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 b specified

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

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

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

Let K = no. of arms

$(\nu_a)_{a \in [K]} \xrightarrow{\text{unknown}}$ $\therefore$ loss/reward $= X \sim \nu_a$, μ_a = mean of ν_a

ν_a distribution associated with arm a

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

Learner chooses $I_t \in [K]$

Earns reward $X_{I_t} \sim \nu_{I_t}$

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

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

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

Total Reward:

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

Goal: maximize $\mathbb{E}[S_n]$ ⟵ Expectation over two random quantit[ies] I_t and randomness in samples itself: X_{I_t}

Now Let: $\mu^* = \max_{a \in [K]} \mu_a$

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

$\therefore$ Regret:

$$\begin{aligned}
\text{Regret}(\pi, n) &= n\mu^* - \mathbb{E}[S_n] \\
&= n\mu^* - \mathbb{E}\left[\sum_{t=1}^{n} X_{I_t}\right] \\
&= n \max_{i \in [K]} \mathbb{E}[X_i] - \mathbb{E}\left[\sum_{t=1}^{n} X_{I_t}\right]
\end{aligned}$$

↑ playing a single arm which gives highest reward

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

... from an environment class, and once

But they come ...

fixed will define the Bandit Instance.

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

$$\underset{\substack{\uparrow \\ \text{environment class}}}{\mathcal{E}} = \{ \overset{\text{Bandit Instance}}{\nu} = (\nu_a \; a \in [K]) \; ; \; \nu_a \in \underset{\substack{\downarrow \\ \text{set of distributions}}}{\mathcal{M}_a} \; \forall \; a \in [K] \}$$

Typical Environment Classes

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

Example: $K = 4$ arms

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

or

$\nu = \{ \text{Ber}(0.2), \text{Ber}(0.3), \text{Ber}(0.4), \text{Ber}(0.9) \}$

2) Uniform distributions

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

3) Gaussian

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

4) Finite Variance

$$\mathcal{E} = \left\{ (V_i)_{i \in [K]} : \mathbb{V}_{x \sim V_c}[x] \leq \sigma^2 \quad \forall i \right\}$$

variance

Finite Variance Gaussian is a subset of Finite Variance

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

5) Bounded Support

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

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 Varian