- sub-Gaussian R.V.s

  A mean-zero R.V. Z is σ^2-sub-Gaussian if $\mathbb{E}[\exp(\lambda Z)] \leq \exp(\sigma^2\lambda^2/2)$ $\forall \lambda \in \mathbb{R}$

- Chernoff bound w/ sub-Gaussian R.V.s:

  If $X_1, \ldots, X_n$ are IID R.V. w/ $\mathbb{E}[X_1] = \mu$, and $X_1 - \mu$ is σ^2-sub-Gaussian, then

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu > \varepsilon\right) \leq \exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right) =: \delta$$

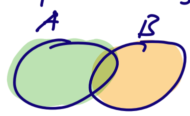
- Hoeffding's lemma: (If $X \in \{0,1\}$, then take $a=0, b=1$ $\Rightarrow$ $X - \mathbb{E}[X]$ is $\frac{1}{4}$-sub-Gaussian)

  If $X \in [a,b]$ then $X - \mathbb{E}[X]$ is $\frac{(b-a)^2}{4}$-sub-Gaussian

- Union Bound (or, sub-additivity of probability)

  Let $A_1, \ldots, A_n$ be events in a well-defined probability space,

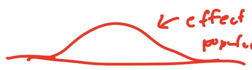
  then $\mathbb{P}\left(\bigcup_{i=1}^{n} \{A_i\}\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$



> Given n treatments., Input horizon $T \in \mathbb{N}$
>
> for $t = 1, 2, \ldots, T$
>
>   Patient t arrives w/ illness
>
>   Agent prescribes treatment $I_t \in [n] = \{1, \ldots, n\}$
>
>   Observe $X_{I_t, t} = \mathbb{1}\{\text{patient } t \text{ recovers given } I_t\}$

Let $X_{i,t} = \mathbb{1}\{\text{patient } t \text{ recovers given } i\}$.

Assume for each $i \in [n]$, $\{X_{i,t}\}_{t=1}^{T}$ are IID w/ $\mathbb{E}[X_{i,1}] = \theta_i$



perhaps $X_{i,t} = \mathbb{1}\{e_{i,t} > 1\}$

Questions:

- How long does it take to identify the best treatment $\arg\max_{i \in [n]} \theta_i$ with probability at least $1-\delta$?
- If $R_T = \max_{i \in [n]} \mathbb{E}\left[\sum_{t=1}^{T} X_{i,t} - X_{I_t,t}\right]$, how can we minimize R_T?

- **Expected Regret**

$$R_T = \max_{i\in[n]} \mathbb{E}\left[\sum_{t=1}^{T} X_{i,t} - X_{I_t,t}\right] = \sum_{i=1}^{n} \Delta_i \mathbb{E}[T_i]$$

where $\Delta_i = \max_{j\in[n]} \theta_j - \theta_i$, $\quad T_i = \sum_{t=1}^{T} \mathbb{1}\{I_t = i\}$

Warm-up: $n=2$. Consider a strategy that plays action $i \in [n]$ round-robin until $T_i = \tau$, then plays $\arg\max_{i\in[n]} \hat{\theta}_{i,\tau}$ for the remainder $T - n\tau$ times.

↖ empirical mean of first τ plays.

Define $\mathcal{E}_i = \left\{ |\hat{\theta}_{i,\tau} - \theta_i| \le \sqrt{\frac{2\sigma^2 \log(2n/\delta)}{\tau}} \right\}$, $i \in [n]$, $\quad \sigma^2 = \frac{1}{4}$

Lemma $\mathbb{P}\left(\bigcap_{i=1}^{n} \{\mathcal{E}_i\}\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{n} \{\mathcal{E}_i^c\}\right) \ge 1 - \delta.$

Proof $\hat{\theta}_{i,\tau} = \frac{1}{\tau} \sum_{s=1}^{\tau} X_{i,s}$, $\quad X_{i,s} \in \{0,1\}$ are iid $\mathbb{E}[X_{i,1}] = \theta_i$

$\Longrightarrow (X_{i,s} - \theta_i)$ is σ^2-sub-Gaussian with $\sigma^2 = 1/4$.

$$\mathbb{P}(\mathcal{E}_i^c) = \mathbb{P}\left(|\hat{\theta}_{i,\tau} - \theta_i| > \sqrt{\frac{2\sigma^2 \log(2n/\delta)}{\tau}}\right)$$

$$\le 2\exp\left(-\frac{\tau}{2\sigma^2}\left(\ \downarrow\ \right)^2\right)$$

$$= 2\exp\left(-\log(2n/\delta)\right)$$

$$= \frac{\delta}{n}$$

$$\mathbb{P}\left(\bigcup_{i=1}^{n}\{\mathcal{E}_i^c\}\right) \le \sum_{i=1}^{n} \mathbb{P}(\mathcal{E}_i^c) \le \sum_{i=1}^{n} \frac{\delta}{n} = \delta.$$

Because $\mathbb{P}\left(\bigcap_{i=1}^{n}\{\mathcal{E}_i\}\right) \ge 1-\delta$, assume $\bigcap_{i=1}^{n} \mathcal{E}_i$ occurs.

Assume WLOG $\theta_1 \ge \theta_2 \ge \dots \ge \theta_n$

If $\tau = \lceil 8\sigma^2 \Delta_2^{-2} \log(2n/\delta) \rceil$ then on events $\mathcal{E}_1 \cap \mathcal{E}_2$

$$\begin{aligned}
\hat{\theta}_{1,\tau} &> \theta_1 - \sqrt{\frac{2\sigma^2 \log(2n/\delta)}{\tau}} \qquad (\mathcal{E}_1) \\
&> \theta_1 - \frac{\Delta_2}{2} \\
&= \theta_2 + \frac{\Delta_2}{2} \\
&\ge \hat{\theta}_{2,\tau} - \sqrt{\frac{2\sigma^2 \log(2n/\delta)}{\tau}} + \frac{\Delta_2}{2} \qquad (\mathcal{E}_2) \\
&\ge \hat{\theta}_{2,\tau}.
\end{aligned}$$

$\Rightarrow$ With probability at least $1-\delta$, strategy identifies $\arg\max_{i \in [n]} \theta_i$.

$\theta_1 \ge \theta_2 \ge \dots \ge \theta_n$ $\quad (n=2 \quad \theta_1 > \theta_2)$

$$R_T = \sum_{i=1}^{n} \Delta_i \mathbb{E}[T_i] = \Delta_2 \mathbb{E}[T_2] \quad \text{since } n=2 \text{ and } \Delta_1 = 0.$$

$$\begin{aligned}
&= \Delta_2 \mathbb{E}[T_2 (\mathbb{1}\{\mathcal{E}_1 \cap \mathcal{E}_2\} + \mathbb{1}\{\mathcal{E}_1^c \cup \mathcal{E}_2^c\})] \\
&= \Delta_2 \mathbb{E}[T_2 \mathbb{1}\{\mathcal{E}_1 \cap \mathcal{E}_2\}] + \Delta_2 \mathbb{E}[T_2 \mathbb{1}\{\mathcal{E}_1^c \cup \mathcal{E}_2^c\}] \\
&\le \Delta_2 \tau \mathbb{E}[\mathbb{1}\{\mathcal{E}_1 \cap \mathcal{E}_2\}] + \Delta_2 (T-\tau) \mathbb{E}[\mathbb{1}\{\mathcal{E}_1^c \cup \mathcal{E}_2^c\}]
\end{aligned}$$

Note: on $\mathcal{E}_1 \cap \mathcal{E}_2$ we proved above ⊛ $\hat{\theta}_{1,\xi} > \hat{\theta}_{2,\xi}$

If $\hat{\theta}_{1,\xi} > \hat{\theta}_{2,\xi}$ then $T_2 = \xi$, $T_1 = T - \xi$

$$\mathcal{E}_1 \cap \mathcal{E}_2 \implies \{T_2 = \xi,\ T_1 = T - \xi\}$$

$$\mathbb{E}[T_2 \mathbb{1}\{\mathcal{E}_1 \cap \mathcal{E}_2\}] = \mathbb{E}[\xi \mathbb{1}\{\mathcal{E}_1 \cap \mathcal{E}_2\}]$$
$$= \xi \cdot \mathbb{E}[\mathbb{1}\{\mathcal{E}_1 \cap \mathcal{E}_2\}]$$
$$\leq \xi$$

$$T_2 \mathbb{1}\{\mathcal{E}_1 \cap \mathcal{E}_2\} = \xi \mathbb{1}\{\mathcal{E}_1 \cap \mathcal{E}_2\}$$

It is true that $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \geq 1 - \delta$, thus

w.p. $\geq 1 - \delta$ $\quad \Delta_2 T_2 \leq \Delta_2 \xi$.

$$\underbrace{\quad}_{\leq 1} \qquad \underbrace{\quad}_{= \mathbb{P}(\bigcup_{i=1}^{n} \{\mathcal{E}_i^c\}) \leq \delta}$$

$$\leq \Delta_2 \, 3 + \Delta_2 T \delta$$

$$\leq 1 + 8 \Delta_2^{-1} \sigma^2 \log(2n/\delta) + \Delta_2 T \delta$$

Idea: choose $\delta = 1/T$ $\qquad (\Delta_2 \leq 1)$

$\hookrightarrow \leq 1 + 8\Delta_2^{-1} \sigma^2 \log(2nT) + 1$

We wanted $R_T = o(T)$ - we have it! $\frac{\log(T)}{T} \to 0$

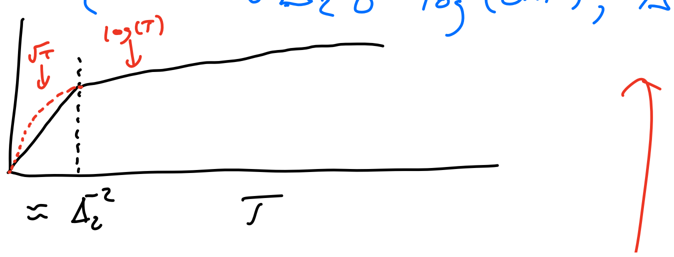
Note: If treatments were nearly identical so that $\Delta_2 = 10^{-10}$

then our regret bound says $R_T \leq 10^{10}$.

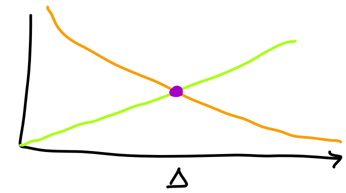
But this is nonsense, b/c it shouldn't matter what is prescribed, they act nearly the same.

We know: $R_T = \Delta_2 \mathbb{E}[T_2] \leq \Delta_2 T$

$$R_T \leq \min\{2 + 8\Delta_2^{-1}\sigma^2 \log(2nT), \Delta_2 T\}$$



maximize over Δ_2



$$\Rightarrow \quad R_T \le 2 + \max_{\Delta \ge 0} \min\{8\Delta^{-1}\sigma^2 \log(2nT), \Delta T\}$$

$$\Delta = \sqrt{\frac{8\sigma^2 \log(2nT)}{T}}$$

$$= 2 + \sqrt{8\sigma^2 T \log(2nT)}.$$

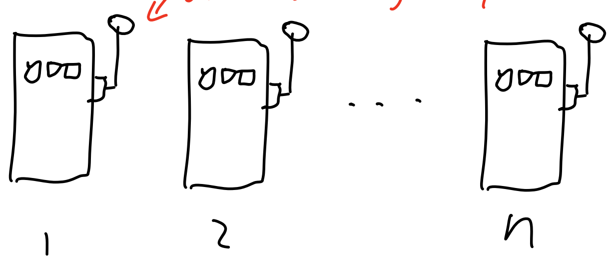
"instance - indep" regret bound

Note $R_T / T \to 0$

Multi-armed Bandit problem.

Imagine Casino

arm that you pull to play. "1-armed bandit"



Each machine i pays out random amount w/ mean θ_i

Elimination-style algorithm

Input: n arms $\mathcal{X} = \{1, \ldots, n\}$, confidence level $\delta \in (0,1)$.
Let $\widehat{\mathcal{X}}_1 \leftarrow \mathcal{X}, \ell \leftarrow 1$
while $|\mathcal{X}_l| > 1$ **do**
  $\epsilon_\ell = 2^{-\ell}$
  Pull each arm in $\mathcal{X}_l$ exactly $\tau_\ell = \lceil 2\epsilon_\ell^{-2} \log(\frac{4\ell^2|\mathcal{X}|}{\delta}) \rceil$ times
  Compute the empirical mean of these rewards $\widehat{\theta}_{i,\ell}$ for all $i \in \mathcal{X}_l$ ← using pulls only from round ℓ.
  $\mathcal{X}_{l+1} \leftarrow \mathcal{X}_l \setminus \{i \in \mathcal{X}_l : \max_{j \in \mathcal{X}_l} \widehat{\theta}_{j,\ell} - \widehat{\theta}_{i,\ell} > 2\epsilon_\ell\}$
  $\ell \leftarrow \ell + 1$
Output: $\widehat{\mathcal{X}}_{\ell+1}$ (or play the last arm forever in the regret setting)

Assume observation $X_{i,t}$ at time t from arm i is 1-sub-Gaussian and iid and $\mathbb{E}[X_{i,t}] \in [0,1]$

Assume WLOG $\theta_1 > \theta_2 \geq \ldots \geq \theta_n$, $\Delta_i = \theta_1 - \theta_i$, $\Delta = \min_{i : \Delta_i > 0} \Delta_i = \Delta_2$

Lemma With probability at least $1-\delta$,

arm $1 \in \mathcal{X}_\ell$ and $\max_{i \in \mathcal{X}_\ell} \Delta_i \leq 8\varepsilon_\ell \quad \forall \ell \in \mathbb{N}$.

Proof For all $\ell \in \mathbb{N}$ and $i \in [n]$ define

$$\mathcal{E}_{i,\ell} = \{ |\hat{\theta}_{i,\ell} - \theta_i| \leq \varepsilon_\ell \}$$

Note that $\varepsilon_\ell \geq \sqrt{\frac{2\log(4\ell^2|\mathcal{X}|/\delta)}{\tau_\ell}}$. Thus

$$\mathbb{P}\left(\bigcup_{\ell=1}^{\infty} \bigcup_{i \in \mathcal{X}_\ell} \mathcal{E}_{i,\ell}^c \right) \leq \mathbb{P}\left(\bigcup_{\ell=1}^{\infty} \bigcup_{i=1}^{n} \mathcal{E}_{i,\ell}^c \right)$$

$$\leq \sum_{\ell=1}^{\infty} \sum_{i=1}^{n} \mathbb{P}(\mathcal{E}_{i,\ell}^c)$$

$$\leq \sum_{\ell=1}^{\infty} \sum_{i=1}^{n} \frac{\delta}{2\ell^2|\mathcal{X}|} \leq \delta \quad (|\mathcal{X}| = n)$$

fact: $\sum_{\ell=1}^{\infty} \frac{1}{\ell^2} = \frac{\pi^2}{6} \leq 2$

$\mathbb{P}\left(\bigcap_{\ell=1}^{\infty} \bigcap_{i=1}^{n} \{\mathcal{E}_{i,\ell}\}\right) \geq 1-\delta$. Thus, assume these events hold.

If arm 1 is kicked out of $\mathcal{X}_\ell$, then $\exists$ arm $j \in \mathcal{X}_\ell$:

$$\hat{\theta}_{j,\ell} - \hat{\theta}_{1,\ell} > 2\varepsilon_\ell.$$

Assume $1 \in \mathcal{X}_\ell$ and show $1 \in \mathcal{X}_{\ell+1}$.

For all $j \in \mathcal{X}_\ell$ we have

$$\hat{\theta}_{j,\ell} - \hat{\theta}_{1,\ell} = (\hat{\theta}_{j,\ell} - \theta_j) + (\theta_1 - \hat{\theta}_{1,\ell}) - \Delta_j$$

$$\leq \varepsilon_\ell + \varepsilon_\ell + 0$$

$$= 2\varepsilon_\ell.$$

$\Rightarrow 1 \in \mathcal{X}_{\ell+1} \Rightarrow$ arm $1 \in \mathcal{X}_\ell \ \forall \ell \in \mathbb{N}$.

Fix any $i \in \mathcal{X}_\ell$ s.t. $\Delta_i > 4\varepsilon_\ell$. (★)

Recall $1 \in \mathcal{X}_\ell \ \forall \ell \in \mathbb{N}$.

Want to show that arm i does not enter $\mathcal{X}_{\ell+1}$.

We need to show $\max_{j \in \mathcal{X}_\ell} \hat{\theta}_{j,\ell} - \hat{\theta}_{i,\ell} > 2\varepsilon_\ell$.

$$\max_{j \in \mathcal{X}_\ell} \hat{\theta}_{j,\ell} - \hat{\theta}_{i,\ell} \geq \hat{\theta}_{1,\ell} - \hat{\theta}_{i,\ell}$$

$$= (\hat{\theta}_{1,\ell} - \theta_1) + (\theta_i - \hat{\theta}_{i,\ell}) + \Delta_i$$

$$\overset{(\mathcal{E}_1 \cap \mathcal{E}_\ell)}{\geq} -\varepsilon_\ell - \varepsilon_\ell + \Delta_i$$

$$\overset{(\not A)}{>} -2\varepsilon_\ell + 4\varepsilon_\ell$$

$$\geq 2\varepsilon_\ell.$$

$$\Longrightarrow \max_{j \in \mathcal{X}_\ell} \hat{\theta}_{j,\ell} - \hat{\theta}_{i,\ell} > 2\varepsilon_\ell \Longrightarrow i \notin \mathcal{X}_{\ell+1}$$

$$\max_{j \in \mathcal{X}_{\ell+1}} \Delta_j \leq 4\varepsilon_\ell = 8\varepsilon_{\ell+1} \qquad (\varepsilon_\ell = 2^{-\ell})$$

$$\Longrightarrow \max_{j \in \mathcal{X}_\ell} \Delta_j \leq 8\varepsilon_\ell \quad \forall \ell \in \mathbb{N}.$$

∎

Sample complexity. Let T_i be the total pulls from arm i at the end of game.

Total # pulls $= \sum_{i=1}^{n} T_i$.

Claim: $|\mathcal{X}_\ell| = 1$ for all ℓ: $8\varepsilon_\ell \le \Delta$

$8 2^{-\ell} \le \Delta \iff 8\Delta^{-1} \le 2^{\ell} \Leftarrow \ell \ge \lceil \log_2(8\Delta^{-1}) \rceil$

$\varepsilon_\ell = 2^{-\ell} \Rightarrow |\mathcal{X}_\ell| = 1$ for $\ell \ge \lceil \log_2(8\Delta^{-1}) \rceil$

For any $i > 1$

$$T_i = \sum_{\ell=1}^{\lceil \log_2(8\Delta^{-1}) \rceil} \tau_\ell \mathbb{1}\{i \in \mathcal{X}_\ell\}$$

$$\le \sum_{\ell=1}^{\lceil \log_2(8\Delta^{-1}) \rceil} \tau_\ell \mathbb{1}\{\Delta_i \le 8\varepsilon_\ell\}$$

$$= \sum_{\ell=1}^{\lceil \log_2(8\Delta_i^{-1}) \rceil} \tau_\ell$$

$\lceil \log_2(a) \rceil \le \log_2(a) + 1 = \log_2(2a)$

$$= \sum_{\ell=1}^{\lceil \log_2(8\Delta_i^{-1}) \rceil} \left\lceil 2\varepsilon_\ell^{-2} \log\left(\frac{4\ell^2 |\mathcal{X}|}{\delta}\right) \right\rceil$$

$$\le \left\lceil 2 \log\left(\frac{4 \lceil \log_2(8\Delta_i^{-1}) \rceil^2 |\mathcal{X}|}{\delta}\right) \right\rceil \sum_{\ell=1}^{\lceil \log_2(8\Delta_i^{-1}) \rceil} \varepsilon_\ell^{-2}$$

$$\le 2 \log\left(\frac{8 \log_2^2(16\Delta_i^{-1}) |\mathcal{X}|}{\delta}\right) \underbrace{\sum_{\ell=1}^{\lceil \log_2(8\Delta_i^{-1}) \rceil} 4^{\ell}}_{c\,\Delta_i^{-2}}$$

$$\leq c' \Delta_i^{-2} \log\left(\frac{\log_2(\Delta_i^{-1})|\mathcal{X}|}{\delta}\right).$$

$$\lceil \log_2(8\Delta_i^{-1}) \rceil \leq 1 + \log_2(8\Delta_i^{-1}) = \log_2(16\Delta_i^{-1})$$

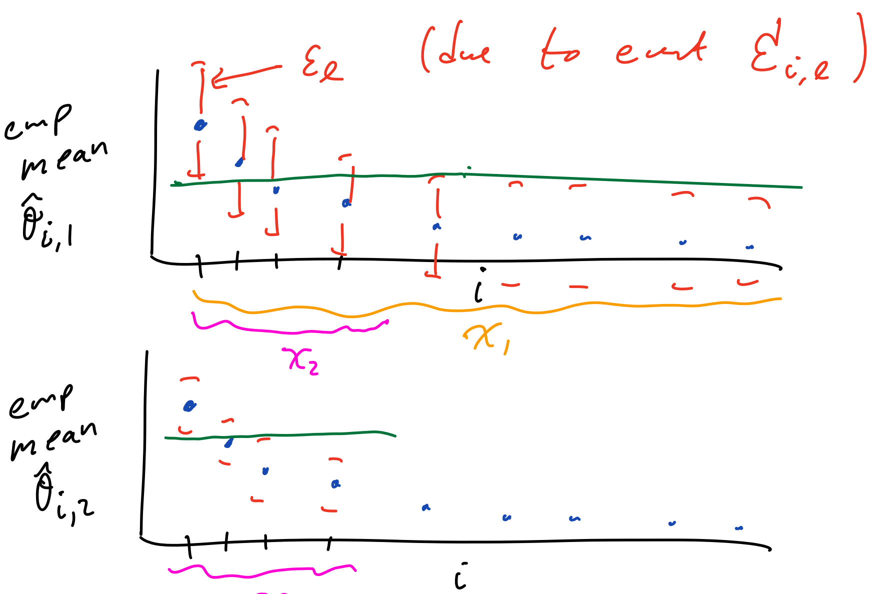
$$\sum_{\ell=0}^{N} a^\ell = \frac{1-a^{N+1}}{1-a}$$

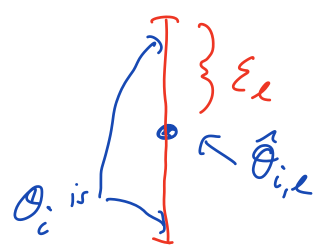
For any $c > 0$, $\log(a) \geq 1$ $\exists c' > 0$ s.t. $\log(ca) \leq c' \log(a)$

$$T_1 = \max_{i>1} T_i \qquad \sum_{i=1}^{n} T_i = T_1 + \sum_{i=2}^{n} T_i \leq 2 \sum_{i=2}^{n} T_i$$

Thm) W.p. $\geq 1-\delta$ after at most

$$c'' \sum_{i=2}^{n} \Delta_i^{-2} \log\left(\frac{\log_2(\Delta_i^{-1})|\mathcal{X}|}{\delta}\right)$$ total pulls we have that alg exits w/ best arm 1.





x_2

x_3

$$\mathcal{E}_{i,t} = \left\{ |\hat{\theta}_{i,t} - \theta_i| \leq \sqrt{\frac{2\log(4t^2|\mathcal{X}|/\delta)}{t}} \right\}$$

Claim: $\mathbb{P}\left(\bigcap_{t=1}^{\infty} \bigcap_{i=1}^{n} \mathcal{E}_{i,t} \right) \geq 1-\delta.$

Consider UCB (upper confidence bound alg)

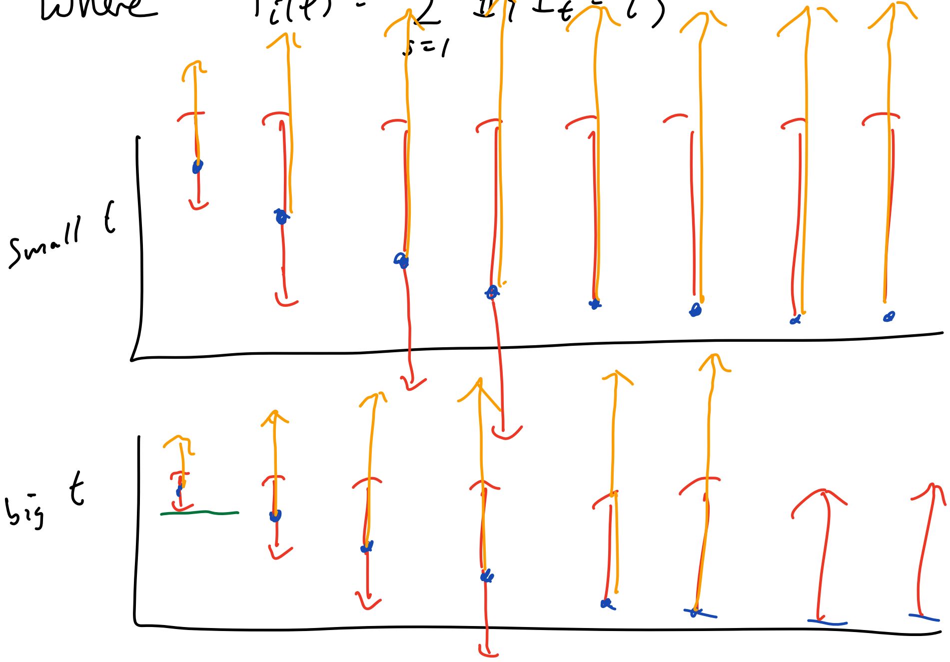
Pull every arm once

for $t = n+1, n+2, \ldots$

pull $I_t = \arg\max_{i \in [n]} \hat{\theta}_{i,T_i(t)} + 2\sqrt{\frac{2\log(4T_i(t)^2|\mathcal{X}|/\delta)}{T_i(t)}}$

$\geq \theta_i$

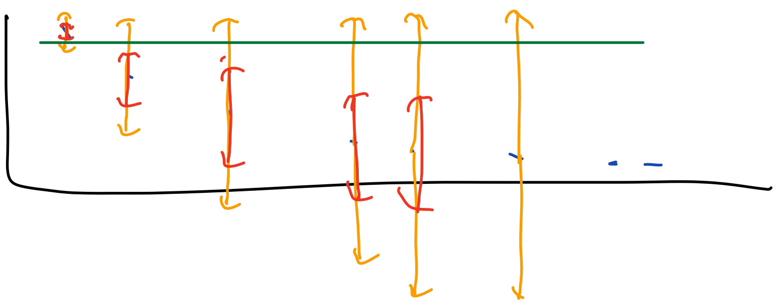
where $T_i(t) = \sum_{s=1}^{t-1} \mathbb{1}\{I_t = i\}$



On your homework, you show that $T_i(t) \le c\,\Delta_i^{-2} \log\left(\frac{T}{\delta}\right)$

$\forall t \in [T]$

$$R_T = \sum_{i=2}^{n} \Delta_i \mathbb{E}[T_i]$$



We've demonstrated the existence of an algorithm that satisfies $\mathbb{E}[T_i(T)] \le c\Delta_i^{-2} \log(T) \Rightarrow R_T \le c \sum_{i=2}^{n} \Delta_i^{-1} \log(T)$

?
↓

(Lai-Robbins '83)

Thm Any strategy that satisfies $\mathbb{E}[T_i(t)] = o(t^a)$ for any $a>0$ w/ $\Delta_i > 0$, we have $\lim_{T\to\infty} \frac{R_T}{\log(T)} \ge \sum_{i=2}^{n} \frac{2}{\Delta_i}$

$\Rightarrow$ Opt algorithm achieves regret $R_T \le \sum_{i=2}^{n} \frac{2}{\Delta_i} \log(T) + c$

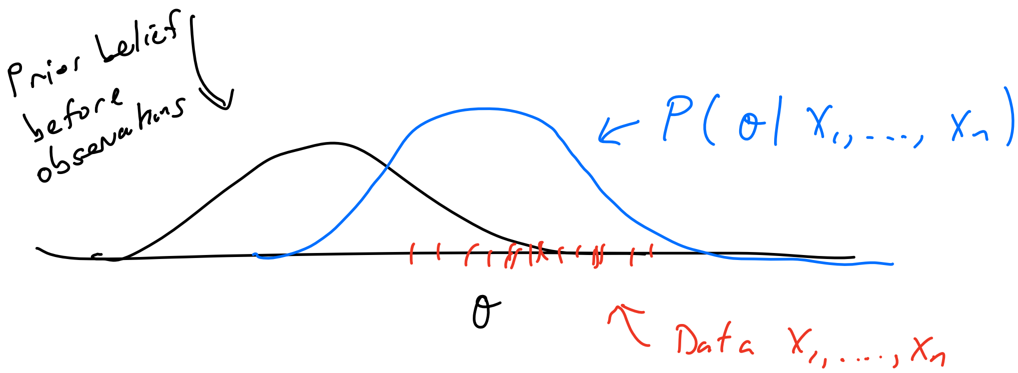
$$I_t = \arg\max_{i \in [n]} \hat{\theta}_{i,T_i(t)} + \sqrt{\frac{2\alpha \log(t)}{T_i(t)}}$$

Thm As $T \to \infty$, $\alpha \searrow 1$ UCB-α satisfies
$$\lim_{T\to\infty} \frac{R_T}{\log(T)} \le \sum_{i=2}^{n} \frac{2}{\Delta_i}.$$

+ Thompson Sampling

Thompson Sampling (1930's Thompson).

$\theta_* \sim P(\theta)$ Observe $X_1, \dots, X_n \sim \mathcal{N}(\theta_*, 1)$



$$P(\theta \mid \{X_i\}_{i=1}^n) \propto P(\{X_i\}_{i=1}^n \mid \theta)\, P(\theta)$$

Given posterior distribution for ith arm $P(\theta_i \mid \{X_{i,1}, \dots X_{i,T_i(t)}\})$

I can answer the question $\mathbb{P}\left(i = \arg\max_j \theta_j \mid \{\{X_{i,t}\}_{t=1}^{T_i(t)}\}_{i=1}^n\right)$

Thompson Sampling

for $t=1,2\dots$

Pull arm $I_t = i$ w/ prob $\mathbb{P}\left(i = \arg\max_j \theta_j \mid \{\{X_{i,t}\}_{t=1}^{T_i(t)}\}_{i=1}^n\right)$

↗

equivalent

$$\equiv \quad \theta'_t \sim P\left(\theta \mid \{\{X_{i,t}\}_{t=1}^{T_i(T)}\}_{i=1}^{n}\right)$$

$$I_t = \operatorname*{argmax}_{i \in [n]} [\theta'_t]_i$$

$$\mathbb{P}\left(i = \operatorname*{argmax}_j \theta_j \mid \{\{X_{i,t}\}_{t=1}^{T_i(T)}\}_{i=1}^{n}\right) = \int \mathbb{1}\{i = \operatorname*{argmax}_j \theta_j\} \, dP\left(\theta \mid \{\{X_{i,t}\}_{t=1}^{T_i(T)}\}_{i=1}^{n}\right)$$

In MAB, typically no sharing of information:

$$P(\theta) = \prod_{i=1}^{n} P_i(\theta_i)$$