Recall $\underline{8/4/20}$.

UCB:

Let $(X_t)_{t=1}^{n}$ be a sequence of independent $\boxed{\text{1-subGaussian}}$ r.v with mean μ;

Let $\hat{\mu} = \frac{1}{n} \sum_{t=1}^{n} X_t$

$$\mathbb{P}\left[\mu \geq \hat{\mu} + \underbrace{\sqrt{\frac{2\log(1/\delta)}{n}}}\right] \leq \delta \quad \forall \delta$$

Define: $UCB_i(t-1,\delta) = \begin{cases} \infty & \text{if } T_i(t-1) = 0 \\ \underbrace{\hat{\mu}_i(t-1)} + \sqrt{\frac{2\log(1/\delta)}{T_i(t-1)}} & \text{o.w} \end{cases}$

Input k, δ

for $t \in 1, \ldots, n$ do {

Choose $A_t = \arg\max_i UCB_i(t-1,\delta)$

Observe $\underline{X_t}$ & update UCB_i $\forall i$ }

end

Terminology: argument inside argmax is called index of arm i

The algorithm chooses arm with largest index.

- W.l.o.g arm 1 has largest μ;

Suppose until round $t-1$, arm 1 was played more often than the others.
We may expect $\hat{\mu}_1(t-1) \approx \mu_1$

If $$\underbrace{\hat{\mu}_i(t-1) + \sqrt{\frac{2\log(1/\delta)}{T_i(t-1)}}}\le \mu_1 \approx \underline{\hat{\mu}_1(t-1) + \sqrt{\frac{2\log(1/\delta)}{T_1(t-1)}}}$$

THM: for any horizon n, if $\delta = 1/n^2$ then

$$R_n \le 3 \sum_{i=1}^{k} \Delta_i + \sum_{i:\Delta_i > 0} \frac{16 \log(n)}{\Delta_i}$$

$$\Rightarrow R_n \le O\left(\sqrt{n \log n \, k}\right)$$

Notation... $\underbrace{(X_{ti})_{t\in[n],\, i\in[k]}}$ be a coll of ind r-v, with dist of $X_{t,i}$ = that of P_i

Set $\boxed{\hat{\mu}_{is} := \frac{1}{s}\sum_{u=1}^{s} X_{ui}}$ empirical mean based on first s samples where i was played

Set Reward at time t,

$$\boxed{X_t = X_{T_{A_t}(t)\, A_t}}.$$

$$\begin{array}{c|ccc}
 & 1 & 2 & k \\
\hline
 & X_{11} & X_{12} & X_{1k} \\
 & \vdots & \vdots & \vdots \\
n & X_{n1} & X_{n2} & X_{nk}
\end{array}$$

Note: for an action a,

$$T_a(t) = \sum_{s=1}^{t} \mathbb{1}\{A_s = a\}.$$

times a is played in rounds $1, 2, \dots, t$,

Set $\underbrace{\hat{\mu}_i(t)} = \hat{\mu}_{iT_i(t)}$ — empirical mean of arm i, after round t.

Recall: $\rightarrow$ $\boxed{R_n = \sum_{i=1}^{k} \underbrace{\Delta_i} \mathbb{E}[T_i(n)].}$

We bound $\mathbb{E}[T_i(n)]$ for all subopt i

"decouple the randomness from the behavior of UCB".

Let G_i - "good event";

$$G_i = \left\{ \mu_1 < \min_{t \in [n]} UCB_1(t) \right\} \cap \left\{ \hat{\mu}_{i u_i} + \sqrt{\frac{2 \log(1/\delta)}{u_i}} < \mu_1 \right\}$$

where u_i is a constant to be chosen later; $u_i \in [n]$

$\therefore$ μ_1 never underestimated by the index of arm 1

Index of $i < \mu_1$

We show:

1) If G_i occurs $T_i(n) \le u_i$.

2) G_i^c occurs with low prob.

then
$$\mathbb{E}[T_i(n)] = \mathbb{E}[\mathbb{I}\{G_i\} T_i(n)] + \mathbb{E}[\mathbb{I}\{G_i^c\} T_i(n)]$$
$$\le u_i + \mathbb{P}(G_i^c) \cdot n \qquad \because T_i(n) \le n \ \forall i;$$

- Assume that G_i holds but $T_i(n) > u_i$

(ie) arm i was played more than u_i times over n rounds $\therefore \exists\, t \in [n]$ where $T_i(t-1) = u_i$ & $A_t = i$;

But:
$$UCB_i(t-1) = \hat{\mu}_i(t-1) + \sqrt{\frac{2\log(1/\delta)}{T_i(t-1)}}$$
$$= \hat{\mu}_{iu_i} + \sqrt{\frac{2\log(1/\delta)}{u_i}}$$

$< \mu_1$ (by def of G_i)

$< UCB_1(t-1)$

But then arm i would **NOT** be played in round t

$\therefore$ if G_i occurs then $\underline{T_i(n) \leq u_i}$.

$\boxed{\mathbb{P}(G_i^c)}$. :

$$G_i^c = \left\{ \mu_1 \geq \min_{t \in [n]} UCB_1(t) \right\} \cup$$

$$\left\{ \hat{\mu}_{i u_i} + \sqrt{\frac{2\log(1/\delta)}{u_i}} \geq \mu_1 \right\}$$

$\mu_1 = \mu_i + \Delta_i$

$$\left\{ \mu_1 \geq \min_{t \in [n]} UCB_1(t) \right\}$$

$$\subset \left\{ \mu_1 \geq \min_{s \in [n]} \hat{\mu}_{1s} + \sqrt{\frac{2\log(1/\delta)}{s}} \right\}$$

$$= \bigcup_{s \in [n]} \left\{ \mu_1 \geq \hat{\mu}_{1s} + \sqrt{\frac{2\log(1/\delta)}{s}} \right\} \leftarrow \boxed{3}$$

$$\therefore \mathbb{P}\left(\mu_1 \geq \min UCB_1(t) \right)$$

$$\leq \mathbb{P}(3) \leq \sum_{s=1}^{n} \mathbb{P}\Big(\underbrace{\mu_1 \geq \hat{\mu}_{1s} + \sqrt{\frac{2\log(1/\delta)}{s}}}_{\delta} \Big)$$

$$\leq n\delta.$$

The second term

$$\left\{ \hat{\mu}_{i u_i} + \sqrt{\frac{2\log(1/\delta)}{u_i}} \geq \mu_1 \right\}$$

Choose u_i so that

$$\Delta_i - \sqrt{\frac{2\log(1/\delta)}{u_i}} \geq c\Delta_i$$

for $c \in (0,1)$;

Note $\mu_1 = \mu_i + [\underbrace{\mu_1 - \mu_i}_{\Delta_i}] = \mu_i + \Delta_i$

$$\therefore \mathbb{P}r\left(\hat{\mu}_{iu_i} + \sqrt{\frac{2\log(1/\delta)}{u_i}} \geq \mu_1\right)$$

$$= \mathbb{P}r\left(\hat{\mu}_{iu_i} + \sqrt{\frac{2\log(1/\delta)}{u_i}} \geq \mu_i + \Delta_i\right)$$

$$= \mathbb{P}r\left(\hat{\mu}_{iu_i} - \mu_i \geq \underbrace{\Delta_i - \sqrt{\frac{2\log(1/\delta)}{u_i}}}_{= c\Delta_i}\right)$$

$$= \mathbb{P}r\left(\hat{\mu}_{i,u_i} - \mu_i \geq c\Delta_i\right)$$

$$\leq \exp\left(-\frac{u_i c^2 \Delta_i^2}{2}\right)$$

$$\therefore \mathbb{P}(G_i^c) \leq n\delta + \exp\left(-\frac{u_i c^2 \Delta_i^2}{2}\right)$$

$$\therefore \mathbb{E}[T_i(n)] \leq u_i + n\left(n\delta + \exp\left(-\frac{u_i c^2 \Delta_i^2}{2}\right)\right)$$

Want:

$$\Delta_i - \sqrt{\frac{2\log(1/\delta)}{u_i}} \geq c\Delta_i$$

$$\therefore \Delta_i(1-c) \geq \sqrt{\frac{2\log(1/\delta)}{u_i}}$$

$$\therefore u_i = \left\lceil \frac{2\log(1/\delta)}{\Delta_i^2(1-c)^2} \right\rceil$$

Next choose $\delta = 1/n^2$

$$\therefore \mathbb{E}[T_i(n)] \leq u_i + 1 + n\exp\left(-\frac{\cancel{2}\log(n^2)}{\cancel{\Delta_i^2}(1-c)^2 \cdot \cancel{2}} \cdot c^2 \cancel{\Delta_i^2}\right)$$

$$\le u_i + 1 + n \exp\left(\log n^{-2\frac{c^2}{(1-c)^2}}\right)$$

$$= u_i + 1 + n^{1-\frac{2c^2}{(1-c)^2}}.$$

$$= \left\lceil \frac{2\log n}{(1-c)^2\Delta_i^2} \right\rceil + 1 + n^{1-\frac{2c^2}{(1-c)^2}}$$

want: $\frac{2c^2}{(1-c)^2} \ge 1$ so $c \ge \sqrt{2}-1$.

Choose $c = 1/2$: get:

$$\boxed{E(T_i(n)) \le \frac{16\log n}{\Delta_i^2} + 3}$$

$$\therefore \; R_n \le 3\sum \Delta_i + \sum_{i:\Delta_i>0} \frac{16\log n}{\Delta_i}$$

THM. $\delta = 1/n^2$; then the regret is bounded by

$$8\sqrt{nk\log n} + 3\sum_{i=1}^{L} \Delta_i$$

Proof:

$$\mathbb{E}(T_i(n)) \le \frac{16\log n}{\Delta_i^2} + 3.$$

let Δ be a value;

$$R_n = \sum_{i:\Delta_i<\Delta} \Delta_i \underbrace{\mathbb{E}(T_i(n))} + \sum_{i:\Delta_i\ge\Delta} \Delta_i\, \mathbb{E}(T_i(n))$$

$$\le n\Delta + \sum_{\Delta_i\ge\Delta}\left(3\Delta_i + \frac{16\log n}{\Delta_i}\right)$$

$$\le n\Delta + 3\sum_i \Delta_i + \frac{16k\log n}{\Delta}$$

we get

Set $\Delta = \sqrt{\frac{16k\log n}{n}}$

Sublinear regret

$$\le 8\sqrt{nk\log n} + 3\sum_{i=1}^{k}\Delta_i$$

$\boxed{X_{11}}$ X_{21} X_{31} . . . X_{n1}

X_{42} X_{22} X_{32} X_{n2}

$\vdots$

X_{1K} X_{2K} X_{nK}

$X_1 = X_T$

$T_{A_t}(t)$. A_t.

$T_a(t) = \#$ a has been.

Suppose we play 1 1 3 2 1 1 4.