→ With prob $1-\delta$, we give right arm (after stopping) or we dont stop (for sound algorithm)

→ If we prove that we stop with prob 1 then we give right answer with prob $1-\delta$.

Goal: Minimize $\mathbb{E}[\tau]$, or high prob upper bound on τ.

Theorem: Consider 1-Gaussian instance μ (arm 1 is optimal)
For any sound algorithm,

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

Convention: $\Delta_1 = \Delta_2$

or $\Delta_i = \begin{cases} \mu_1 - \mu_2 & i=1 \\ \mu_1 - \mu_i & i \neq 1 \end{cases}$

• This is the first instance where we need that only one arm is optimal. If there are even 2 optimal arms, $\mathbb{E}[\tau]$ blows to ∞. Earlier, we used only 1 optimal arm only for cosmetic reasons (not needed). Here, the algo cannot differentiate whether there are 2 optimal arms or the the other arm is arbitrarily close to the first.

EE6106 LECTURE 15 (Date: 19 th March 2023)

Recall, FIXED ~~BANDIT~~ BAI CONFIDENCE

- K arms, 1-subGaussian
- Unique best arm
- Sound algorithm:

$$\mathbb{P}(\underbrace{\tau}_{\text{stopping time}} < \infty, \underbrace{\hat{a}_\tau}_{\text{algorithm output}} \neq 1) \leq \underbrace{\delta}_{\text{error threshold}}$$

τ is a causal random variable.

Thm: Consider 1-Gaussian instance μ, with unique best arm 1. On any algo,

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

where $\Delta_i = \mu_1 - \mu_i$ if $i \neq 1$ and $\mu_1 - \mu_2$ if $i = 1$

Note that the $\sum \frac{1}{\Delta_i^2}$ parameter is very similar to 'H_1'.

Proof : Fix arm a.

Define alternative instance $\mu^{[a]}$ as follows :

$$\text{for } a \neq 1, \quad \mu_i^{[a]} = \begin{cases} \mu_i & i \neq a \\ \mu_a + (\Delta_a + \varepsilon) & i = a \end{cases}$$

All arms except a are being kept the same, while the mean of arm a is ramped

For $a = 1$,

$$\mu_i^{[1]} = \begin{cases} \mu_i & i \neq 1 \\ \mu_1 - (\Delta_1 + \varepsilon) & i = 1 \end{cases}$$

Here we are bringing the mean down so that it is now suboptimal.

Also note that $\Delta_1 = \Delta_2$ by definition

We have :

$$\mathbb{P}_\mu(A^c) + \mathbb{P}_{\mu^{[a]}}(A) \geq \frac{1}{2}\left[e^{-\mathbb{E}_\mu[N_a(\tau)]\, D(\mu_a, \mu_a^{[a]})}\right]$$

Note : ① τ is a stopping time, so ideally we should be using it directly as such.

We proved DDL for cases where T is an exogenous variable.

T however is not Exogenously fixed.

② It turns out that DDL still holds for stopping times

Here, A is the event that we do stop and give an answer different ~~that~~ than the optimal arm for μ^a.

i.e.

$$A = \{\tau < \infty,\ \hat{a}_\tau \neq i^*(\mu^{[a]})\}$$

This implies, $\quad \mathbb{P}_{\mu^{[a]}}(A) \leq \delta$

$$A^c = \{\tau = \infty\} \cup \{\hat{a}_\tau = i^*(\mu^{[a]}), \tau < \infty\}$$

* If $\mathbb{P}_\mu(\tau = \infty) > 0$, $\mathbb{E}_\mu[\tau] = \infty$, bound holds trivially

$\therefore$ We assume that $\mathbb{P}_\mu(\tau = \infty) = 0$

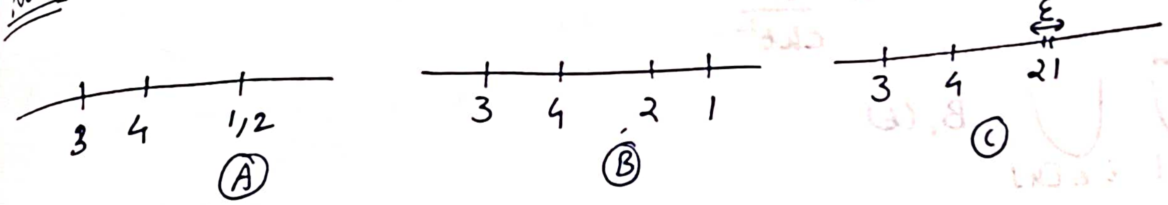
$$\therefore \mathbb{P}_\mu(A^c) = \mathbb{P}(\tau < \infty, \hat{a}_\tau = i^*(\mu^{[a]})) \leq \delta$$

$$\therefore 2\delta \geq \frac{1}{2} e^{-\mathbb{E}_\mu[N_a(\tau)]\frac{(\Delta_a + \varepsilon)^2}{2}}$$

Where the $\frac{\Delta_a + \varepsilon}{2}$ term arises due to the relative entropy between two Gaussians

Take logarithms, and let $\epsilon \to 0$, will give the required bound ■

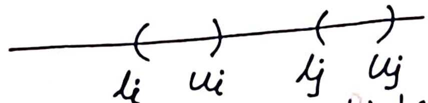
Note : If we allow the algorithm to output a $\mathcal{E}$-optimal, the algorithm works.



On Ⓑ, the algo works well

$\lim_{\mathcal{E}\to 0}$ Ⓒ = Ⓐ. Hence the algorithm fails on Ⓐ because it cannot disambiguate between Ⓐ and Ⓒ.

Algorithm Design : we would like to use confidence intervals.



We would like the LCB of one of the arms to be higher than the UCBs of all the other arms

$\boxed{\text{ACTION ELIMINATION}}$

- Start with set $A = [K]$ of active arms
- For $t \geq 1$, • pull each arm in A once
  - • update empirical means $\hat{\mu_i}$
  - • define $\alpha_t = \sqrt{\frac{2}{t}\log\left(\frac{2kt^2c}{\delta}\right)}$ ... radius of confidence interval.
  - • $E = \{ i \in A : \exists j \in A \text{ s.t. } \hat{\mu_j} - \alpha_t > \hat{\mu_i} + \alpha_t \}$
  - • $A = A \setminus E$
  - • If $|A| = 1$, stop, output ~~and~~ sole element A

Here t is actually a ROUND counter rather than a TIME counter

$\hat{\mu_j} - \alpha_t = LCB$, $\hat{\mu_j} + \alpha_t = UCB$ of some other arm.

"E" collects the arms whose LCB has been separated from the UCB of some other arm.

Any arm worse than any other arm is ejected.

c is a constant (scalar value)

$B_i(t) = \{ |\hat{\mu_i}(t) - \mu_i| \geq \alpha_t \}$

This is the event that after t pulls, arm i has a deviation of at least α_t from its true mean. This is a bad event because the confidence ~~of~~ interval of arm i will be invalid at time t.

$$\mathbb{P}(B_i(t)) \leq 2\exp\left(-\frac{t\alpha_t^2}{2}\right) \quad \ldots \text{ subGaussian C.I.}$$

$$= \frac{\delta}{ckt^2}$$

$$B = \bigcup_{t=1}^{\infty} \bigcup_{i \in [k]} B_i(t)$$

B is the event that at some point the true mean & the empirical mean of ~~the~~ any of arms differs by the α at that point & so has an invalid confidence interval.

B^c is the event that all arms have their C.I.s to be valid at all times.

$$\therefore \mathbb{P}(B) \leq \sum_{t=1}^{\infty} \sum_{i=1}^{k} \frac{\delta}{ckt^2}$$

$$= \frac{\delta}{c} \sum_{i=1}^{\infty} \frac{1}{t^2} \leq \delta \qquad \left(\text{set } c = \frac{\pi^2}{6}\right)$$

In practice, this algorithm is extremely conservative
It rarely ever makes a mistake, and runs way higher than it should
On B^c, algo will not stop and give a wrong answer.

$\Rightarrow$ AE is δ-sound.

We try to now state a higher-probability upperbound on the stopping time of the AE algo.

Claim : Under AE,

$$\mathbb{P}(\tau = \infty) = 0$$

* Assume all arms to have distinct means
  - Otherwise, in case the optimal arm is removed, two of the remaining arms may coalesce.

• Bound on τ on B^c (Good event)

In the good event, we give the correct answer, with probability 1

We want to bound the number of rounds for the sub-optimal arm i.

A sufficient condition for arm i elimination is

$$4\alpha_t < \Delta_i$$

we would need to solve a "Lambert equation" to solve this. Instead,
we note that,

T_i = min # of pulls satisfies

$$T_i = O\left(\frac{1}{\Delta_i^2}\log\left(\frac{K}{\delta\Delta_i}\right)\right)$$

$$\tau \leq \sum_{1}^{K} O\left(\frac{1}{\Delta_i^2}\log\left(\frac{K}{\delta\Delta_i}\right)\right)$$

Inefficiency in AE

1. The idea is that there is not really a need to sample every arm
2. we can just focus on better arms more frequently

LUCB Algorithm (Shivaram Kalyanakrishnan)