Week 6 - Lecture 29

R.V.s. that satisfy subgaussianity

Example:

1) If X has zero mean and $|X| \leq B$ (almost surely) for some B, then X is B-subgaussian

2) If X has zero mean and $X \in [a, b]$ almost surely, then X is $(b-a)/2$-subgaussian

Henceforth focus will be on subgaussian r.v.s.

$$\therefore \quad \mathcal{E} = \left\{ \nu = (\nu_a)_{a \in [K]} : \left(\nu_a - \mathbb{E}_{X \sim \nu_a}[X]\right) \text{ is } \sigma\text{-subgaussian } \forall a \in [K] \right\}$$

Environment class

Stochastic Bandits

Input : No. of arms K

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

learner selects $I_t \in [K]$

Environment $X_{I_t t} \sim \nu_{I_t}$

Policy $\pi_t(\cdot \mid X_{I_1 1}, I_1, X_{I_2 2}, I_2, \dots\dots X_{I_{t-1} t-1}, I_{t-1}$

$\pi = (\pi_1, \pi_2 \dots\dots \pi_n)$

$$\text{Regret}(\pi, n) = \mathbb{E}\left[\max_i \sum_{t=1}^{n} X_{it} - \sum_{t=1}^{n} X_{I_t t}\right]$$

$$= \underbrace{\mathbb{E}\left[\max_i \sum_{i=1}^{n} X_{it}\right]}_{\text{This is harder to deal with, thus will relax it}} - \mathbb{E}\left[\sum_{t=1}^{n} X_{I_t t}\right]$$

I_t has randomness and X_{I_t} also has randomness

$$\geq \max_i \mathbb{E}\left[\sum_{t=1}^{n} X_{it}\right] - \mathbb{E}\left[\sum_{t=1}^{n} \mu_{I_t t}\right]$$

$$= \max_i \sum_{t=1}^{n} \mu_i - \mathbb{E}\left[\sum_{t=1}^{n} \mu_{I_t}\right]$$

$$\boxed{\nu_i \text{ has a mean } \mu_i \quad \therefore X_{it} \sim \nu_i}$$

$$= \max_i n\mu_i - \mathbb{E}\left[\sum_{t=1}^{n} \mu_{I_t}\right]$$

Let $\mu^* = \max_i \mu_i$

$$= n\mu^* - \mathbb{E}\left[\sum_{t=1}^{n} \mu_{I_t}\right]$$

$= \bar{R}(\pi, n) \longrightarrow$ Pseudo-regret is a lower bound on actual regret

$$R(\pi, n) \geq \bar{R}(\pi, n)$$

We will upper bound $\bar{R}(\pi, n)$, but it will not be an upper bound on $R(\pi, n)$; but we will ignore this

We are interested in making this learnable, which mea

$$\frac{\bar{R}(\pi, n)}{n} \longrightarrow 0$$

Decomposition of regret

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

μ_{I_t} depends on π

Defining no. of pulls :

$$T_i(n) = \sum_{t=1}^{n} \mathbb{1}\{I_t = i\}$$

$T_i(n)$ is a random variable

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

$$\therefore \bar{R}(\pi, n) = \sum_{i=1}^{K} \mathbb{E}[T_i(n)] \mu^* - \mathbb{E}\left[\sum_{t=1}^{n} \sum_{i=1}^{K} \mu_i \mathbb{1}\{I_t = i\}\right]$$

$$= \sum_{i=1}^{K} \mathbb{E}\left[T_i(n)\right]\mu^* - \mathbb{E}\left[\sum_{i=1}^{K}\sum_{t=1}^{n} \mu_i \mathbb{1}_{\{I_t = i\}}\right]$$

$$= \sum_{i=1}^{K} \mathbb{E}\left[T_i(n)\right]\mu^* - \mathbb{E}\left[\sum_{i=1}^{K} \mu_i T_i(n)\right]$$

$$= \sum_{i=1}^{K} \mathbb{E}\left[T_i(n)\right]\mu^* - \sum_{i=1}^{K} \mu_i \, \mathbb{E}\left[T_i(n)\right]$$

$$= \sum_{i=1}^{K} \mathbb{E}\left[T_i(n)\right]\left(\mu^* - \mu_i\right)$$

Let $\mu^* - \mu_i = \Delta_i$ $\leftarrow$ gap between best arm and ith arm

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

Regret decomposition formula

If i is optimal then no Δ_i. But for sub-optimal arm Δ_i will have a value. This formula calculates the gap for all arms w.r.t. the optimal arm, and the gap with optimal arm would be zero.

Let $i^* = \text{argmax}\, \mu_i$

So our policy should be such that T_i for i^* should be higher than T_i for sub-optimal arms.

The bound for $\bar{R}(\pi, n)$ would thus be decided by no. of pulls of the sub-optimal arms.

Thus we need a policy that will help us identify the arm with the highest mean quickly.