We progress in rounds and @ time s, an adversary picks $l_s \in \mathbb{R}^K$, learner pick $I_s \in [K]$, learner observes l_{s,I_s}.

$$R_t = \sum_{s=1}^{t} l_{s,I_s} - \min_{i \in [K]} \sum_{s=1}^{t} l_{s,i}$$

Two types of adversaries

- adaptive adversary: loss @ time t is dependent on the players actions up to time t
- oblivious adversary: Fixes all of it's losses ahead of time.

<u>Goal</u> $\mathbb{E} R_s \to$ expectation being taken over the randomness in learner and adversary.

Pseudo-Regret:

$$PR_t = \mathbb{E}\left[\sum_{s=1}^{t} l_{s,I_s} - \min_{i \in [K]} \mathbb{E}\left[\sum_{s=1}^{t} l_{s,i} \right] \right]$$

$$= \max_{i \in [K]} \mathbb{E}\left[\sum_{s=1}^{t} l_{s,I_s} - l_{s,i} \right]$$

In general: $PR_t \leq \mathbb{E} R_t$

Stochastic Setting $\to PR_t$

For oblivious adversaries: $\min_{i \in [K]} \mathbb{E}\left[\sum_{s=1}^{t} l_{s,i} \right] = \min_{i \in [K]} \sum_{s=1}^{t} l_{s,i}$

$\rightarrow \quad PR_t = \mathbb{E} R_t$

<u>Remark</u> Any deterministic learner will incur linear regret against some adversary.

$\rightarrow$ at time t the learner has to choose a distribution $q_t \in \Delta_K = \{ (x_1, \dots, x_K) : \sum_{i=1}^{K} x_i = 1, \ x_i \geq 0 \}$ and then play $I_t \sim q_t$.

@ time s you see loss ℓ_{s, I_s}

$$\forall j \in [K] \quad \hat{\ell}_{s,j} = \frac{\mathbb{1}\{ I_t = j \} \ell_{s, I_s}}{q_{s,j}}, \quad q_{s,j} \neq 0$$

$$\mathbb{E}[\hat{\ell}_{s,j}] = \mathbb{E}\left[\mathbb{E}\left[\hat{\ell}_{s,j} \mid I_1, \ell_{1,I_1}, \dots, I_{s-1}, \ell_{s-1, I_{s-1}} \right] \right]$$

$$= \mathbb{E}\left[q_{s,j} \cdot \frac{\ell_{s,j}}{q_{s,j}} + (1 - q_{s,j}) \cdot 0 \right]$$

$$= \mathbb{E}[\ell_{s,j}] \rightarrow \text{IPS estimator}$$

$\rightarrow$ Estimator for the reward of total reward for each arm

$$\hat{S}_{tj} = \sum_{s=1}^{t} \hat{\ell}_{sj}$$

$\rightarrow \mathbb{E}\left[\hat{S}_{t,j}\right] = \mathbb{E}\left[\sum_{s=1}^{t} \ell_{s,j}\right]$

Algorithm 7 EXP3

Input: $\gamma \in [0,1], \eta > 0, k$

Initialize $\lambda = (\frac{1}{k}, \cdots, \frac{1}{k})$. , $p_0 = \lambda$

for $s = 1, \cdots, t$ **do**

  Let $q_s = (1-\gamma)p_s + \gamma\lambda$

  Draw $I_s \sim q_s$ and observe loss ℓ_{s,I_s}

  Calculate the estimated total rewards for each $i \in [k]$

$$\hat{S}_{si} = \sum_{r=1}^{s} \frac{\mathbf{1}(I_r = i)\ell_{r,I_r}}{q_{ri}}$$

.

  Calculate the sampling distribution

$$p_{s+1,i} = \frac{\exp(-\eta \hat{S}_{si})}{\sum_{j=1}^{k} \exp(-\eta \hat{S}_{sj})} \text{ for } i \in [k].$$

end for

Theorem 6. *Assume $\gamma \in [0,1]$, $\eta > 0$ and for all $i \in [k]$ and $s \leq t$, we have $|\ell_{si}| \leq 1$. Then for all $i \in [k]$*

$$\mathbb{E}\left[\sum_{s=1}^{t} \ell_{s,I_s} - \ell_{s,i}\right] \leq \frac{\log(k)}{\eta} + 2\gamma t + \frac{1}{\eta}\mathbb{E}\left[\sum_{s=1}^{t}\sum_{j=1}^{k} q_{s,j}\psi(-\eta\hat{\ell}_{s,j})\right]$$

$$\downarrow \quad q_{s,j} \cdot \eta^2 \hat{\ell}_{s,j}^2$$

where $\psi(x) = e^x - 1 - x.$ $\leq x^2, \; x \leq 1$

<u>Example</u> $\gamma = \eta k, \quad \eta = \sqrt{\frac{3k\log(k)}{t}} \quad \Big| \; q_s = (1-\gamma)p_s + \gamma\lambda$

$$|\eta \hat{\ell}_{s,j}| \leq \left|\eta \frac{\ell_{s,j}}{q_{s,j}}\right| \leq \eta \frac{k}{\gamma} \leq 1$$

We also have $\phi(x) \leq x^2, \quad x \leq 1$

$$\frac{\log k}{\eta} + 2\eta k t + \frac{1}{\eta}\mathbb{E}\left[\sum_{s=1}^{t}\sum_{j=1}^{k} q_{s,j}\eta^2 \hat{\ell}_{s,j}^2\right]$$

$$\leq \frac{\log K}{\eta} + 2\eta K t + \frac{1}{\eta}\mathbb{E}\left[\sum_{s=1}^{t}\sum_{j=1}^{K}\eta^2 \hat{\ell}_{s,j}^2\right]$$

$$\leq \frac{\log K}{\eta} + 3\eta K t$$

$$\leq \sqrt{3Kt\log K} \rightarrow$$ even against oblivious adversary!

$$\mathbb{E}\,\hat{\ell}_{s,j}^2 = \mathbb{E}\,\frac{\mathbb{1}\{I_t = j\}\,\ell_{s,j}^2}{q_{s,j}^2}$$

$$= \frac{\ell_{s,j}^2}{q_{s,j}} \leq \frac{1}{q_{s,j}}$$

Proof

$$\hat{S}_t := \sum_{s=1}^{t}\sum_{j=1}^{K} q_{sj}\,\hat{\ell}_{sj}, \quad W_t = \sum_{j=1}^{K}\exp(-\eta\,\hat{S}_{tj})$$

$$\rightarrow \quad \hat{S}_t = \sum_{s=1}^{t}\ell_{s,I_s}$$

① $$\exp(-\eta\,\hat{S}_{t,i}) \leq \sum^{K} \exp(-\eta\,\hat{S}_{\cdot\cdot})$$

[illegible]

$$= W_t$$

$$= W_0 \cdot \frac{W_1}{W_0} \cdot \frac{W_2}{W_1} \cdots \frac{W_t}{W_{t-1}}$$

$$= K \prod_{s=1}^{t} \frac{W_s}{W_{s-1}}$$

$$\frac{W_s}{W_{s-1}} = \sum_{j=1}^{K} \frac{\exp(-\eta \hat{S}_{s_j})}{W_{s-1}}$$

$$= \sum_{j=1}^{K} \frac{\exp(-\eta \hat{S}_{s-1,j})}{W_{s-1}} \cdot \exp(-\eta \hat{\ell}_{s,j})$$

$\psi(x) = e^x - 1 - x$

$$= \sum_{j=1}^{K} p_{s,j} \exp(-\eta \hat{\ell}_{s,j})$$

$$= \sum_{j=1}^{K} p_{s_j} \left(\psi(-\eta \hat{\ell}_{s,j}) + 1 - \eta \hat{\ell}_{s_j} \right)$$

$1 + x \le e^x$

$$= 1 - \eta \sum_{j=1}^{K} p_{s_j} \hat{\ell}_{s_j} + \sum_{j=1}^{K} p_{s,j} \psi(-\eta \hat{\ell}_{s_j})$$

$$\le \exp\left(-\eta \sum^{K} p_{s_j} \hat{\ell}_{s_j} + \sum^{K} p_{s,j} \psi(-\eta \hat{\ell} \ldots)\right)$$

$$\ldots \Big(\sum_{j=1} p_{s_j} \hat{\ell}_{s_j} + \sum_{j=1} p_{s_j} \psi(-\eta \hat{\ell}_{s_j}) \Big)$$

-> plug into telescoping product

$$\exp(-\eta \hat{S}_{ti}) \le K \exp\left(-\eta \sum_{s=1}^{t} \sum_{j=1}^{K} p_{s,j} \hat{\ell}_{s,j} + \sum_{s=1}^{t} \sum_{j=1}^{K} p_{s_j} \psi(-\eta \hat{\ell}_{s_j}) \right)$$

$$\sum_{s=1}^{t} \sum_{j=1}^{K} p_{s_j} \hat{\ell}_{s_j} \le \hat{S}_{ti} + \frac{\log K}{\eta} + \frac{1}{\eta} \sum_{s=1}^{t} \sum_{j=1}^{K} p_{s_j} \psi(-\eta \hat{\ell}_{s_j})$$

$$q_s = (1-\gamma) p_s + \gamma \lambda \rightarrow p_{s_j} = \frac{q_{s_j} - \gamma \lambda_j}{1-\gamma}$$

$$\le \frac{q_{s_j}}{1-\gamma}$$

Mulitply both sides by $1-\gamma$, add

$$\gamma \sum_{s=1}^{t} \sum_{j=1}^{K} \frac{1}{K} \hat{\ell}_{t,j}$$

$$\sum_{s=1}^{t} \sum q_{s_j} \hat{\ell}_{s_j} - \sum_{s=1}^{t} \hat{\ell}_{s,i} \le \underbrace{-\gamma \hat{S}_{ti} + \gamma \sum_{s=1}^{t} \sum_{j=1}^{K} \lambda_j \hat{\ell}_{s_j}}_{2\gamma t} + \frac{\log(K)}{\eta}$$

$$+ \frac{1}{\eta} \sum_{s=1}^{t} \sum_{j=1}^{K} q_{s_j} \psi(-\eta \hat{\ell}_{s,j})$$

$$\mathbb{E}\left[-\gamma\, \ell_{t,i} + \gamma \sum_{s=1} \sum_{j=1} \frac{1}{K} \ell_{t,j}\right]$$

$$|\ell_{s,j}| \leq 1 \quad = \gamma \, \mathbb{E}\left[\sum_{s=1}^{t} \sum_{j=1}^{K} \frac{1}{K} \hat{\ell}_{s,j} - \frac{1}{K} \hat{\ell}_{si}\right]$$

$$\leq 2\gamma t$$