today: learning without full information

recall: used to compute

$$\sum_{\tau=1}^{t-1} u^{\tau}(s) \quad \text{all strategies } s$$

full information.

propensity scoring:

s^t choosen strategy

$$\bar{u}^t(s) = \begin{cases} u^t(s^t)/\text{prob}(s^t) & \text{choosen} \\ 0 & \text{if } s \neq s^t \end{cases}$$

Good news

$$E(\bar{u}^t(s)) = Pr(s^t \text{ choosen}) \cdot \bar{u}(s^t) + 0$$
$$= u(s^t)$$

trouble: $\bar{u}(s)$ can be big!

if $\bar{u}(s) \in [0, m]$ range

$\Longrightarrow$ regret $mO(\sqrt{T \ln k})$

$k = \#$ strategies

new version:

prob δ choose strategy at random

$\Rightarrow$ Prob(s choosen) $\geq \frac{\delta}{K}$

if $\boxed{u(s) \in [0,1]} \Rightarrow \bar{u}(s) \in [0, \frac{K}{\delta}]$

New alg: ~~use~~ Hedge $(1-\delta)$ each step

Prob δ: choose at random

$$\underset{\delta, H}{E}\left[\sum_{\tau=1}^{t} \bar{u}^{\tau}(s^{\tau})\right) \geq (1-\delta) \underset{H}{E}\left(\sum_{\tau=1}^{t} \bar{u}^{\tau}(\bar{s}^{\tau})\right)$$

as choosen by Hedge

$$\geq (1-\delta) E\left(\sum_{\tau=1}^{t} \bar{u}^{\tau}(s)\right) + (1-\delta)\frac{K}{\delta} O(\sqrt{T \ln K})$$

all $s \in S$

what we really want

$$\underset{alg}{E}\left[\sum_{\tau=1}^{T} u^{\tau}(s^{\tau})\right] = \underset{\substack{alg \\ \bar{u}}}{E}\left(\sum_{\tau=1}^{t} \bar{u}^{\tau}(s^{\tau})\right) > \cdots >$$

$$(1-\delta) E\left(\sum_{\tau=1}^{t} \bar{u}^{\tau}(s)\right) + \text{regret} =$$

$$(1-\delta) E\left(\sum_{\tau=1}^{t} u^{\tau}(s)\right) + \text{regret}$$

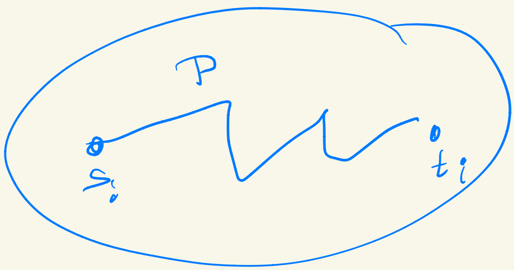
Follow perturbed leader:

Why is this better?

Hedge: needs $p_s \geq 0$ prob all $s \in S$

Follow Perturbed leader (FPL) ξ_s all $s \in S$

FPL for choosing path



idea: choose ξ_e all $e \in E$

$$P^{t-1} = \underset{\substack{P \\ s\text{-}t \text{ path}}}{\operatorname{argmin}} \left[\sum_{e \in P} \left(\sum_{\tau=1}^{t-1} c^{\tau}(e) + \xi_e \right) \right]$$

Claim: $\Pr(P^{t-1} = P^{t}) \geq (1-\varepsilon)^{n}$

$n \geq$ length of path