Lecture 6.4

External Regret Framework

The setting: Iterative process between a player and an adversary.

- For each time step $t = 1, 2, \ldots, T$ $\{$
  - the player picks a probability distribution $p_t \in \Delta(A)$. // A is the set of actions available to the player
  - Adversary picks a utility function $\pi_t : A \to [0,1]$

- The player samples an action $a_t \sim p_t$ and receives a reward $\pi_t(a_t)$.

- The player gets to know π_t.

}

Total expected utility : $\sum_{t=1}^{T} \sum_{a \in A} \pi_t(a) \cdot p_t(a)$

Regret : $\underbrace{\boxed{\left(\sum_{t=1}^{T} \max_{a \in A} \pi_t(a) \right)}}_{B} - \sum_{t=1}^{T} \sum_{a \in A} \pi_t(a)\, p_t(a)$

Time-averaged regret : $\frac{1}{T}\left[\left[\sum_{t=1}^{T} \max_{a \in A} \pi_t(a)\right] - \left[\sum_{t=1}^{T} \sum_{a \in A} \pi_t(a)\, p_t(a)\right]\right]$

No regret dynamic/algorithm: if time-averaged regret goes to zero when $T \longrightarrow \infty$

Q: Does there exist any no regret algorithm?

Ans. NO!

$n = |A|$, $n = 2$, $A = \{a_1, a_2\}$

Adversary: $\left.\begin{array}{l}\pi_t(a_1) = 1 \\ \pi_t(a_2) = 0\end{array}\right\}$ if $p_t(a_1) < p_t(a_2)$

$\left.\begin{array}{l}\pi_t(a_1) = 0 \\ \pi_t(a_2) = 1\end{array}\right\}$ if $p_t(a_1) \geq p_t(a_2)$

$B = T$,

Expected utility of the player $\leq \frac{T}{2}$

Time averaged - regret $\geq \frac{1}{T}\left[T - \frac{T}{2}\right] = \frac{1}{2}$

Weaken the benchmark:

$$B = \max_{a \in A} \sum_{t=1}^{T} \pi_t(a)$$: External regret benchmark.

Q. Does there exist any no-external regret algorithm?

Ans. YES!

No-Regret Algorithm

Theorem: Let $|A| = n$. Then there exists a no-regret algorithm whose time averaged regret is $O\left(\sqrt{\frac{\log n}{T}}\right)$.

Corollary: There exists a no-regret algorithm whose expected time averaged regret is at most ε, for any $\varepsilon > 0$, after $O\left(\frac{\log n}{\varepsilon^2}\right)$ iterations.

The algorithm is called multiplicate weight (MW)/ hedge algorithm.