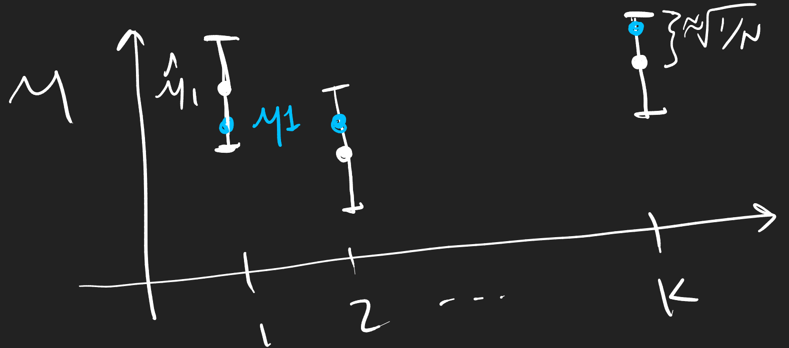
1) Explore-then-Commit

$$R(T) = \sum_{t=1}^{T} \mu^* - \mu_{a_t} = \underbrace{\sum_{t=1}^{NK} (\mu^* - \mu_{a_t})}_{R_1 \leq NK} + \underbrace{\sum_{t=1+NK}^{T} (\mu^* - \mu_{a_t})}_{R_2}$$

assume $r_t \in [0, 1]$

Consider the difference $\hat{\mu}_a - \mu_a$



$$\mu_a \in \left[\hat{\mu}_a \pm \sqrt{\frac{\log K/\delta}{N}}\right]$$

Lemma: After exploration

$$|\hat{\mu}_a - \mu_a| \lesssim \sqrt{\frac{\log(K/\delta)}{N}} \quad \text{w.p. } 1-\delta$$

Proof: Hoeffding & union bound $\quad P(A \cap B) \leq P(A) + P(B)$

Hoeffding's Bound: if $r_i \in [0, 1]$ and $\mathbb{E}[r_i] = \mu$

Then $r_1, \ldots, r_N$ iid

$$\left|\frac{1}{N}\sum_{i=1}^{N} r_i - \mu\right| \lesssim \sqrt{\frac{\log(1/\delta)}{N}}$$

$$R_2 = \sum_{t=NK+1}^{T} \mu^* - \mu_{\hat{a}^*} = (T-NK)(\overset{= \mu_{a^*}}{\mu^*} - \mu_{\hat{a}^*})$$

$$\leq (T-NK)\left(\hat{\mu}_{a^*} + \sqrt{\frac{\log(K/\delta)}{N}}\right.$$

w.p $1-\delta$

$$\left. - \left(\hat{\mu}_{\hat{a}^*} - \sqrt{\frac{\log(K/\delta)}{N}}\right)\right)$$

$$\leq (T-NK)\left[\underbrace{\hat{\mu}_{a^*} - \hat{\mu}_{\hat{a}^*}}_{\leq 0} + 2\sqrt{\frac{\log K/\delta}{N}}\right]$$

$$R(T) = R_1 + R_2 \leq \underline{NK} + \underline{2T\sqrt{\frac{\log K/\delta}{N}}} \quad \text{w.p. } 1-\delta$$

minimize w.r.t. N

$$N = \left(\frac{T}{2K}\sqrt{\log(K/\delta)}\right)^{2/3}$$

$$R(T) \lesssim T^{2/3} K^{1/3} \left(\log\frac{K}{\delta}\right)^{1/3}$$

$$\frac{R(T)}{T} = T^{-1/3} \to 0 \quad \text{as } T \to \infty \quad \checkmark$$

---

2) UCB Algorithm

$$\frac{\delta}{KT} \cdot K \cdot T = \delta$$

Alg 4: UCB

Initialize $\hat{u}_0^a = \infty$

for $t = 1, \ldots, T$

$a_t = \arg\max_a \hat{u}_t^a \leftarrow \hat{\mu}_t^a + \underbrace{\sqrt{\frac{\log(KT/\delta)}{N_t^a}}}_{\text{exploration bonus}}$ w.p. $1-\delta$

update $\hat{\mu}_{t+1}^{a_t}$ and $N_{t+1}^{a_t}$

\# times arm a pulled: $N_t^a = \sum_{k=1}^{t} \mathbb{1}\{a_t = a\}$

Empirical mean: $\hat{\mu}_t^a = \sum_{k=1}^{t} r_t \mathbb{1}\{a_t = a\} / N_t^a$

UCB Analysis:

Intuition: case 1) a_t has a large CI (high uncertainty) $\rightarrow$ exploring

case 2) a_t has a small CI (good arm) $\rightarrow$ exploiting

Regret-at-t: $\mu^* - \mu_{a_t} \leq \underline{\hat{u}_t^{a^*}} - \mu_{a_t}$ $\quad \mu^*$ in CI

$$\leq \underline{\hat{U}_t^{a_t}} - \mu_{a_t} \qquad a_t \text{ argmax}$$

$$= \underline{\hat{\mu}_t^{a_t}} + \sqrt{\frac{\log(KT/\delta)}{N_t^{a_t}}} - \mu_{a_t}$$

$$\leq 2\sqrt{\frac{\log(KT/\delta)}{N_t^{a_t}}}$$

$$R(T) = \sum_{t=1}^{T} \mu^* - \mu_{a_t} \leq 2\sqrt{\log(KT/\delta)} \underbrace{\sum_{t=1}^{T} \sqrt{1/N_t^{a_t}}}$$

<u>Claim:</u> $\sum_{t=1}^{T} \sqrt{1/N_t^{a_t}} \leq \sqrt{KT}$

$$R(T) \leq 2\sqrt{KT \log(KT/\delta)} \qquad O(T^{1/2})$$

sublinear! E-F-C $O(T^{2/3})$

<u>Proof of Claim:</u> $\sum_{t=1}^{T} \sqrt{1/N_t^{a_t}} = \sum_{t=1}^{T} \sum_{a=1}^{K} \mathbb{1}\{a_t = a\} \sqrt{\frac{1}{N_t^a}}$

$$= \sum_{a=1}^{K} \left(\sum_{t=1}^{T} \mathbb{1}\{a_t = a\} \sqrt{\frac{1}{N_t^a}} \right)$$

$$= \sum_{a=1}^{K} \underbrace{\sum_{i=1}^{N_T^a} \sqrt{1/i}}$$

$$\leq \sum_{a=1}^{K} \sqrt{N_T^a} \leq \sqrt{KT}$$

Notice: $\sum_{a=1}^{K} N_T^a = T$ one arm per timestep

$$\frac{1}{K}\sum_{a=1}^{K}\sqrt{N_T^a} \underset{\text{Jensen's}}{\leq} \sqrt{\frac{1}{K}\sum_{a=1}^{K} N_T^a} = \sqrt{T/K}$$

Principle: "Optimism in the Face of Uncertainty"