EXP3(γ): Exponential Weights for Exploration Exploitation

Input: Time horizon T, n arms, $\eta > 0$, $\gamma \in [0,1]$, $\lambda \in \triangle_n$.

Initialize: Player sets $p_1 = (1/n, \dots, 1/n) \in \triangle_n$. Adversary chooses $\{\ell_t\}_{t=1}^T \subset [-1,1]^n$.

for: $t = 1, \cdots, T$

Player draws $I_t \sim q_t := (1-\gamma)p_t + \gamma\lambda$ and suffers (and observes) loss $\ell(I_t, \ell_t) = \ell_{t,I_t}$

Player computes $\widehat{\ell}_{t,i} = \frac{\mathbf{1}\{I_t = i\}}{q_{t,i}} \ell_{t,i}$

Update iterates:

$$w_{t+1,i} = w_{t,i} \exp(-\eta \widehat{\ell}_{t,i}) \qquad p_{t+1,i} = w_{t+1,i} / \sum_{j=1}^n w_{t+1,j}.$$

Theorem 17. *Fix any sequence $\ell_t \in [-1,1]$ for all t and assume $\eta, \gamma \geq 0$ satisfy $-\eta \widehat{\ell}_{t,i} \leq 1$ for all i, t. Then we have*

$\gamma = \xi d$

$$\max_{i=1,\dots,n} \mathbb{E}\Big[\sum_{t=1}^T \ell_{t,I_t} - \ell_{t,i}\Big] \leq 2\gamma T + \frac{\log(n)}{\eta} + (1-\gamma)\eta \mathbb{E}\underbrace{\Big[\sum_{t=1}^T \sum_{j=1}^n p_{t,j} \widehat{\ell}_{t,j}^2\Big]}_{\frac{dT}{1-\gamma}}.$$

$$= 2\xi dT + \frac{\log(n)}{\xi} + \xi dT$$

$$= 3\xi dT + \frac{\log(n)}{\xi}$$

$$\leq \sqrt{12 dT \log(n)}$$

Linear Bandits - Adversarial Setting

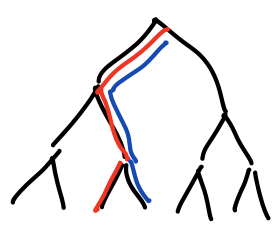
Input $\mathcal{X} = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$

Adversary chooses $\{\theta_t\}_{t=1}^T \subset \mathbb{R}^d$: $\max_{x \in \mathcal{X}} |\langle x, \theta_t \rangle| \leq 1 \quad \forall t$

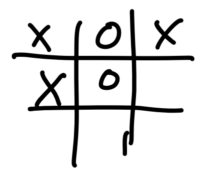
for $t = 1, 2, \dots, T$

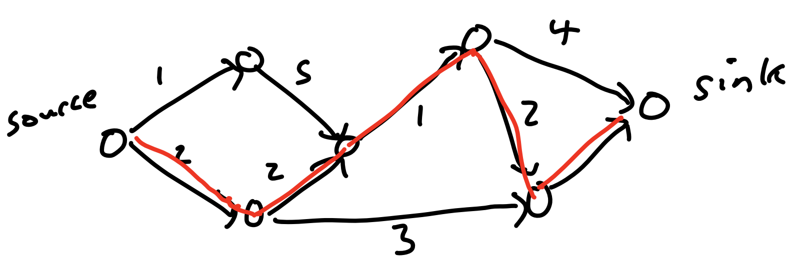
Player chooses $x_t \in \mathcal{X}$

Adversary reveals $\langle \theta_t, x_t \rangle$ (not θ_t!)



$$\text{Regret} = \max_{x \in \mathcal{X}} \sum_{t=1}^T \langle \theta_t, x_t - x \rangle$$





$x \in \{0,1\}^{|E|}$

For $x_i \in \mathcal{X}$ define $\ell_{t,i} = \langle x_i, \theta_t \rangle$

$$\hat{\ell}_{t,i} = x_i^T A(q_t)^{-1} x_{I_t} \ell_{t,I_t}, \qquad A(\lambda) = \sum_{i=1}^n \lambda_i x_i x_i^T$$

$$\begin{aligned}\mathbb{E}[A(q_t)^{-1} x_{I_t} \ell_{t,I_t}] &= \sum_{j=1}^n q_{t,j} A(q_t)^{-1} x_j \ell_{t,j} \\ &= \sum_{j=1}^n q_{t,j} A(q_t)^{-1} x_j x_j^T \theta_t \\ &= I \cdot \theta_t = \theta_t\end{aligned}$$

$$\Longrightarrow \quad \mathbb{E}[\hat{\ell}_{t,i}] = \ell_{t,i} \quad \forall i$$

Aside $\quad x_t \sim \lambda, \quad y_t = \langle x_t, \theta_* \rangle + \eta_t \qquad t = 1, \dots, T$

$$\hat{\theta}_{LS} = \left(\frac{1}{T}\sum_{t=1}^T x_t x_t^T\right)^{-1} \frac{1}{T}\sum_{t=1}^T x_t y_t$$

$$\mathbb{E}[\hat{\theta}_{LS}] = \theta_*, \quad \mathrm{Cov}(\hat{\theta}_{LS}) = \left(\sum_{t=1}^T x_t x_t^T\right)^{-1}$$

$$\hat{\theta}_{IPS} = \left(\sum_{x \in \mathcal{X}} \lambda_x x x^T\right)^{-1} \frac{1}{T}\sum_{t=1}^T x_t y_t$$

$$\mathbb{E}[\hat{\theta}_{IPS}] = \theta_*, \quad \mathrm{Cov}(\hat{\theta}_{IPS}) = \left(T \sum_{x \in \mathcal{X}} \lambda_x x x^T\right)^{-1}$$

$$\mathcal{X} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

$$\lambda_1, \quad \lambda_2 = 1 - \lambda_1$$

$$T_1 = \sum_{t=1}^{T} \mathbb{1}\{I_t = 1\}$$

$$\hat{\theta}_{LS} = \begin{bmatrix} \frac{1}{T_1} \sum_{t=1}^{T} \mathbb{1}\{I_t = 1\} y_t \\ \frac{1}{T - T_1} \sum_{t=1}^{T} \mathbb{1}\{I_t = 2\} y_t \end{bmatrix}$$

$$\hat{\theta}_{IPS} = \begin{bmatrix} \frac{1}{\lambda_1 T} \sum_{t=1}^{T} \mathbb{1}\{I_t = 1\} y_t \\ \frac{1}{(1-\lambda_1) T} \sum_{t=1}^{T} \mathbb{1}\{I_t = 1\} y_t \end{bmatrix}$$

$$\hat{\ell}_{t,i} = x_i^\top A(q_t)^{-1} x_{I_t} \ell_{t,I_t}$$

$$-\eta \hat{\ell}_{t,i} \leq \eta |x_i^\top A(q_t)^{-1} x_{I_t}| \qquad (|\ell_{t,i}| \leq 1)$$

$$= \eta \left| \langle A(q_t)^{-1/2} x_i, A(q_t)^{-1/2} x_{I_t} \rangle \right|$$

$$\leq \eta \, \|x_i\|_{A(q_t)^{-1}} \cdot \|x_{I_t}\|_{A(q_t)^{-1}}$$

$$\leq \max_{j=1,\dots,n} \eta \cdot \|x_j\|^2_{A(q_t)^{-1}}$$

$$\leq \max_{j=1,\dots,n} \frac{\eta}{\gamma} \|x_j\|^2_{A(\lambda)^{-1}}$$

$$= \frac{\eta}{\gamma} \cdot d \qquad \text{if } \lambda \text{ G-optimal.}$$

Suggests picking $\gamma = \eta d$ s.t. $-\eta \hat{\ell}_{t,i} \leq 1 \quad \forall i, t$.

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} p_{t,i}\,\hat{\ell}_{t,i}^{2}\right] = \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} p_{t,i}\, x_i^T A(q_t)^{-1} x_{I_t} x_{I_t}^T A(q_t)^{-1} x_i \,\ell_{t,I_t}^2\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} p_{t,i}\, x_i^T A(q_t)^{-1} x_{I_t} x_{I_t}^T A(q_t)^{-1} x_i\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} p_{t,i}\; x_i^T A(q_t)^{-1} x_i\right]$$

$$\leq \frac{1}{1-\gamma}\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} p_{t,i}\; x_i^T A(p_t)^{-1} x_i\right]$$

$$= \frac{1}{1-\gamma}\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} p_{t,i}\; \mathrm{Trace}\left(x_i^T A(p_t)^{-1} x_i\right)\right]$$

$$= \frac{1}{1-\gamma}\,\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i} p_{t,i}\; \mathrm{Trace}\left(A(p_t)^{-1} x_i x_i^T\right)\right]$$

$$= \frac{dT}{1-\gamma}$$

Contextual Bandits

for $t=1,2,\ldots,T$

Nature reveals context $c_t \overset{iid}{\sim} \nu$

Player chooses $I_t \in [n]$

and gets reward $y_t \in [-1,1]$, s.t. $\mathbb{E}[y_t | c_t] = r(c_t, I_t)$

Suppose support$(\nu) = \mathcal{C}$ and $|\mathcal{C}| < \infty$.

Treat each context as indep. bandit.

$$\sum_{t=1}^{T} \max_i \left(r(c_t, i) - r(c_t, I_t) \right)$$

$$= \sum_{c \in \mathcal{C}} \max_i \sum_{t=1}^{T} \mathbb{1}\{c_t = c\} \left(r(c_t, i) - r(c_t, I_t) \right)$$

$$\lesssim \sum_{c \in \mathcal{C}} \sqrt{n T_c}$$

$\leftarrow$ If playing MAB w/ $\sqrt{nT}$ regret

$T_c : \sum_t \mathbb{1}\{c_t = c\}$

$$\leq \sqrt{\left(\sum_{c \in \mathcal{C}} (\sqrt{T_c})^2 \right) \cdot \left(\sum_{c \in \mathcal{C}} (\sqrt{n})^2 \right)}$$

$$= \sqrt{n |\mathcal{C}| T}$$

Alternatively, what if we ignore context?

$$\max_i \sum_{t=1}^{T} r(c_t, i) - r(c_t, I_t) \leq \sqrt{nT}$$