Tail Probability Given n and ε

$$\mathbb{P}(|\hat{\mu} - \mu| \geq \varepsilon) \leq 2 e^{-\left(\frac{n\varepsilon^2}{2\sigma^2}\right)}$$

Confidence Interval Given n and δ, with prob $> 1-\delta$

$$\hat{\mu} \in \left[\mu - \sqrt{\frac{2\sigma^2 \log(2/\delta)}{n}}, \mu + \sqrt{\frac{2\sigma^2 \log(2/\delta)}{n}} \right]$$

Number of samples Given ε and δ, we need

$$n = \frac{2\sigma^2 \log(2/\delta)}{\varepsilon^2}$$

Assumption:

From now on, we will assume without loss of generality $\sigma = 1$

Arm 1 . . . Arm k

P_1 . . . P_k

* at time t,

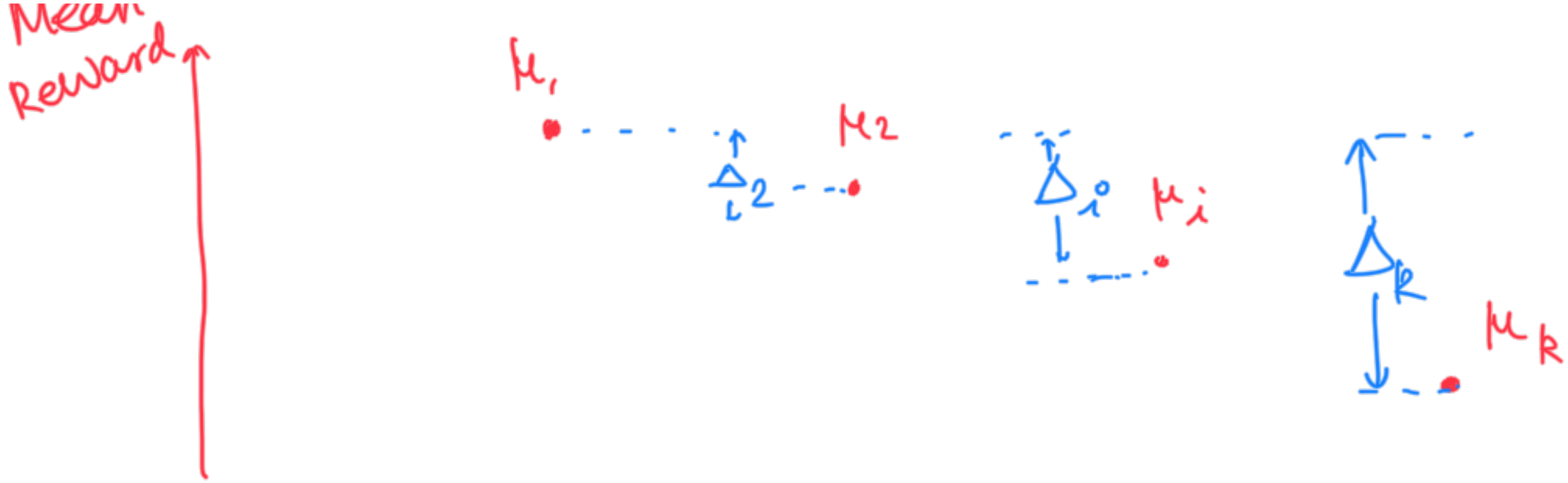
we play A_t and get $X_t \sim P_{A_t}$

* $\mu_i = \mathbb{E}_{P_i}[X_t]$, $\mu_* = \max_i \mu_i$

* Regret : $R_n = \mathbb{E}\left[\sum_{t=1}^{n} (\mu_* - X_t)\right]$

Notational Convention: $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_k$ (for purpose of analysis)

Arm 1 Arm k



Explore - Then- Commit

Exploration phase • play each arm m times and obtain the rewards

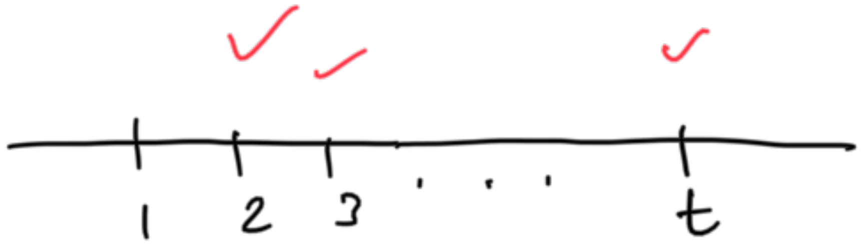
Exploitation phase • play the arm with best sample mean

not 't' samples

Sample mean $\hat{\mu}_i(t) = \dfrac{\sum_{s=1}^{t} X_s \cdot \mathbb{I}\{A_s = i\}}{T_i(t)}$

Total number of times we

Pick Arm 3



picked arm i within time t

$\mathbb{I}\{A_s = 3\}$

$\ast$ For $1 \leq t \leq mk$

$$A_1 = 1,\ A_2 = 1,\ \ldots,\ A_m = 1$$

Deterministic exploration

$$A_{m+1} = 2,\ A_{m+2} = 2,\ \ldots,\ A_{2m} = 2$$

$$\vdots$$

$$A_{m(k-1)+1} = k,\ \ldots,\ A_{mk} = k$$

$\ast$ for $t > mk$

$$A_t = \underset{i}{\operatorname{argmax}}\ \hat{\mu}_i(mk)$$

Exploitation (or) commit

(break ties arbitrarily)

Regret analysis of Explore - Then - Commit

Theorem: Let $n > mk$ (we have explored for mk rounds and then we are exploiting)

$$R_n \leq \underbrace{m \sum_{i=1}^{k} \Delta_i}_{\text{exploration}} + \underbrace{(n-mk) \sum_{i=1}^{k} \Delta_i e^{-\left(\frac{m\Delta_i^2}{4}\right)}}_{\text{exploitation}}$$

Proof: For sake of analysis ($\mu_1 \geq \mu_2 \geq \cdots \geq \mu_k$)

we know that $R_n = \sum_{i=1}^{k} \Delta_i \mathbb{E}[T_i(n)]$

For first mk (during exploration) we choose for first mk (during exploration), we choose

each action deterministically 'm' times

out of mk rounds

$$\mathbb{E}[T_i(n)] = \underbrace{m}_{\text{out of } mk \text{ rounds}} + \underbrace{(n-mk)}_{\text{exploitation}} \mathbb{P}(i^{th} \text{ arm was chosen after } mk \text{ rounds of exploration})$$

exploitation

arbitrary ties

arbitrary ties

$$\le m + (n-mk)\, \mathbb{P}(i^{th} \text{ arm was one of the best arms after } mk \text{ rounds})$$

$$= m + (n-mk)\, \mathbb{P}\left(\hat{\mu}_i(mk) \geq \max_{j \neq i} \hat{\mu}_j(mk)\right)$$

every other arm

$$\mathbb{P}(\underbrace{\text{arm } i \text{ beats all other arms}}_{\text{Event A}}) \le \mathbb{P}(\underbrace{\text{arm } i \text{ beats arm } 1}_{\text{Event B}})$$

$A \subseteq B$

$$\mathbb{P}(\text{arm } i \text{ beats arm } 1) = \mathbb{P}(\hat{\mu}_i(mk) \geq \hat{\mu}_1(mk))$$

$$= \mathbb{P}(\hat{\mu}_i(mk) - \hat{\mu}_1(mk) \geq 0)$$

$$= \mathbb{P}(\underbrace{\hat{\mu}_i(mk) - \mu_i}_{\substack{\text{Sample mean} \\ - \\ \text{True Mean} \\ \text{for arm } i}} + \underbrace{\mu_i - \mu_1}_{-\Delta_i} + \underbrace{\mu_1 - \hat{\mu}_1(mk)}_{\substack{\text{Sample mean} \\ - \\ \text{True Mean} \\ \text{for arm } 1}} \geq 0)$$

$$Y = \hat{\mu}_i(mk) - \mu_i \qquad Y' = \mu_1 - \hat{\mu}_1(mk)$$

Y and Y' are $\frac{1}{\sqrt{m}}$ subgaussian

$Y + Y'$ to be $\sqrt{\frac{2}{m}}$ subgaussian

$$= \mathbb{P}\Big(\underbrace{\hat{\mu}_i(mk) - \mu_i}_{Y} + \underbrace{\mu_i - \mu_1}_{-\Delta_i} + \underbrace{\mu_1 - \hat{\mu}_1(mk)}_{Y'} \geq 0\Big)$$

$$= \mathbb{P}(Y + Y' \geq \Delta_i)$$

$$\leq e^{-\left(\frac{m\Delta_i^2}{4}\right)}$$

$$R_n \leq \underbrace{m\sum_{i=1}^{k} \Delta_i}_{\text{exploration}} + \underbrace{(n - mk)\sum_{i=1}^{k} \Delta_i\, e^{-\left(\frac{m\Delta_i^2}{4}\right)}}_{\text{exploitation}}$$

Case

Gap Dependent Bound: Consider $k = 2$, $\Delta_1 = 0$, $\Delta_2 = \Delta$

Let us assume we know the gap Δ

$$R_n \le m\Delta + (n-2m)\Delta e^{-\left(\frac{m\Delta^2}{4}\right)} \quad -$$

(let us also say n is much larger than $2m$)

$$R_n \le m\Delta + n\Delta e^{-\left(\frac{m\Delta^2}{4}\right)} \quad -$$

to minimise the R.H.S, diff w.r.t m

$$\Delta + n\Delta\left(-\frac{\Delta^2}{4}\right) e^{-\left(\frac{m_*\Delta^2}{4}\right)} = 0$$

$$\cancel{\Delta} = n\cancel{\Delta}\left(\frac{\Delta^2}{4}\right) e^{-\left(\frac{m_*\Delta^2}{4}\right)}$$

$$e^{\left(\frac{m_*\Delta^2}{4}\right)} = n\Delta^2$$

$$e \quad = \quad \frac{n\Delta^2}{4}$$

$$\frac{m_* \Delta^2}{4} = \log\left(\frac{n\Delta^2}{4}\right)$$

Optimal exploration

$$m_* = \frac{4}{\Delta^2} \log\left(\frac{n\Delta^2}{4}\right)$$

notice the Δ^2

$$m_* = \max\left\{1, \left\lceil \frac{4}{\Delta^2} \log\left(\frac{n\Delta^2}{4}\right) \right\rceil \right\}$$

$$-\left(\frac{m\Delta^2}{4}\right)$$

$$R_n \leq m\Delta + n\Delta e$$

$$R_n \leq \underbrace{\Delta \max\left\{1, \left\lceil \frac{4}{\Delta^2}\log\left(\frac{n\Delta^2}{4}\right)\right\rceil\right\}}_{\text{Term I}}$$

$$+ \underbrace{n\Delta e^{-\left(\frac{\Delta^2}{4}\max\left\{1, \left\lceil \frac{4}{\Delta^2}\log\left(\frac{n\Delta^2}{4}\right)\right\rceil\right\}\right)}}_{\text{Term II}}$$

$$e^{-5} \leq e^{-\frac{6}{2}}$$

$$\lceil x \rceil < 1+x$$

$$\Delta\lceil x\rceil \leq \Delta + \Delta x$$

Term I

$$\Delta \quad \text{or} \quad \Delta + \frac{4}{\Delta}\log\left(\frac{n\Delta^2}{4}\right)$$

Term II

$$n\Delta e^{-\left(\cancel{\frac{\Delta^2}{4}}\cancel{\frac{4}{\Delta^2}}\log\left(\frac{n\Delta^2}{4}\right)\right)}$$

$$= n\Delta \frac{4}{n\Delta^2} = \frac{4}{\Delta}$$

$$R_n \leq \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, \log\left(\frac{n\Delta^2}{4}\right)\right\}\right)$$

as $\Delta \to 0$, above bound blows to ∞

$$R_n \leq \min\left\{n\Delta, \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, \log\left(\frac{n\Delta^2}{4}\right)\right\}\right)\right\}$$