

Week 5 - Lecture 26

Stochastic Bandits

Assume that the losses that the environment assigns, it generates from a fixed distribution.

Input : No. of arms, no. of rounds n [horizon may or may not be specified]

In each round $t = 1, 2, \dots, n$

Learner selects an arm $[a \in [K]]$

Environment assigns a loss to arms drawn from a distribution (associated with arm)

Let $K = \text{no. of arms}$, unknown

$(V_a)_{a \in [K]}$ $\therefore \text{loss/reward} = X \sim V_a$, $\mu_a = \text{mean of } V_a$

V_a distribution associated with arm a

$t \in \{1, 2, \dots, n\}$

Learner chooses $I_t \in [K]$

Earn reward $X_{I_t} \sim V_{I_t}$

Policy of the learner [how to select arm at round t]

$$\pi_t \left(\cdot \mid \underbrace{I_1, X_1, I_2, X_2, \dots, I_{t-1}, X_{t-1}}_{\mathcal{H}_t = \text{history}} \right)$$

$\pi = \{\pi_1, \pi_2, \pi_3, \dots\}$ constitutes a policy

Total Reward:

$$S_n = \sum_{t=1}^n X_{I_t}$$

Goal: maximize $\mathbb{E}[S_n]$ \leftarrow Expectation over two random quantities

Now let: $\mu^* = \max_{a \in [K]} \mu_a$ I_t and randomness in samples itself: X_{I_t}

$$\therefore \mathbb{E}[S_n] = n\mu^* \leftarrow \text{Benchmark}$$

Regret:

$$\text{Regret}(\pi, n) = n\mu^* - \mathbb{E}[S_n]$$

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

$$= n \max_{i \in [K]} \mathbb{E}[X_i] - \mathbb{E}\left[\sum_{t=1}^n X_{I_t}\right]$$

\uparrow

playing a single arm which gives highest reward

$(\mu_a)_{a \in [K]}$ \rightarrow the distributions are unknown

But they come from an environment class, and once fixed will define the Bandit Instance.

Thus these Bandit instances can be said to be drawn from an environment.

$$\begin{array}{ccc} \mathcal{E} = \left\{ \overset{\text{Bandit Instance}}{V} = (V_a \mid a \in [K]) ; V_a \in \mathcal{M}_a \forall a \in [K] \right\} \\ \uparrow \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \downarrow \\ \text{environment class} \qquad \qquad \qquad \qquad \qquad \qquad \text{set of distributions} \end{array}$$

Typical Environment Classes

$$1) \mathcal{E} = \left\{ (\text{Ber}(\mu_i))_i : \mu_i \in [0, 1], \forall i \right\}$$

Example: $K = 4$ arms

$$\text{Then: } V = \left\{ \text{Ber}(0.5), \text{Ber}(0.3), \text{Ber}(0.2), \text{Ber}(0.6) \right\}$$

$$\text{or} \\ V = \left\{ \text{Ber}(0.2), \text{Ber}(0.3), \text{Ber}(0.4), \text{Ber}(0.9) \right\}$$

2) Uniform distributions

$$\mathcal{E} = \left\{ (U(a_i, b_i))_i, a_i, b_i \in \mathbb{R}, a_i \leq b_i \forall i \right\}$$

3) Gaussian

$$\mathcal{E} = \left\{ (N(\mu_i, \sigma_i^2))_i, \mu_i \in \mathbb{R}, \sigma_i^2 \in \mathbb{R}, \forall i \right\}$$

4) Finite Variance

$$\mathcal{E} = \left\{ (V_i)_{i \in [K]} : \underset{\substack{\uparrow \\ \text{variance}}}{V_{x \sim V_i}}[x] \leq \sigma^2 \quad \forall i \right\}$$

Finite Variance Gaussian is a subset of Finite Variance

$$\mathcal{E} = \left\{ (N(\mu_i, \sigma^2))_i, \mu_i \in \mathbb{R}, \quad \forall i \right\}$$

5) Bounded Support

$$\mathcal{E} = \left\{ (V_i)_{i \in [K]}, \text{Supp}(V_i) \subseteq [a, b] \right\}$$

If $[a, b] = [0, 1]$ then Bernoulli can be considered a subset of Bounded Support

6) Sub-Gaussian

$$\mathcal{E} = \left\{ (V_i)_{i \in [K]}, V_i \text{ is } \sigma^2\text{-sub Gaussian } \forall i \right\}$$

This class is a more ^{of a} generalization of Gaussian random variable. This class includes the Finite Variance