n alternatives / treatments / arms

Input: n arms described by distributions ν_i, $i=1,\dots,n$

for $t = 1, 2, \dots, T$ ← user t arrives at nytimes.com

Player chooses $I_t \in [n] := \{1,\dots,n\}$

Nature reveals $X_{I_t,t} \overset{iid}{\sim} \nu_{I_t}$

Example: $\nu_i = \mathcal{N}(\theta^*_i, 1)$ $X_{i,t} \in \mathbb{R}$ $\forall t$

Example: $\nu_i = \text{Bernoulli}(\theta^*_i)$ $X_{I_t,t} \in \{0,1\}$ $\forall t$

We do not know $\{\nu_i\}_{i=1}^n$ but we typically assume some property. For example $\text{support}(\nu_i) \in [0,1]$, ν_i is 1-sub-Gaussian.

Regret Minimization / Reward Accumulation Maximization

Goal: maximize $\mathbb{E}\left[\sum_{t=1}^{T} X_{I_t,t}\right]$

<u>Def</u> Regret $= \max_{i=1,\dots,n} \mathbb{E}\left[\underbrace{\sum_{t=1}^{T} X_{i,t}}_{\text{Best action in hindsight}} - \underbrace{\sum_{t=1}^{T} X_{I_t,t}}_{\text{What player recieves}}\right]$

Goal: minimize regret.

Obtain sub-linear regret:

$\text{Regret}(T) = o(T)$ (e.g. $R_T = O(\sqrt{T})$)

Proposition $R_T := \max_{i=1,\dots,n} \mathbb{E}\left[\sum_{t=1}^{T} X_{i,t} - \sum_{t=1}^{T} X_{I_t,t}\right]$

$$= \sum_{i=1}^{n} \Delta_i \, \mathbb{E}[T_i]$$

where $\Delta_i := \max_{j=1,\dots,n} \theta_j^* - \theta_i^*$, T_i is # times arm i is played up to time T. $\theta_i^* = \mathbb{E}_{X\sim\nu_i}[X]$.

Proof

$$\max_{i=1,\dots,n} \mathbb{E}\left[\sum_{t=1}^{T} X_{i,t} - \sum_{t=1}^{T} X_{I_t,t}\right]$$

$$= \max_i \sum_{t=1}^{T} \theta_i^* - \sum_{t=1}^{T} \mathbb{E}[X_{I_t,t}]$$

$$= T \cdot \max_i \theta_i^* - \sum_{t=1}^{T} \mathbb{E}\Big[\underbrace{\mathbb{E}[X_{I_t,t} \mid I_t = i]}_{\theta^*_{I_t}}\Big]$$

$$= T \cdot \max_i \theta_i^* - \sum_{t=1}^{T} \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{1}\{I_t = i\}\, \theta_i^*\right]$$

$$= T \cdot \max_i \theta_i^* - \sum_{i=1}^{n} \theta_i^* \underbrace{\sum_{t=1}^{T} \mathbb{E}[\mathbb{1}\{I_t = i\}]}$$

$$= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\{I_t = i\}\right]$$

$$= \mathbb{E}[T_i]$$

$$= \underbrace{T}_{= \sum_{i=1}^{n} \mathbb{E}[T_i]} \cdot \max_i \theta_i^* - \sum_{i=1}^{n} \theta_i^* \, \mathbb{E}[T_i]$$

$$= \sum_{i=1}^{n} \underbrace{\left(\max_j \theta_j^* - \theta_i^* \right)}_{=\Delta_i} \mathbb{E}[T_i]$$

Best arm identification

(adaptive A/B/n testing)

Input confidence $\delta \in (0,1)$.

Objective is for an algorithm to pull arms and stop ASAP and output $\hat{i} \in [n]$ where $\mathbb{P}\left(\hat{i} = \arg\max_{i \in \{1,\dots,n\}} \theta_i^* \right) \geq 1 - \delta$.

(ε, δ)- PAC identification

Identify an arm $\hat{i}$: $\max_j \theta_j^* - \theta_{\hat{i}}^* \leq \varepsilon$

Top-k : identify top k means

Multiple testing / threshold bandits

Combinatorial bandits

$c \in \mathcal{C}$ $\quad c \subset [n]$

$$\operatorname*{argmax}_{c \in \mathcal{C}} \sum_{i \in c} \theta_i^*$$

Top-k
Threshold bandits
Matching
Routing

Max-bandits Pull k arms $|S| = k$ and observe $\max_{i \in S} X_{i,t}$