No External Regret Algorithm

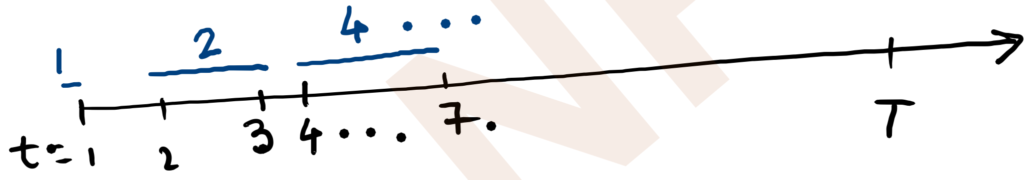
Lecture 7.1

Theorem: $|A| = n$. There exists an algorithm with external regret (time-averaged) $O\left(\sqrt{\frac{\log n}{T}}\right)$.

Multiplicate weight algorithm. Set $\varepsilon = \sqrt{\frac{\ln n}{T}}$. The problem with setting $\varepsilon = \sqrt{\frac{\ln n}{T}}$ is that we need to know the time horizon T in advance.

Remark: $|A| = n$. Then there exists an algorithm with time-averaged external regret $O\left(\sqrt{\frac{\ln n}{T}}\right)$. Moreover the algorithm does not need to know T a-priori.

Proof:



Group the iretations into epochs. ξ_i, $i = 0, \dots, l$

$$\xi_i = \left[2^i , \min\{ 2^{i+1} - 1 , T \} \right]$$

In the begining of each epoch, we reset the weight vector of the MW algorithm to the all 1's vector.

In the i-th epoch, we use $\varepsilon = \varepsilon_i = \sqrt{\frac{\ln n}{2^i}}$. Let OPT_i be the pay off of any fixed action in the i-th epoch.

$$OPT - \sum_{i=1}^{T} r_i \leq \sum_{i=0}^{l} \left[OPT_i - \sum_{t \in \varepsilon_i} r_i \right]$$

$$\leq \sum_{i=0}^{l} \left[\varepsilon_i \cdot 2^i + \frac{\ln n}{\varepsilon_i} \right]$$

$$= \sqrt{\ln n} \sum_{i=0}^{\ell} 2^{\frac{i}{2}+1}$$

$$\leq 2^{\frac{\ell}{2}+2} \cdot \sqrt{\ln n}$$

$$\leq 4\sqrt{T \ln n} \qquad \blacksquare$$

$$\frac{1}{T}\left(OPT - \sum_{i=1}^{T} r_i \right) \leq 4\sqrt{\frac{\ln n}{T}}$$

Combining Expert Advice: The problem of designing a no-regret algorithm is also called "combining expert advice" - the mixed strategy of every iteration is like "combining experts' advices" and the goal is to perform as good as an expert asymptotically.

Combining exploration and exploitation.

Connection between no-external-regret dynamic and equilibrium concepts:

Let $\Gamma = \langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$ be any normal form game.

Each player runs a no-external regret algorithm.

$$A_i = S_i$$

$$p_t \in \Delta(S_i)$$

$$\pi_i(s_i) = u_i(s_i, p_{-i}) = \sum_{s_{-i} \in S_{-i}} \left[u_i(s_i, s_{-i}) \prod_{\substack{j \in [n] \\ j \neq i}} p_j(s_j) \right]$$

Theorem: Let $\varepsilon > 0$, players run their no-external-regret algorithm for T iterations such that their time-averaged external regret is at most ε. Define $\sigma_i = \prod_{i \in N} p_i^t$,

$\sigma = \frac{1}{T} \sum_{i=1}^{T} \sigma_i$. Then σ is an ε-CCE of Γ.

That is $$\mathbb{E}_{s \sim \sigma}\left[u_i(s)\right] \geq \mathbb{E}_{s \sim \sigma}\left[u_i(s_i', s_{-i})\right] - \varepsilon \quad \forall i \in N \quad \forall s_i' \in S_i$$