Week 7 - Lecture 33

We know:

$$\bar{R}_n = \sum_{i=1}^{K} \mathbb{E}[T_i(n)]\Delta_i = \sum_{i:\Delta_i>0} \mathbb{E}[T_i(n)]\Delta_i$$

Theorem: UCB pull each sub-optimal arm K in expectation at most

$$\Delta_i = \max_j \mu_j - \mu_i$$

$$\mathbb{E}[T_k(n)] \leq \frac{6\log n}{\Delta_k^2} + \pi^2/3 + 1 \quad \text{times.}$$

And Psuedo-regret is bounded as:

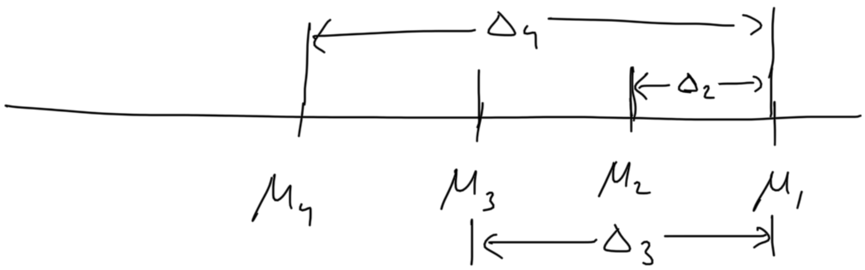
$$\bar{R}_n \leq 6 \sum_{i:\Delta_i>0} \frac{6\log K}{\Delta_K} + \sum_{i:\Delta_i>0} (\pi^2/3 + 1)\cdot\Delta_i \quad \text{—— (1)}$$

$$\Delta_i = \max_j \mu_j - \mu_i$$

Assuming arm 1 as optimal

then $\Delta_1 = 0$, $\Delta_2 > 0$, $\Delta_3 > 0$ . . .

Defn: $\Delta = \min_{i>1} \Delta_i$



Δ is smallest between Δ_2, Δ_3 & Δ_4

$\therefore$ the gap between highest mean and next highest mean

$\therefore \Delta < \Delta_i \;\; \forall \; i > 1$

Plugging in defn of Δ in ~~eq~~ eq 1

$$\bar{R}_n \leq 6(K-1)\frac{\log n}{\Delta} + \sum_{i:\Delta>0} \left(\pi^2/3 + 1\right)\Delta_i \qquad (2)$$

$$= O\left(\frac{K \log n}{\Delta}\right)$$ $\longrightarrow$ Sub-linear in n

$\longrightarrow$ If Δ is smaller then identifying the best arm is more challenging than if Δ is larger.
If the sub-optimal μs are closer to the optimal μ then figuring out the one optimal arm becomes harder.

$\longrightarrow$ The $\bar{R}_n$ bound given in the theorem captures this as $\bar{R}_n = O\left(\frac{K \log n}{\Delta}\right) \therefore$ as Δ becomes smaller, regret grows.

$$\hat{\mu}_i(t) = \frac{1}{T_i(t-1)} \underbrace{\sum_{s=1}^{T_i(t-1)} X_{si}}_{\substack{\text{first } T_i(t-1) \\ \text{observed samples}}}$$

→ s does not run like t, it is to be taken only for the rounds where i was played

---

$$UCB_i(t-1) = \hat{\mu}_i(t-1) + \sqrt{\frac{2\log t}{T_i(t-1)}}$$

Index of arm i

→ Suppose in round t sub-optimal arm i is played, this could have happened because for that round t, index of arm i is greater than the index of all other arms and in particular is greater than the index of arm 1 (assumed to be optimal)

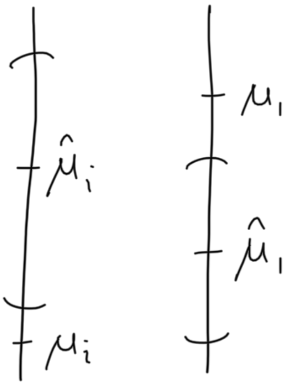
At round t, $I_t = i$ (i is sub-optimal)

if:
$$\hat{\mu}_i(t-1) + \sqrt{\frac{2\log t}{T_i(t-1)}} \geq \hat{\mu}_j(t-1) + \sqrt{\frac{2\log t}{T_j(t-1)}}$$

$\forall j$

$$\hat{\mu}_i(t-1) + \sqrt{\frac{2\log t}{T_i(t-1)}} \geq \hat{\mu}_1(t-1) + \sqrt{\frac{2\log t}{T_1(t-1)}}$$

→ Thus the index of i-th arm was overestimated and index of optimal arm was underestimated, because i-th arm was not played enough no. of times



→ It may have so happened that the estimates were not estimated properly and μ_1 lies above and μ_i lies below the respective confidence bounds.