# Week 6 - Lecture 30

## Stochastic Bandits

$\nu = (\nu_a)_{a \in [K]}$ ← Bandit Instance

$X \sim \nu_a$

$\mathbb{E}_{X \sim \nu_a}[X] = \mu_a$

Goal: Identify the arm with largest $\mu_a$, quickly.

$$\bar{R}(\pi, n) = (\max_a \mu_a) n - \mathbb{E}\left[\sum_{t=1}^{n} \mu_{I_t}\right]$$

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

where $\Delta_i = (\max_a \mu_a) - \mu_i$

We will assume that arm 1 is optimal, algorithm doesn't know this. ↓ For ease in calculations

$\therefore \Delta_1 = 0$

$\Delta_2 = +ve$

$\Delta_3 = +ve$

Also we will assume that optimal arm is unique

$\therefore \mu_1 > \mu_i \;\; \forall i \neq 1$

## Explore then Commit (ETC) Algorithm

Play all arms to get the estimates of the arms and then play the arm with highest mean.

But key is, how many rounds to explore

Input : $m$ – no. of rounds to explore each time
$K$ – no. of arms

For $t = 1, 2, \dots$

$$I_t = \begin{cases} t \bmod K + 1 & \text{if } t \le mk \\ \arg\max_i \hat{\mu}_i(mk) & \text{if } t > mk \end{cases}$$

← (play arms in a round robin fashion)

$$\hat{\mu}_i(t) = \frac{\sum_{s=1}^{t} 1\{I_s = i\} X_{I_s}}{\sum_{s=1}^{t} 1\{I_s = i\}}$$

← (rewards of arm $i$ till round $t$)

← (counting the number of plays of arm $i$ till round $t$)

---

$$T_i(n) = m + \sum_{s=mk+1}^{n} 1\{I_s = i\}$$

$$T_i(n) = m + (n - mk)\, 1\left\{ \hat{\mu}_i(mk) \ge \max_{j \ne i} \hat{\mu}_j(mk) \right\}$$

$$\mathbb{E}\left[T_i(n)\right] \leq m + (n-mk) \times$$

$$\Pr\left\{ \hat{\mu}_i(mk) \geq \max_{j \neq i} \hat{\mu}_j(mk) \right\}$$

Now need to figure this

$$\Pr\left\{ \hat{\mu}_i(mk) \geq \max_{j \neq i} \mu_j(mk) \right\} \text{ — } (A)$$

[Assume $i \neq 1$, i.e. it is not the optimal arm]

Thus, [an event that shouldn't happen, but checking now what is the probability of that happening

$$\Pr\left\{ \hat{\mu}_i(mk) \geq \hat{\mu}_1(mk) \right\} \text{ — } (B)$$

$\hat{\mu}_1(mk) < \max_{j \neq i} \mu_j(mk)$, $\hat{\mu}_1$ is contained in $\max \mu_j(mk)$

thus A is more stringent than B $\therefore$

$$\Pr\left\{ \hat{\mu}_i(mk) \geq \max_{j \neq i} \mu_j(mk) \right\} \leq \Pr\left\{ \hat{\mu}_i(mk) \geq \hat{\mu}_1(mk) \right\}$$

---

$$= \Pr\left\{ \hat{\mu}_i(mk) - \mu_i \, ( \hat{\mu}_1(mk) - \mu_1 ) \geq \mu_1 - \mu_i \right\}$$

$$\Rightarrow \Pr\left\{ \hat{\mu}_i(mk) \geq \max_{j \neq i} \hat{\mu}_j(mk) \right\}$$

$$\leq \Pr\left\{ \hat{\mu}_i(mk) - \mu_i - ( \hat{\mu}_1(mk) - \mu_1 ) \geq \Delta_i \right\}$$

[ $\hat{\mu}_i(mk)$ is mean of arm $i$ played $m$ times;

$$\therefore \hat{\mu}_i(mK) = \frac{\sum_{s=1}^{m} X_{is}}{m}$$ ← all samples from arm $i$, all samples are iid

$$\therefore \hat{\mu}_i(mK) - \mu_i = \frac{\sum_{s=1}^{m} X_{is}}{m} - \mu_i$$

$$= \frac{\sum_{s=1}^{m} (X_{is} - \mu_i)}{m}$$ ← Each quantity is 1-subgaussian

$= \frac{1}{\sqrt{m}}$ -subgaussian

$$\left[\begin{aligned} &\sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3 + \dots \sigma_m^2} \\ &= \sqrt{\frac{1}{m^2} + \frac{1}{m^2} + \dots \frac{1}{m^2}} \\ &= \frac{1}{\sqrt{m}} \end{aligned}\right]$$

$(\hat{\mu}_1(mK) - \mu_1)$ will also be $\frac{1}{\sqrt{m}}$ -subgaussian

as all arms $- \mu$ is 1-subgaussian

Thus:

$$\Pr\{ \underbrace{\hat{\mu}_i(mK) - \mu_i}_{\frac{1}{\sqrt{m}}\text{-subgaussian}} - (\underbrace{\hat{\mu}_1(mK) - \mu_1}_{\frac{1}{\sqrt{m}}\text{-subgaussian}}) \geq \Delta_i \}$$

Also these two are independent

$$\therefore \Pr\{ \underbrace{\hat{\mu}_i(mK) - \mu_i - (\hat{\mu}_1(mK) - \mu_1)}_{\sqrt{\frac{2}{m}}\text{-subgaussian}} \geq \Delta_i \}$$

**Lemma of Subgaussian :** inputting)

$$\Pr\{X \geq \varepsilon\} \leq \exp\{-\varepsilon^2/2\sigma^2\}$$

$$\therefore \Pr\{\hat{\mu}_i(mK) - \mu_i - (\hat{\mu}_1(mK) - \mu_1) \geq \Delta_i\}$$

$$\leq \exp\left\{\frac{-\Delta_i^2 m}{4}\right\}$$

Thus :

$$\mathbb{E}[T_i(n)] \leq m + (n - mk)\exp\left\{\frac{-\Delta_i^2 m}{4}\right\}$$

$$\therefore R(ETC, n) = \sum_{i=1}^{K} \mathbb{E}[T_i(n)]\Delta_i$$

$$= m\sum_{i=1}^{K}\Delta_i + (n - mK)\sum_{i=1}^{K}\Delta_i \exp\left\{\frac{-\Delta_i^2 m}{4}\right\}$$

---