EE6106 LECTURE 16 (Dated: 26 March 2024)

We have seen that for any Gaussian instance, for any δ-sound algorithm,

$$\mathbb{E}[\tau] \geq 2\log\left(\frac{1}{4\delta}\right)\left(\sum_{1}^{k} \frac{1}{\Delta_i^2}\right)$$

We then also saw the Action Elimination algorithm.
Now we will study LUCB. The innovation is in terms of the sampling strategy
LUCB : Shivaram Kalyanakrishnan (ICML 2012)

LUCB

For $t \geq 1$, where t is the round counter

$$h_t = \arg\max_i \hat{\mu}_i$$

$$l_t = \arg\max_{i \neq h_t} \underbrace{\hat{\mu}_i + \alpha_{i,t}}_{\text{UCB of } i}$$

If $\hat{\mu}_{h_t} - \alpha_{h_t,t} \geq \hat{\mu}_{l_t} + \alpha_{l_t,t}$ ... (1)

- STOP
- output h_t

Else

- Pull h_t and l_t

Note

- $\alpha_{i,t}$ can be thought of as the confidence interval's radius
- Condition (1) ensures that LCB of h_t ensures that the is greater than the UCBs of all arms
- We are NEVER rejecting any arms in this algorithm, EXCEPT when we get the best two arms.

Shivaram uses :

$$\alpha_{i,t} = \sqrt{\frac{1}{2N_i(t-1)} \log\left(\frac{ckt^4}{\delta}\right)}$$

$N_i(t-1)$: number of pulls given ~~by~~ to the i^{th} arm

δ : the algorithm is δ-sound

We used :

$$\alpha_{i,t} = \sqrt{\frac{2\log}{N_i(t-1)}\left(\frac{2KCN_i^2(t-1)}{\delta}\right)}$$

in action elimination

<u>Note</u> ① We do not use the same c in both the definition

② The second $\alpha_{i,t}$ comes through a blatant use of the sub-Gaussian inequality. It depends on the number of pulls received by that arm.

③ $N_i(t-1)$ is the "local clock" of the arm i, while t is the "algo clock". These two aren't directly related. The first $\alpha_{i,t}$ also depends on the global clock time.

④ In the second case, if an arm is not pulled, then the confidence interval of the arm does not change. In the first case, it keeps growing until the arm is pulled.

⑤ The idea is that the first α promotes exploration, in that it will keep on increasing a not-so-much pulled optimal arm's C.I.

⑥ Under the second $\alpha_{i,t}$, there is not a guarantee that we will terminate with probability 1

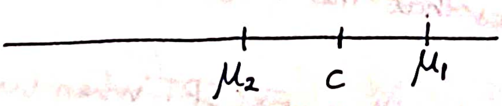
EXERCISE 1

Show ① results in sound algo

EXERCISE 2

Show ① implies $\mathbb{P}(T<\infty) = 1$

We shall now obtain a high probability bound under ②



$c = \frac{\mu_1 + \mu_2}{2}$, μ_2 is the next best ~~bad~~ arm

<u>Def</u> : ARM 1 is BAD if $\hat{\mu}_1 - \alpha_{1,t} < c$.
Arm $i \neq 1$ is BAD if $\hat{\mu}_1 + \alpha_{1,t} > c$

Claim : On B^c [bad event's complement] - if algo is not stopped, either h_t or ℓ_t is BAD.

Proof : Left as an EXERCISE

Note : Are we sure $h_t = 1$ and $\ell_t = 2$? No! The claim says that at least one of the arms must be bad.

Hint : Consider 3 cases

i) $h_t = 1$, $\ell_t \neq 1$, both GOOD $\Rightarrow$ then algorithm stops

ii) $\ell_t = 1$, $h_t \neq 1$, both GOOD $\Rightarrow$ UCB of arm 1 is less than empirical mean ~~LCB~~ of some other arm, which is not possible

iii) $\ell_t, h_t \neq 1$, both GOOD $\Rightarrow$ at least one arm in the $\{\ell_t, h_t\}$, but still the event is good, so there has to be some contradiction

For $i \neq 1$, let τ_i be the minimum number of pulls s.t. $\alpha_{i,t} < \Delta_i/4$ and $\tau_1 := \tau_2$

Note that once an arm has got ~~least~~ more than τ_i number of pulls, it can not be BAD.

Claim : On B^c, $\tau \leq \sum_1^K \tau_i$. Note that τ is the round counter, the pull counter would be given by 2τ.

Proof

$$\tau \leq \sum_1^\infty \mathbb{1}\{h_t \text{ is BAD or } \ell_t \text{ is BAD}\}$$

$$\leq \sum_{i=1}^k \underbrace{\sum_1^\infty \mathbb{1}\{(h_t = i \text{ or } \ell_t = i) \text{ AND } i \text{ is BAD}\}}_{\leq \tau_i}$$

$$\therefore \tau \leq \sum_1^K \tau_i$$

There is an algorithm called TRACK & STOP, which we will talk about in next class.

$\mathcal{E}$-class of MAB instances

Thm : Let π - sound be δ-sound on $\mathcal{E}$. Then for $\nu \in \mathcal{E}$,

$$\mathbb{E}_{\nu,\pi}[\tau] \geq C^*(\nu) \log\left(\frac{1}{4\delta}\right),$$

where

$$C^*(\nu)^{-1} = \sup_{\alpha \in P_k} \inf_{\nu' \in \mathcal{E}_{alt}(\nu)} \left[\sum \alpha_i D(\nu_i, \nu_i')\right]$$

P_k is the probability simplex of dimension of k.

$\mathcal{E}_{alt}$ has a different correct answer than those in $\mathcal{E}$. Further, $\mathcal{E}$ has a unique correct answer ~~and all~~

Thus if ν has optimal arm 1, then $\mathcal{E}_{alt}(\nu')$ has optimal arm different than 1.

$\sum \alpha_i D_i(\nu_i, \nu_i')$ can be thought of as a convex combination of the relative entropies

Proof :

The proof is trivial if $\mathbb{E}_{\nu,\pi}[\tau] = \infty$

Assume then that $\mathbb{E}_{\nu,\pi}[\tau] < \infty \Rightarrow \tau < \infty$ w.p. 1

Pick $\nu' \in \mathcal{E}_{alt}(\nu)$, then we claim that the following holds

$$2\delta \geq \frac{1}{2} \exp\left[-\sum_1^K \mathbb{E}_{\nu,\pi}[N_i(\tau)] D(\nu_i, \nu_i')\right]$$

$$\Downarrow$$

$$\sum \mathbb{E}_\nu[N_i(\tau)] D(\nu_i, \nu_i') \geq \log\left(\frac{1}{4\delta}\right) \forall \; \nu' \in \mathcal{E}_{alt}(\nu)$$

$$\frac{\mathbb{E}_\nu[\tau]}{c^*(\nu)} = \mathbb{E}_\nu[\tau] \sup_\alpha \inf_{\nu'} \sum \alpha_i D(\nu_i, \nu_i')$$

$$\geq \mathbb{E}_\nu[\tau] \inf_{\nu'} \sum \frac{\mathbb{E}_\nu[N_i(\tau)]}{\mathbb{E}_\nu[\tau]} D(\nu_i, \nu_i')$$

$$= \inf_{\nu'} \sum \mathbb{E}_\nu[N_i(\tau)] D(\nu_i, \nu_i')$$

$$\geq \log\left(\frac{1}{4\delta}\right) \quad \blacksquare$$

$\star$ Turns out.

For Gaussian instances,

$$2 \sum_1^K \frac{1}{\Delta_i^2} \leq c^*(\nu)$$

$\hookrightarrow$ Try to prove

For "optimal algo",

$$\Rightarrow \alpha_i^* \approx \frac{\mathbb{E}_\nu[N_i(\tau)]}{\mathbb{E}_\nu[\tau]}$$

$\Rightarrow$ Stop when $\inf_{\nu'} \left(\sum \mathbb{E}_\nu[N_i(\tau)] D(\nu_i, \nu_i')\right) \approx \log\left(\frac{1}{4\delta}\right)$