Week 3

Lecture 13

The expected regret can also be :

$$\mathbb{E}\left[\sum_{t=1}^{n} x_{tI_t}\right] - \underbrace{\mathbb{E}\left[\sum_{t=1}^{n} \min_i x_{ti}\right]}_{} \quad \text{——— (2a)}$$

Here the benchmark is that we choose the best action in each round

$$\mathbb{E}\left[\sum_{t=1}^{n} x_{tI_t}\right] - \underbrace{\min_i \mathbb{E}\left[\sum_{t=1}^{n} x_{ti}\right]}_{} \quad \text{——— (2)}$$

Here the benchmark is the <u>one</u> arm which if played for all rounds will give min loss

Though eq 2a is a better criterion, but it is harder to achieve and hence expected regret is defined by eq

As long as x_t are generated randomly the regret generated by eq 2a $\geq$ regret by eq 2

This is because :

$$\mathbb{E}\left[\sum_{t=1}^{n} \min_i x_{ti}\right] \leq \min_i \mathbb{E}\left[\sum_{t=1}^{n} x_{ti}\right]$$

If x_t are generated according to some sequence then eq 2 becomes :

$$R(n,\pi) = \mathbb{E}\left[\sum_{t=1}^{n} x_{tI_t}\right] - \min_i \left[\sum_{t=1} x_{ti}\right] \quad \text{—— (2b)}$$

In eq 2, the first term has 2 levels of randomness : → randomness of loss by the adversar

→ —— do —— arm selection by learner

$$\mathbb{E}\left[R(n,\pi)\right] = \mathbb{E}\left[\sum_{t=1}^{n} x_{tI_t}\right] - \min_i \mathbb{E}\left[\sum_{t=1}^{n} x_{ti}\right. \quad \text{—— (2)}$$

↑

Psuedo regret, also known as

The learner chooses the next action based on history, hence the expected regret should ideally be over a particular sequence.

$$\therefore \mathbb{E}\left[R(n,\pi,\{x_t\})\right] = \mathbb{E}\left[\sum_{t=1}^{n} x_{tI_t}\right] - \min_i \left[\sum_{t=1}^{n} x_{ti}\right] \quad \text{—— (2c)}$$

Assuming $\{X_t\}$ to be stochastic, $\mathbb{E}$ have to be on both the terms on the R.H.S. of eq 2...

Hence revisiting the definitions again :

Expected Regret :

$$\mathbb{E}\left[R(n,\pi)\right] = \mathbb{E}\left[\sum_{t=1}^{n} X_{tI_t}\right] - \mathbb{E}\left[\sum_{t=1}^{n} \min_i X_{ti}\right]$$

$$\geq \mathbb{E}\left[\sum_{t=1}^{n} X_{tI_t}\right] - \min_i \mathbb{E}\left[\sum_{t=1}^{n} X_{ti}\right]$$

↑

Psuedo regret

Whatever has been discussed till now about bandits is the <u>Adversarial Bandi setting</u> [starting from Lec 12]

<u>Importance Sampling</u>

At round t

$P_t = (P_{t1}, P_{t2}, \ldots P_{tk})$

Learner chooses $I_t \sim P_t$

Let $I_t = i$,

Learner suffers loss x_{ti}

At round t we can define an estimator for all $\{x_t\}$

Importance weighted estimator :

$$\forall j \in [K] \quad \hat{x}_{tj} = \frac{x_{tj}}{P_{tj}} \mathbb{1}\{I_t = j\} = \begin{cases} \frac{x_{ti}}{P_{ti}} & \text{for } j = i \\ 0 & \text{for } j \neq i \end{cases}$$

$$\mathbb{E}[\hat{x}_{tj}] = \sum_{i=1}^{K} \frac{x_{tj}}{P_{tj}} \mathbb{1}\{i = j\} \times P_{ti}$$

$$= \frac{x_{tj}}{P_{tj}} \times P_{tj}$$

$$= x_{tj}$$

Hence the importance weighted estimator is an unbiased estimator of the losses.

Hence though we don't have information for all the arms/actions, but we have unbiased estimaters for all losses, which we can use to update the weights.