is best single action in hindsight, can compete: no-regret algorithms!

What about other actions between these: stronger than best single action, weaker than best sequence?

Intuition: what if in hindsight every time we played a we instead played b, every time we played b instead played c, etc.?

Def: The swap regret of a sequence of actions $a^1, a^2, a^3, \dots, a^T$ with respect to switching function

actions in online learning

$\delta : A \to A$ is

$$S_T(\delta) = \frac{1}{T}\left(\sum_{t=1}^{T} c^t(a^t) - \sum_{t=1}^{T} c^t(\delta(a^t)) \right)$$

Note: $R_T(a) = S_T(\delta)$ for $\delta(x) = a \;\; \forall x \in A$

So if low swap-regret $\forall \delta$ then low regret $\forall a$, but not necessarily vice versa

Def: Let $\mathcal{A}$ be an online learning algorithm. Then its expected swap regret with respect to $\delta : S_i \to S_i$ is

$$E[S_T^{\mathcal{A}}(\delta)] = \frac{1}{T}\left(\sum_{t=1}^{T} \mathop{E}_{a^t \sim p^t}[c^t(a^t)] - \sum_{t=1}^{T} \mathop{E}_{a^t \sim p^t}[c^t(\delta(a^t))] \right)$$

Def: $\mathcal{A}$ has no-swap-regret if $E[S_T^{\mathcal{A}}(\delta)] = o(1)$ as $T \to \infty$ for all $\delta : A \to A$

Re-do no-regret $\Rightarrow$ CCE proof to show no-swap-regret $\Rightarrow$ CE!

- $\mathcal{A}_i$: algorithm used by player i
- p_i^t: distribution (mixed strategy) used by player i at time t (generated by $\mathcal{A}_i$)
- $\sigma^t = p_1^t \times p_2^t \times \cdots \times p_k^t$ product distribution over S generated by players at time t
- $\sigma = \frac{1}{T} \sum_{t=1}^{T} \sigma^t$ time-averaged distribution.

Thm: Suppose $E[S_T^{\mathcal{A}_i}(\delta)] \le \varepsilon$ $\forall i \in [k]$, $\forall \delta : S_i \to S_i$.

Then σ is an ε-approximate correlated equilibrium:

$$\mathop{E}_{s \sim \sigma}[C_i(s)] \le \mathop{E}_{s \sim \sigma}[C_i(s_{-i}, \delta(s_i))] + \varepsilon$$

$\forall i \in [k]$, $\forall \delta : S_i \to S_i$

Pf: Let $i \in [k]$, $\delta: S_i \to S_i$

Define $c^t(a) = \mathbb{E}_{s \sim \sigma^t}[C_i(s_{-i}, a)]$.

learning cost

game cost

$$\Rightarrow \mathbb{E}_{a \sim p^t}[c^t(a)] = \mathbb{E}_{s \sim \sigma^t}[C_i(s)]$$

So:

$$\mathbb{E}_{s \sim \sigma}[C_i(s)] - \mathbb{E}_{s \sim \sigma}[C_i(s_{-i}, \delta(s_i))]$$

$$= \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{s \sim \sigma^t}[C_i(s)] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{s \sim \sigma^t}[C_i(s_{-i}, \delta(s_i))]$$

(def of σ)

$$= \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{a^t \sim p^t}[c^t(a^t)] - \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{a^t \sim p^t}[c^t(\delta(a^t))]$$

(def of c^t)

$$= \mathbb{E}[S_T^{A_i}(\delta)]$$

(def of swap regret)

$$\leq \varepsilon$$

Back to online learning: Can we build a no-swap-regret algorithm?

Idea: don't start from scratch. Use no-regret algorithms!

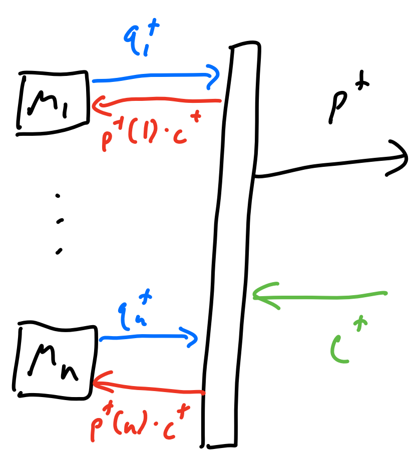
Thm [Blum, Mansour]: If there is a no-regret algorithm, then there is a no-swap-regret algorithm.

Black-box reduction!

Pf sketch:

- $A = [n]$
- $M_1, M_2, \ldots, M_n$ no-regret algorithms
  (could be different instantiations of same alg)

At time t:



- Receive distributions $q_1^t, q_2^t, \ldots, q_n^t$ from $M_1, M_2, \ldots, M_n$
- Compute "consensus distribution" p^t
- receive c^t from adversary
- Give M_j vector $p^t(j) \cdot c^t$

Time-averaged expected cost of algorithm:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n} p^t(i)\cdot c^t(i) \qquad \mathrm{I}$$

Time-averaged expected cost if we switched using switching function $\delta: A \to A$:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n} p^t(i)\, c^t(\delta(i)) \qquad \mathrm{II}$$

So want to prove $\mathrm{I} - \mathrm{II} = o(1)$ as $T \to \infty$

From perspective of M_j:

- no regret with respect to any fixed action, in particular $\delta(j)$
- But using perceived cost vectors, not real ones

more formally:

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n} q_j^t(i)\, p^t(j)\, c^t(i)}_{M_j\text{'s perceived expected cost}} - \underbrace{\frac{1}{T}\sum_{t=1}^{T} p^t(j)\, c^t(\delta(j))}_{M_j\text{'s perceived cost if always played } \delta(j)} \le R_T^{M_j}(\delta(j)) = o(1)$$

$\le O\left(\sqrt{\frac{\ln n}{T}}\right)$ for MW

Sum over all $j \in [n]$:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n} q_j^t(i)\, p^t(j)\, c^t(i) - \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{n} p^t(j)\, c^t(\delta(j)) \leq \sum_{j=1}^{n} R_T^{M_j}(\delta(j))$$

$$= o(1)$$

$$\leq O\left(\sqrt{\frac{n^2 \ln n}{T}}\right)$$

Second term = II !

So just want first term = I ; want

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n} q_j^t(i)\, p^t(j)\, c^t(i) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n} p^t(i) \cdot c^t(i)$$

Let's try to set p^t to make this true!

want $\sum_{j=1}^{n} q_j^t(i)\, p^t(j) = p^t(i) \quad \forall i, t$

New type of math to use: Markov Chains!

Markov chain with:

- states $A = [n]$,
- transition probability from $j \to i$ $= q_j^t(i)$

$\left(\sum_{i=1}^{n} q_j^t(i) = 1 \right)$

Since finite Markov chain, exists stationary distribution:

$$p^t(i) = \sum_{j=1}^{n} q_j^t(i)\, p^t(j)$$

Done!