EE6106 Lecture 11 (Date : 5th March 2024)

EE6106 $\frac{2n\Delta^2}{\cdot -1} = 0$

Recall

- $P, Q \sim$ measures on $(\Omega, \mathcal{F})$.
  $P(A) + Q(A^c) \geq \frac{1}{2}\exp(-D(P,Q))$

Moral : If P and Q are "close", then $P(A)+Q(A^c)$ cannot be small

- $D(P_\nu, P_{\nu'}) = \sum_i D(\nu_i, \nu_i') E_\nu[T_i(n)]$

- $\mathcal{E} \sim$ set of arm distributions
  $\Rightarrow \mathcal{E}^k \sim$ set of MAB instances

$$R_n(\pi, \mathcal{E}^k) = \sup_{\nu \in \mathcal{E}^k} R_n(\pi, \nu)$$

$\downarrow$ worst case regret of π on $\mathcal{E}^k$

- $R_n^*(\mathcal{E}^k) = \inf_\pi R_n(\pi, \mathcal{E}^k)$ ... most robust worst case policy

$\hookrightarrow$ minimax regret

$$\therefore R_n^*(\mathcal{E}^k) = \inf_\pi \sup_{\nu \in \mathcal{E}^k} R_n(\pi, \nu)$$

, Lower bound on minimax regret : means that any policy in the worst case will have at least as much regret

THEOREM Let $K > 1$, $n \geq K-1$. Then for any policy π, $\exists\, \mu \in \mathcal{E}_{\mathcal{N}}^k(1)$ s.t. $\mu_i \in [0,1]\ \forall i$, s.t.

$$R_n(\pi, \mu) \geq \frac{1}{27}\sqrt{n(K-1)}$$

$\mathcal{E}_{\mathcal{N}}^k(1)$: Gaussian distributions with variance 1 for all arms' rewards

* Effectively, this theorem gives a lower bound on the minimax regret

Note : When we did UCB, there was a $\sqrt{nk\log k}$ upper bound that we had got

Proof

Fix policy π, let $\Delta \in [0, 1/2]$ ... open interval

Consider $\mu = (\Delta, 0, \ldots, 0)$

Let
$$i = \arg\min_{j>1} E_\mu[T_j(n)]$$

Note : i is the arm which on average gets pulled the least.

Construct a new arm μ'

$$\mu' = (\Delta, 0, \ldots, 0, 2\Delta, 0, \ldots, 0)$$

i^{th} arm

Note : $R_n(\pi,\mu) \geq \underbrace{\mathbb{P}_\mu\left(T_1(n) \leq \frac{n}{2}\right)}_{A} \frac{n\Delta}{2}$

This is a bad event since we are pulling the optimal arm less than $n/2$ times. Idea: Lower bound regret in terms of a bad/rare event

Similarly,

$$R_n(\pi,\mu') \geq \underbrace{\mathbb{P}_{\mu'}\left(T_1(n) > \frac{n}{2}\right)}_{B} \frac{n\Delta}{2}$$

$$\Rightarrow R_n(\pi,\mu) + R_n(\pi,\mu') \geq \frac{n\Delta}{2}\left[\mathbb{P}_\mu(A) + \mathbb{P}(B)\right]$$

$$= \frac{n\Delta}{2}\left[\mathbb{P}_\mu(A) + \mathbb{P}(A^c)\right]$$

Using the BH inequality,

$$R_n(\pi,\mu) + R_n(\pi,\mu') \geq \frac{n\Delta}{4}\exp\left(-D(\mathbb{P}_\mu, \mathbb{P}_{\mu'})\right)$$

Now, $D(\mathbb{P}_\mu, \mathbb{P}_{\mu'}) = \mathbb{E}_\mu[T_i(n)]D(v_i, v_i')$ (because for all $j \neq i$, the $D(v_j, v_j')$ term is 0)

$\therefore D(\mathbb{P}_\mu, \mathbb{P}_{\mu'}) = \mathbb{E}_\mu[T_i(n)]\frac{(2\Delta)^2}{2} \ldots$ (since $v_i \sim (0,1)$, $v_i' \sim (2\Delta, 1)$ have $D(v_i, v_i') = \frac{(0-2\Delta)^2}{2(1)^2} = 2\Delta^2$)

① Now $\mathbb{E}_\mu[T_i(n)] \leq \frac{n}{k-1}$ since the way i was chosen by taking the argmin of the number of pulls across all suboptimal arm

$$\therefore D(\mathbb{P}_\mu, \mathbb{P}_{\mu'}) \leq \frac{n}{k-1}(2\Delta^2)$$

$$\therefore R_n(\pi,\mu) + R_n(\pi,\mu') \geq \frac{n\Delta}{2}\exp\left(-\frac{2n\Delta^2}{k-1}\right)$$

Explanation of ① : $\sum_{j>1} \mathbb{E}_\mu[T_i(n)] \leq n$

$$\therefore \operatorname*{argmin}_{j>1} \mathbb{E}[T_j(n)] = i \Rightarrow \mathbb{E}_\mu[T_i(n)] \leq \frac{n}{k-1}$$

Set $\Delta = \sqrt{\frac{k-1}{4n}} \leq \frac{1}{2}$

Then,

$$R_n(\pi,\mu) + R_n(\pi,\mu') \geq \frac{\sqrt{n(k-1)}}{8\sqrt{e}}$$

$\therefore \max\left(R_n(\pi,\mu) + R_n(\pi,\mu')\right)$

$$\geq \frac{\sqrt{n(k-1)}}{16\sqrt{e}} \geq \frac{\sqrt{n(k-1)}}{27}$$

$$\frac{d}{d\Delta}\left(\frac{n\Delta}{2} e^{-\frac{2n\Delta^2}{k-1}}\right) = 0$$

$$\Rightarrow \frac{n}{2} e^{-\frac{2n\Delta^2}{k-1}} + \frac{n\Delta}{2}\left[-\frac{4n\Delta}{k-1}\right] e^{-\frac{2n\Delta^2}{k-1}} = 0$$

$$\Rightarrow \frac{n}{2} - \frac{2n^2\Delta^2}{k-1} = 0$$

$$\Rightarrow \Delta = \sqrt{\frac{k-1}{4n}}$$

$\boxtimes$

Idea: Construct μ' in a very particular way, so that we know there will be some misbehaviour on at least one of the two instances.

Idea: Choose event A s.t. its bad on one of the two instances, while its complement is bad on the other instances.

"Change of Measure" arguments

Recall

Under UCB,

$$R_n(UCB) \leq 3\sum\Delta_i + 8\sqrt{nk\log k}$$

Note: The regret bound we have obtained until now is not an instance-dependent bound.

Def: Policy π is CONSISTENT on $\mathcal{E}^k$ if $\forall v \in \mathcal{E}^k$,

$$R_n(\pi, v) = O(n^a) \ \forall\, a > 0$$

Note: In terms of estimation jargon, consistency tells that the estimator will converge to true means as the number of samples goes to infinity.

Note:

$R_n(\pi, v) = O(n^a)$ H O is "Little O".

Hence, logarithmic and poly-logarithmic policies are consistent

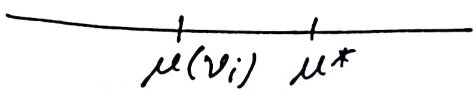
Let $\mu^* \in \mathbb{R}$, arm i s.t.

$$\mu(v_i) < \mu^*$$

$$d_{inf}(v_i, \mu^*, \mathcal{E}) = \inf\left\{D(v_i, v_i') : \mu(v_i') > \mu_i^*\ v_i' \subset \mathcal{E}\right\} \quad (1)$$

Interpretation of (1):

Pick an arm i. μ^* is some greater number. What is the slightest perturbation requeried ~~to push the~~ KL divergence ~~higher~~? to push its to the mean to higher than μ^*?



Note that the perturbation is in the KLDivergence and not in the mean.

Q : Why is this not just the $KLD(\mu_i, \mu^*)$:

Ans : Because we are working in a general family and hence the distributions are parametrized by not just the means, but other parameters as well.

Note : It is possible for a "rich" ~~even~~ family that even a slight perturbation in the KLD can push the means far far away.

THEOREM : Let π be a consistent policy on $\mathcal{E}^k$. For $\nu \in \mathcal{E}^k$,

$$\liminf_{n \to \infty} \frac{R_n(\pi, \nu)}{\log n} \geq \sum_{i : \Delta_i > 0} \frac{\Delta_i}{d_{inf}(\nu_i, \mu^*, \mathcal{E})}$$

where μ^* is the optimal mean in ν

Note : This tells us that logarithmic regret is the best that we can do

Corner case : If the family of distributions is very rich, then $d_{inf}(\nu_i, \mu^*, \mathcal{E}) = 0$. Then the regret becomes super-logarithmic & then logarithmic regret becomes impossible.

Announcements

1. HW2 is due on next Tuesday (12th March)
2. Extra class on a handful of Saturdays (9th March ~~is~~ may be the ~~first~~ second)
3. Extra class tomorrow 530 PM (6th March)