Recap Adversarial full-feedback setting ≡ Online learning with experts

N ~~action~~ /experts, T rounds

At each round

adversary picks $c_t(a)$ $\forall a$

algorithm picks action/expert a_t

costs of all actions/experts $c_t(a)$ $\forall a$ ≡ full feedback.

Hedge

$w_1(e) = 1$ for each expert

Sample expert e_t from distribution $p_t(e) = \frac{w_t(e)}{\sum_e w_t(e)}$

Observe costs for all experts

multiplicative $w_{t+1}(e) \leftarrow w_t(e)\,(1-\varepsilon)^{c_t(e)}$ $\quad \varepsilon \in (0, 1/2)$

$$\text{Regret(Hedge)} = O\left(\sqrt{\textstyle\sum_t c_t(e^*) \log N}\right)$$

exponential $w_{t+1}(e) \leftarrow w_t(e)\, e^{-\varepsilon c_t(e)}$

Adversarial Bandits - Multiarmed bandit with adversarially chosen costs for each arm.

Reduction to full feedback

Create expert corresponding to each action.

↳ always picks a specific action

Algorithm: At each round t

Use Hedge to draw an expert e_t from p_t ~~$\propto w_t$~~

Use e_t to pick arm a_t (TBD)

Observe cost $c_t(a_t)$ - only for a_t, NOT $\forall a$.

Define fake costs $\hat{c}_t(e)$ $\forall e$. (TBD)

Return to Hedge

More general problem:

Adversarial bandits with expert feedback — each expert can pick a different arm in different rounds.

K arms/actions, N experts, T rounds

In each round $t \in [T]$

Adversary picks cost $c_t(a) \geq 0 \quad \forall a$

Each expert recommends arm $a_{t,e}$

Algorithm picks arm $a_t \in [K]$ & receives cost $c_t(a_t)$

Total cost of each expert $\text{cost}(e) = \sum_{t=1}^{T} c_t(a_{t,e})$

Goal: Minimize regret relative to best expert — (than best arm)

$$R(T) = \text{cost(Alg)} - \min_e \text{cost}(e)$$
$$= \sum_{t=1}^{T} c_t(a_t) - \min_e \sum_t c_t(a_{t,e})$$

Note: If each expert recommends fixed arm then $N = K$

$$R(T) = \sum_t c_t(a_t) - \min_a \sum_t c_t(a) \leftarrow \text{best arm in hindsight.}$$

Assumptions: experts can't learn over time $\Rightarrow$ all $a_{t,e}$ are chosen in advance.

focus on oblivious adversary $\Rightarrow$ all costs are selected in advance.

bounded costs $c_t(a) \leq 1 \quad \forall t, a$.

Thm: Alg achieves $E[R(T)] = O(\sqrt{KT \log N})$

lower bound = $\Omega(\sqrt{KT})$

$\Omega(\min(T, \sqrt{KT \log N / \log K}))$

Proof: $\hat{R}_{\text{Hedge}}(T) = \hat{\text{cost}}(\text{Hedge}) - \min_e \hat{\text{cost}}(e)$

$R_{\text{Hedge}}(T) = \text{cost(Hedge)} - \min_e \text{cost}(e)$ $\quad \text{cost} = \sum_{t=1}^{T} c_t(\,)$

If $\star\, E[\hat{c}_t(e) \mid p_t] = c_t(e) \quad \forall e$ then $E[R_{\text{Hedge}}(T)] \leq E[\hat{R}_{\text{Hedge}}(T)]$

Proof: $E[\hat{c}_t(e) \mid p_t] = \sum_e \Pr(e_t = e \mid p_t)\, E[\hat{c}_t(e_t) \mid p_t]$

$$= \sum_e p_t(e) \; c_t(e)$$

$$= E[c_t(e) \mid p_t]$$

$$\Rightarrow E[\hat{cost}(Hedge)] = E[cost(Hedge)]$$

$$E[\min_e \hat{cost}(e)] \le \min_e E[\hat{cost}(e)] = \min_e E[cost(e)]$$

$$\Rightarrow E[R_{Hedge}(T)] \le E[\hat{R}_{Hedge}(T)] \quad .$$

(1) ★ Define fake costs.

Inverse Propensity Score

$$\hat{c}_t(a) = \begin{cases} c_t(a)/q_t(a) & a_t = a \\ 0 & \text{o.w.} \end{cases}$$

$$q_t(a) = Pr(a_t = a \mid p_t)$$

Fake cost of each expert $\hat{c}_t(e) = \hat{c}_t(a_{t,e})$

$$E[\hat{c}_t(e) \mid p_t] = E[\hat{c}_t(a_{t,e}) \mid p_t] = Pr(a_t = a_{t,e} \mid p_t) \frac{c_t(a_{t,e})}{q_t(a_{t,e})} + Pr(a_t \neq a_{t,e} \mid p_t) \cdot 0.$$

$$= c_t(a_{t,e}) = c_t(e).$$

(2) Define arm selection.

Hedge guarantee requires bounded fake costs.

$\Rightarrow q_t(a_t)$ needs to be sufficiently large.

But also want to follow e_t to ensure low costs.

w.p. $1-\gamma$ follow $e_t \Rightarrow a_t = a_{t,e_t}$

γ choose a_t randomly. $\Rightarrow q_t(a) \ge \frac{\gamma}{K} \; \forall a.$

Algorithm (EXP4) - exploration, exploitation, experts, exponentiation.

Inputs: N experts, K actions, ε for Hedge, γ for exploration.

At each round t

Call Hedge to draw an expert e_t using p_t over w_t

w.p. $1-r$ follow expert e_t, $a_t = a_{t,e_t}$

else pick a_t uniformly at random

Observe $c_t(a_t)$

Define fake costs $\forall e$

$$\hat{c}_t(e) = \begin{cases} \frac{c_t(a_t)}{\Pr(a_t = a_{t,e} | p_t)} & a_t = a_{t,e} \\ 0 & \text{o.w.} \end{cases}$$

Return fake costs to Hedge.

Note: fake costs are NOT Oblivious.

bcos depend on observed costs of choices made by Alg in past.

need adaptive adversary guarantees for Hedge

$$E[\text{cost}(\text{Exp4})] \le E[\text{cost}(\text{Hedge})] + rT.$$

$$\Rightarrow E[\text{Regret}_{\text{Exp4}}(T)] \le E[R_{\text{Hedge}}(T)] + rT$$

$$\le E[\hat{R}_{\text{Hedge}}(T)] + rT$$

Hedge bound $\sum_{t=1}^{T} \hat{c}_t(e^*) \le uT$. $\leftarrow$ $\hat{c}_t(e) \le \frac{1}{q_t(a_{t,e})} \le \frac{K}{r}$

Then $E[\hat{R}_{\text{Hedge}}(T)] \le \sqrt{uT \log N}$

$$\Rightarrow E[\text{Regret}_{\text{Exp4}}(T)] \le \sqrt{\frac{K}{r} T \log N} + rT = T^{2/3} (K \log N)^{1/3}$$

$$r^3 = \frac{K \log N}{T}$$

$$\frac{KT \log N}{r} \asymp r^2 T^2 \Rightarrow r^3 \asymp \frac{K}{T} \log N \quad rT = \frac{K^{1/3} (\log N)^{1/3} T}{T^{1/3}}$$

Hedge bound If $\sum_t E[\hat{c}_t^2(e_t) | p_t] \le uT$ $\approx$ var of costs

$$E[\hat{R}_{\text{Hedge}}(T)] \le O\left(\sqrt{uT \log N}\right) = O\left(\sqrt{KT \log N}\right)$$

$E[R_{\text{Hedge}}(T)] = O(\sqrt{N \log N}) = O(\sqrt{\quad})$ ✓

better bound # $E[\hat{c}_t^2(e_t) \mid p_t] \le \frac{K}{1-\gamma} =: u$ allow for $\gamma = 0$.

⇒ skip uniform exp step in Exp(4)

for each arm a, let $\mathcal{E}_a$ be set of experts that recommend arm a.

Let $p_t(a) = \sum_{e \in \mathcal{E}_a} p_t(e)$

$q_t(a) \ge p_t(a)(1-\gamma)$ ←

$\Rightarrow \hat{c}_t(e) = \hat{c}_t(a_{t,e}) \le \frac{c_t(a_{t,e})}{q_t(a_{t,e})} \le \frac{1}{q_t(a_{t,e})} \le \frac{1}{(1-\gamma)p_t(a_{t,e})}$ ↙

$$\sum_e \hat{c}_t^2(e_t) p_t(e) = \sum_a \sum_{e \in \mathcal{E}_a} p_t(e) \hat{c}_t(e) \hat{c}_t(e)$$

$$\le \sum_a \sum_{e \in \mathcal{E}_a} \frac{p_t(e)}{(1-\gamma)p_t(a_{t,e})} \hat{c}_t(a_{t,e}^{a})$$

$$= \frac{1}{1-\gamma} \sum_a \frac{\hat{c}_t(a)}{p_t(a)} \underbrace{\sum_{e \in \mathcal{E}_a} p_t(e)}_{p_t(a)}$$

$$= \frac{1}{1-\gamma} \sum_a \hat{c}_t(a)$$

$$E[\hat{c}_t^2(e_t) \mid p_t] = \frac{1}{1-\gamma} \sum_a E[\hat{c}_t(a) \mid p_t]$$

$$= \frac{1}{1-\gamma} \sum_a c_t(a) \le \frac{K}{1-\gamma}.$$