Minimax Lower Bounds

Source: Chap 15, Bandit Algorithms, L & S.

Recall from Notes (Set 10) — Minimax bound (i.e., showing existence of an environment for which regret is lower bounded by $\Theta(\sqrt{n})$, where n is the time horizon).

This set (Chap 15): Formally deriving the result for a k-arm unstructured bandit environments.

First, divergence decomposition lemma.

Setting: k-armed, environment $\mathcal{E}$.

$$\nu = (P_1\ P_2\ \cdots\ P_k) \in \mathcal{E}, \text{ and}$$
$$\nu' = (P_1'\ P_2'\ \cdots\ P_k') \in \mathcal{E}$$

Fix a policy π, and a time-horizon n.

Let $P_\nu \equiv P_{\nu\pi} \equiv P_{\nu\pi}^{(n)} \longrightarrow$ probability measure induced on $(A_1, X_1, A_2, X_2, \cdots, A_n, X_n)$

under environment $\nu \in \mathcal{E}$ and policy π (i.e., randomness due to environment ν as well as policy π, and with dependence over time steps induced by π).

Similarly, $P_{\nu'} \equiv P_{\nu'\pi} \equiv P^{(n)}_{\nu'\pi} \longrightarrow$ prob. measure due to (ν', π) over n time-steps.

Let $E_\nu[\cdot] = E_{P^{(n)}_{\nu\pi}}[\cdot]$,

and $E_{\nu'}[\cdot] = E_{P^{(n)}_{\nu'\pi}}[\cdot]$.

Lemma (13.1 in text): Setting as above. Then,

$$D(P_\nu, P_{\nu'}) = \sum_{j=1}^{k} E_\nu[T_j(n)]\, D(P_j, P_j').$$

Proof: Assume $D(P_j, P_j') < \infty \quad \forall\, j = 1, 2, \ldots, k.$

$$D(P_\nu, P_{\nu'}) = E_\nu\left[\ln\left(\frac{P_\nu}{P_{\nu'}}\right)\right]$$

(see text for more careful reasoning of existance/finiteness)

$$P_{\nu\pi}(A_1=a_1, X_1=x_1, \dots, A_n=a_n, X_n=x_n)$$

$$= P_{\nu\pi}(a_1, x_1, a_2, x_2, \dots, a_n, x_n)$$

$$= \prod_{t=1}^{n} \pi_t(a_t \mid a_1, x_1, \dots, a_{t-1}, x_{t-1}) \cdot P_{a_t}(x_t),$$

and $\pi_1(a_1 \mid a_0, x_0) \equiv \pi_1(a_1)$ (slightly misusing notation).

Similarly,

$$P_{\nu'\pi}(a_1, x_1, \dots, a_n, x_n) = \prod_{t=1}^{n} \pi_t(a_t \mid a_1, x_1, \dots, a_{t-1}, x_{t-1}) P'_{a_t}(x_t)$$

only place where ν' enters

Note: Policy is same in both environments. Thus, given the same sequence of observations of past actions and rewards, the next action has the same dist. under either ν or ν' environments.

$$\therefore \frac{dP_\nu}{dP_{\nu'}}(a_1, x_1, \dots, a_n, x_n) = \prod_{t=1}^{n} \left(\frac{P_{a_t}(x_t)}{P'_{a_t}(x_t)} \right)$$

$$D(P_v, P_{v'}) = E_v\left[\ln \frac{dP_v}{dP_{v'}}\right] = \sum_{t=1}^{n} E_v\left[\ln\left(\frac{P_{A_t}(x_t)}{P'_{A_t}(x_t)}\right)\right]$$

$$= \sum_{t=1}^{n} E_v\Big[\underbrace{E_v\left[\ln\left(\frac{P_{A_t}(x_t)}{P'_{A_t}(x_t)}\right)\Big| A_t\right]}_{D(P_{A_t}, P'_{A_t})}\Big]$$

$$= \sum_{t=1}^{n} E_v\left[D(P_{A_t}, P'_{A_t})\right]$$

$$= \sum_{t=1}^{n} E_v\left[\sum_{j=1}^{k} \chi_{\{A_t=j\}} D(P_j, P'_j)\right]$$

$$= \sum_{j=1}^{k} E_v\left[\sum_{t=1}^{n} \chi_{\{A_t=j\}} D(P_j, P'_j)\right]$$

$$= \sum_{j=1}^{k} D(P_j, P'_j)\, E_v\Big[\underbrace{\sum_{t=1}^{n} \chi_{\{A_t=j\}}}_{T_j(n)}\Big]$$

$$= \sum_{j=1}^{k} E_v\left[T_j(n)\right] D(P_j, P'_j)$$

$\blacksquare$

Rest of this section: Assume the arms are 1-Gaussian (i.e., $\mathcal{E} = \mathcal{E}_N^k(1)$).

Then, if $P_j = N(\mu_j, 1)$, $P_j' = N(\mu_j', 1)$,

$$D(P_j, P_j') = \frac{(\mu_j - \mu_j')^2}{2}.$$

(Recall KL divergence for Gaussians in Set 10).

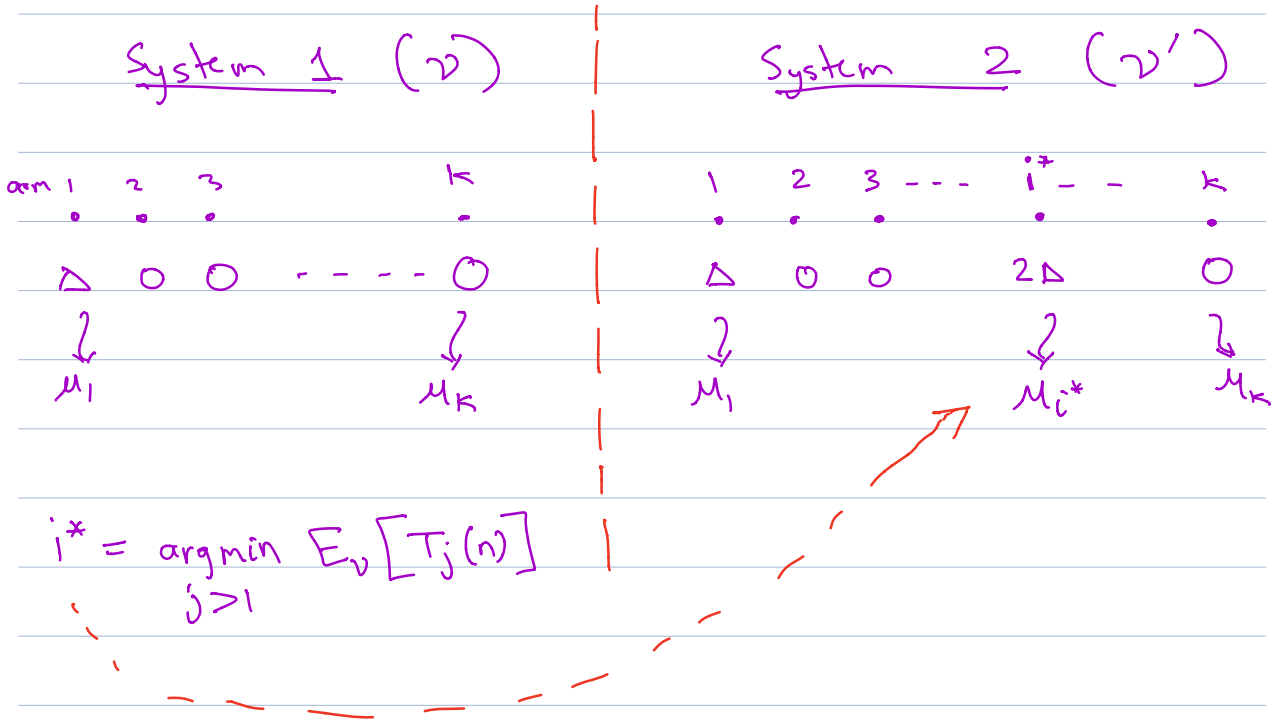
Theorem (15.2 in text): For any policy π, $\exists\ \mu = (\mu_1, \dots, \mu_k) \in [0,1]^k$ with

$$P_{\nu_\mu} = (P_1\ P_2 \cdots P_k), \quad P_j \sim N(\mu_j, 1)$$

such that $R_n(\pi, \nu_\mu) \geq \frac{1}{27}\sqrt{(k-1)n}$

Proof: Fix a policy π.

Let $\Delta = \sqrt{\frac{k-1}{4n}}$ (observe $\Delta < \frac{1}{2}$ $\forall\ k \geq 1, n \geq 1$).



Choose the arm that is EXPECTED to be played least frequently under (π, ν) (system 1) to be the changed arm in system 2.

ν: $(P_1 \cdots P_K) = (N(\Delta,1)\ N(0,1) \cdots N(0,1))$

ν': $(P_1' \cdots P_K') = (N(\Delta,1)\ N(0,1) \cdots \underbrace{N(2\Delta,1)}_{i^*} \cdots N(0,1))$

$1 \quad \cdots \quad i^* \quad \cdots \quad K$

Now, $R_n(\pi, \nu) = \sum_{j=1}^{K} \Delta_j E_\nu[T_j(n)]$

$\geq \Delta \left(\frac{n}{2}\right) P_\nu\left(T_1(n) \leq n/2\right).$

Similarly,

$$R_n(\pi, \nu') > P_{\nu'}\left(T_1(n) > \frac{n}{2}\right) \cdot \left(\frac{n}{2}\right) \cdot \Delta .$$

$P_{\nu'}\left(T_1(n) > \frac{n}{2}\right)$: strict lower bound on expected number of times arm 1 is played.

Δ: regret incurred whenever arm 1 played

Now, let $P = P_\nu$, $Q = P_{\nu'}$,

$$A = \left\{ T_1(n) \leq \frac{n}{2} \right\}.$$

As discussed in Set 10, A is a "bad event" under environment ν (the optimal arm is played for less than half the time steps), and A^c is a bad event under Q (the first arm, which is sub-optimal under ν' is played more than half the time steps).

From B-H inequality, $P(A) + Q(A^c) \geq \frac{e^{-D(P,Q)}}{2}$.

i.e., $P_\nu\left(T_1(n) \leq \frac{n}{2}\right) + P_{\nu'}\left(T_1(n) > \frac{n}{2}\right) \geq \frac{e^{-D(P_\nu, P_{\nu'})}}{2}$

Multiply by $\frac{n\Delta}{2}$:

$$\underbrace{\frac{n\Delta}{2} P_\nu\left(T_1(n) \le \frac{n}{2}\right)}_{\le R_n(\pi,\nu)} + \underbrace{\frac{n\Delta}{2} P_{\nu'}\left(T_1(n) > \frac{n}{2}\right)}_{< R_n(\pi,\nu')} \ge \frac{n\Delta}{4} e^{-D(P_\nu, P_{\nu'})}$$

$$\Rightarrow \quad R_n(\pi,\nu) + R_n(\pi,\nu') \ge \frac{n\Delta}{4} e^{-D(P_\nu, P_{\nu'})}$$

$\longrightarrow$ (*).

Next

$$D(P_\nu, P_{\nu'}) = \sum_{j=1} E_\nu[T_j(n)]\, \underbrace{D(P_j, P_j')}_{= \begin{cases} 0 & j \neq i^* \\ \frac{(2\Delta)^2}{2} & j = i^* \end{cases}}$$

$$= E_\nu[T_{i^*}(n)](2\Delta^2)$$

Finally, observe that

$$\sum_{j=2}^{n} E_\nu[T_j(n)] \leq n.$$

Also, $i^* = \underset{j \geq 2}{\operatorname{argmin}} E_\nu[T_j(n)]$

Pigeon hole principle $\Rightarrow E_\nu[T_{i^*}(n)] \leq \left(\frac{n}{k-1}\right)$.

$\sqrt{\frac{k-1}{4n}}$

$\therefore D(P_\nu, P_{\nu'}) \leq (2\Delta^2) \cdot \frac{n}{k-1} = \frac{1}{2}$.

→ (**)

From (*) and (**),

$$R_n(\pi, \nu) + R_n(\pi, \nu') \geq \frac{n\Delta}{4} \cdot e^{-1/2} \qquad 0.0379$$

$$= \sqrt{(k-1)n}\left(\frac{e^{-1/2}}{8}\right)$$

(pigeon hole principle)

$\Rightarrow$

$$\max\left\{R_n(\pi, \nu), R_n(\pi, \nu')\right\} \geq \frac{1}{2}\left(\sqrt{(k-1)n} \cdot \frac{e^{-1/2}}{8}\right)$$

i.e., either for $\tilde{\nu} = \nu$ or $\tilde{\nu} = \nu'$, we have

$$R_n(\pi, \tilde{\nu}) \geq \sqrt{n(k-1)} \cdot \left(\frac{e^{-1/2}}{16}\right)$$

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

Note: The only place in the proof where Gaussian noise was used was in the step:

$$D(P_{i*}, P'_{i*}) = \frac{(2\Delta)^2}{2}$$

Recall from the KL-UCB discussion (Set 7) that for $P_{i*} \sim \text{Bernoulli}(\mu)$, $P'_{i*} \sim \text{Bernoulli}(\mu')$

$$d(\mu, \mu') = \frac{\Delta^2}{2\mu(1-\mu)} + o(\Delta^2),$$

and thus we again have quadratic scaling in Δ. Thus, we can construct a minimax LB with Bernoulli rvs. Please read 15.3, Note 1 for additional discussion on more general rvs.