- sub-Gaussian R.V.s

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

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

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

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

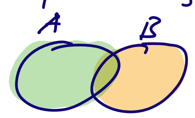
- Hoeffding's lemma: (If $X \in \{0,1\}$, then take $a=0, b=1 \Rightarrow X - \mathbb{E}[X]$ is $\frac{1}{4}$-sub-Gaussian)

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

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



Given n treatments.
for $t = 1, 2, \dots, T$
    Patient t arrives w/ illness
    Agent prescribes treatment $I_t \in [n] = \{1, \dots, n\}$
    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/ $\mathbb{E}[X_{i,1}] = \theta_i$

← effect size $e_{i,t}$ population

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

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

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

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

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

$$\le 2\exp\left(-\frac{\tau}{2\sigma^2}\left(\sqrt{\frac{2\sigma^2 \log(2n/\delta)}{\tau}}\right)^2\right)$$

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

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

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

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

$$\hat{\theta}_{1,\tau} > \theta_1 - \sqrt{\frac{2\sigma^2 \log(2n/\delta)}{\tau}} \qquad (\mathcal{E}_1)$$

$$> \theta_1 - \frac{\Delta_2}{2}$$

$$= \theta_2 + \frac{\Delta_2}{2}$$

$$\ge \hat{\theta}_{2,\tau} - \sqrt{\frac{2\sigma^2 \log(2n/\delta)}{\tau}} + \frac{\Delta_2}{2} \qquad (\mathcal{E}_2)$$

$$\ge \hat{\theta}_{2,\tau}.$$

$\Longrightarrow$ With probability at least $1-\delta$, strategy identifies $\operatorname{argmax}_{i\in[n]} \theta_i$.

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

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

$$= \Delta_2 \mathbb{E}\left[T_2 \left(\mathbb{1}\{\mathcal{E}_1 \cap \mathcal{E}_2\} + \mathbb{1}\{\mathcal{E}_1^c \cup \mathcal{E}_2^c\}\right)\right]$$

$$= \Delta_2 \mathbb{E}[T_2 \mathbb{1}\{\mathcal{E}_1 \cap \mathcal{E}_2\}] + \Delta_2 \mathbb{E}[T_2 \mathbb{1}\{\mathcal{E}_1^c \cup \mathcal{E}_2^c\}]$$

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

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

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

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

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! $\frac{\log(T)}{T} \to 0$

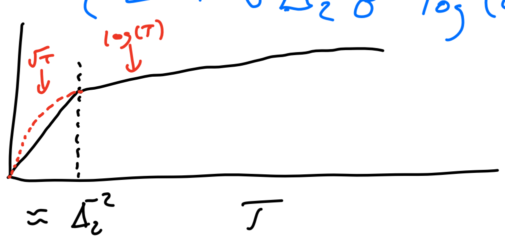
Note: If treatments were nearly identical so that $\Delta_2 = 10^{-10}$

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

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



maximize over Δ_2

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

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

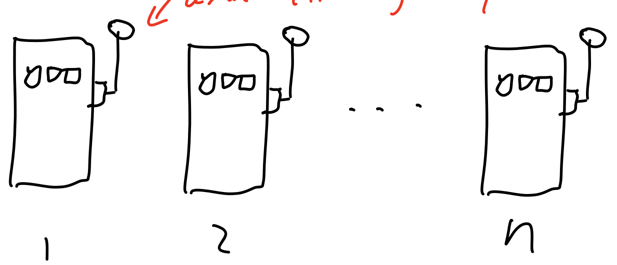
$$= \underbrace{2 + \sqrt{8\sigma^2 T \log(2nT)}}_{\text{"instance - indep" regret bound}}.$$

Note $\quad R_T / T \to 0$

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 $\mathcal{X} = \{1, \dots, n\}$, 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 $i \in \mathcal{X}_l$
  $\mathcal{X}_{l+1} \leftarrow \mathcal{X}_l \setminus \{i \in \mathcal{X}_l : \max_{j\in\mathcal{X}_l} \widehat{\theta}_{j,\ell} - \widehat{\theta}_{i,\ell} > 2\epsilon_\ell\}$
  $\ell \leftarrow \ell + 1$
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 $\theta_1 \geq \theta_2 \geq \dots \geq \theta_n$. $\Delta_i = \theta_1 - \theta_i$

<u>Lemma</u> With probability at least $1-\delta$,

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

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

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

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

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

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

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