Multi-armed bandits

Stochastic: Pulling arm i results in reward R.V. X_i w/ $\mathbb{E}[X_i] = \mu_i$ $\forall$ time

$$\Delta_i = \mu_{i^*} - \mu_i$$

$$\max_i \mathbb{E}\left[\sum_{t=1}^{T} \mu_i - \mu_{I_t}\right] \leq \sqrt{nT\log(nT)} \wedge \sum_{i \neq i^*} \Delta_i^{-1}\log(nT)$$

Adversarial setting: Adversary chooses $l_t \in [-1,1]^n$ $\forall t$ prior to start of game.

Pulling arm i results in loss $l_{t,i}$

$$\max_i \mathbb{E}\left[\sum_{t=1}^{T} l_{t,I_t} - l_{t,i}\right] \leq \sqrt{nT\log(nT)}$$

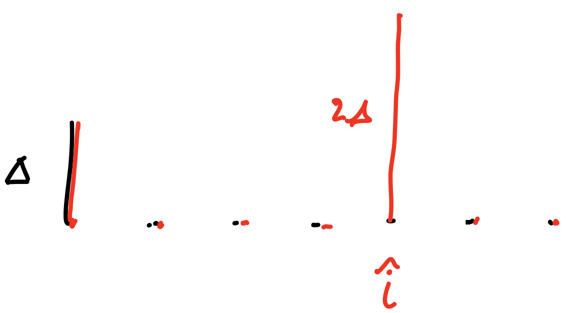
<u>Thm</u> For every $T \geq n$ $\exists$ set of bandit instances w/ $P_\theta = \mathcal{N}(\theta, I)$ s.t. $\sup_{P_\theta} \mathbb{E}_\theta[\text{Regret}] \geq \sqrt{(n-1)T/256}$.

$\theta = (\Delta, 0, \ldots, 0) \in [0,1]^n$

$\theta' = (\Delta, 0, \ldots, 2\Delta, \ldots 0)$

$\exists \hat{i} \in \{2, \ldots, n\}$ s.t. $\mathbb{E}_\theta[T_{\hat{i}}] \leq \frac{T}{n-1}$

$\Rightarrow$ intuitively, we can't tell whether $\theta_{\hat{i}} = 0$ or $\theta_{\hat{i}} \approx \sqrt{\frac{n-1}{T}} \approx \Delta$



$$\text{Regret}(\theta') = \sum_i \Delta_i' \mathbb{E}_{\theta'}[T_i] \geq \Delta_1' \mathbb{E}_{\theta'}[T_1] \geq \Delta \, \mathbb{P}_{\theta'}(T_1 \geq T/2)\, T/2$$

$$\text{Regret}(\theta) = \sum_{i\neq 1} \Delta_i \mathbb{E}_\theta[T_i] = \Delta \sum_{i \neq 1} \mathbb{E}_\theta[T_i]$$

$$\geq \Delta\, \mathbb{P}_\theta\Big(\sum_{i\neq 1} T_i \geq T/2\Big)\, T/2$$

$$= \Delta\Big(1 - \mathbb{P}_\theta(T_1 \geq T/2)\Big)\, T/2$$

$$\max_{\mu \in (\theta, \theta')} \text{Regret}(\mu) \geq \frac{1}{2}\Big(\text{Regret}(\theta) + \text{Regret}(\theta')\Big)$$

$$\geq \frac{T\Delta}{4}\Big(1 + \mathbb{P}_{\theta'}(T_1 \geq T/2) - \mathbb{P}_\theta(T_1 \geq T/2)\Big)$$

$$\geq \frac{\Delta T}{4}\Big(1 - \sup_A |\mathbb{P}_{\theta'}(A) - \mathbb{P}_\theta(A)|\Big)$$

(Pinsker's ineq.) $$\geq \frac{\Delta T}{4}\Big(1 - \sqrt{KL(\mathbb{P}_\theta \,\|\, \mathbb{P}_{\theta'})/2}\Big)$$

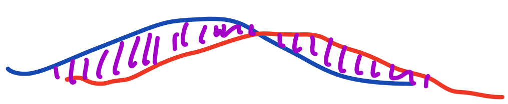
$$= \frac{\Delta T}{4}\Big(1 - \sqrt{\sum_i (\theta_i - \theta_i')^2 \mathbb{E}_\theta[T_i]/4}\Big)$$

$$= \frac{\Delta T}{4}\Big(1 - \sqrt{\Delta^2 \mathbb{E}_\theta[T_i]}\Big)$$

$$\geq \frac{\Delta T}{4}\Big(1 - \sqrt{\Delta^2 T/(n-1)}\Big)$$

$$\Delta = \sqrt{\frac{n-1}{4T}}$$

$\mathbb{P}_{\theta'}, \mathbb{P}_\theta$ defined over 1d



$$\sup_A |\mathbb{P}_{\theta'}(A) - \mathbb{P}_\theta(A)| = \int_x \left| \frac{d\mathbb{P}_\theta(x)}{dx} - \frac{d\mathbb{P}_{\theta'}(x)}{dx} \right| dx$$

$$\leq \int_x \frac{d\mathbb{P}_\theta}{dx} \log\left(\frac{\frac{d\mathbb{P}_\theta(x)}{dx}}{\frac{d\mathbb{P}_{\theta'}(x)}{dx}}\right) dx$$

$$= \mathbb{KL}(\mathbb{P}_\theta \,||\, \mathbb{P}_{\theta'})$$

$\boxtimes$

Consider a game where at each time t row player chooses $I_t \in [m]$, column player chooses $J_t \in [n]$ and row player receives reward $e_{I_t}^T A e_{J_t} + \eta_t$ where A is unknown and $\mathbb{E}[\eta_t] = 0$, $|\eta_t| \leq 1$.

row-player's objective maximize $\mathbb{E}\left[\sum_t e_{I_t} A e_{J_t}\right]$

$$\ell_t = -Ae_{J_t}$$

$$V_A = \max_{x \in \Delta_m} \min_{y \in \Delta_n} x^T A y$$ is the value of the game.

$(x, y) \in \Delta_m \times \Delta_n$ is a ε-Nash equilibrium if they achieve the value of the game. Equivalently:

$$(x - x')^T A y \geq -\varepsilon \ \ \forall x' \quad \text{and} \quad x^T A (y' - y) \geq -\varepsilon \ \ \forall y'$$

$$\text{Nash-Regret} = \mathbb{E}\left[\sum_{t=1}^{T}\left(V_* - e_{I_t}^T A e_{J_t}\right)\right] \qquad I_t \sim x_t, \quad J_t \sim y_t$$

$$= \mathbb{E}\left[\sum_{t=1}^{T} x_*^T A y_* - x_t^T A y_t\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{T} x_*^T A y_t - x_t^T A y_t\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{T} (x_* - x_t)^T A y_t\right] = \text{External regret.}$$

<u>Claim</u> $\exists$ A matrix s.t. for any row-player that achieves $o(T)$ external regret, $\exists$ adversary s.t. external regret $\geq c\sqrt{mT}$ and $\mathbb{E}\left[\sum_{t=1}^{T} x_t^T A y_t - x_*^T A y_*\right] \geq c' T$

Let row-player and column player each play a independent copy of EXP-3. and let $\hat{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t$, $\hat{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t$

$$(x - \hat{x})^T A \hat{y} \leq c\sqrt{\frac{m \log(T)}{T}} \qquad \hat{x}^T A (y - \hat{y}) \geq c\sqrt{\frac{n \log(T)}{T}}$$

$\Rightarrow$ if $T \geq c'(m+n)\,\varepsilon^{-2}\log(1/\varepsilon)$ then $(\hat{x}, \hat{y})$ is an ε-Nash Eq.