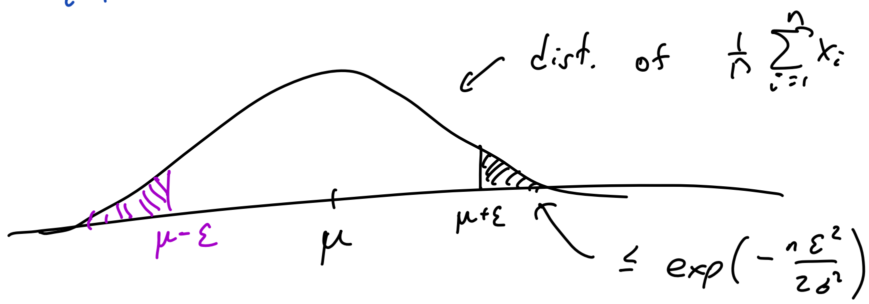
Def) R.V. (mean-0) Z is σ^2-sub-Gaussian if $\forall \lambda \in \mathbb{R}$ $\mathbb{E}[\exp(\lambda Z)] \leq \exp(\lambda^2 \sigma^2 / 2)$

Chernoff Bound) If $X_1, \dots, X_n$ are iid R.V. w/ $\mathbb{E}[X_1] = \mu$ and $\mathbb{E}[\exp(\lambda X_1)] \leq \exp(\lambda^2 \sigma^2 / 2)$ then

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



Note: above bound also applies to $-X_1, \dots, -X_n$, $\mu = \mathbb{E}[X_1]$

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

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_i X_i - \mu \right| > \varepsilon \right) = \mathbb{P}\left(\left\{ \frac{1}{n} \sum_i X_i - \mu > \varepsilon \right\} \cup \left\{ \frac{1}{n} \sum_i X_i - \mu < -\varepsilon \right\} \right)$$

$$\leq 2 \exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right) = \delta \quad \text{solve for } \varepsilon.$$

For any events A, B we have $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$

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

$$\exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right) = \delta/2 \longrightarrow \frac{2}{\delta} = \exp\left(\frac{n\varepsilon^2}{2\sigma^2}\right)$$

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

$$\varepsilon =$$

With probability at least $1-\delta$, we have

$$\left|\frac{1}{n}\sum_i X_i - \mu\right| \leq \sqrt{\frac{2\sigma^2 \log(2/\delta)}{n}}.$$

Patients arrive one at a time. I have a treatment (action 1) and control (action 2).

When I apply action $i \in \{1,2\}$ at the t^{th} time I observe iid R.V. $X_{i,t} \doteq \mathbb{1}\{\text{patient got better}\}$

$$\mathbb{E}[X_{i,t}] = \theta_i \in [0,1]$$

Fix $s \in \mathbb{N}$.

Define $\hat{\theta}_{i,s}$ = empirical mean of first s observations of applying action i.

$(X_{i,t} - \mu) \in [-\mu, 1-\mu] \Rightarrow (X_{i,t} - \mu)$ is $(\frac{1}{4})$-sub-Gaussian by Hoeffding's inequality. (If $X \in (a,b)$ then $X - \mathbb{E}[X]$ is $\frac{(b-a)^2}{4}$)

$$\mathcal{E}_i := \left\{ |\hat{\theta}_{i,s} - \theta_i| \leq \underbrace{\sqrt{\frac{2\sigma^2 \log(4/\delta)}{s}}} \right\} \quad \text{for } i = 1, 2$$

$$\mathbb{P}(\mathcal{E}_i^c) \leq 2\exp\left(-\frac{s(\;)^2}{2\sigma^2}\right) \leq \delta/2$$

$$1 - \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) = \mathbb{P}(\mathcal{E}_1^c \cup \mathcal{E}_2^c)$$

$$\leq \mathbb{P}(\mathcal{E}_1^c) + \mathbb{P}(\mathcal{E}_2^c) \leq \delta/2 + \delta/2 \leq \delta$$

W.p. $\geq 1-\delta$ $\quad \mathcal{E}_1 \cap \mathcal{E}_2$ holds.

$\Delta = \theta_1 - \theta_2$ $\leftarrow$ Assumption.

$$\tau = \lceil 8\sigma^2 \Delta^{-2} \log(4/\delta) \rceil$$

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_i^c) &= \mathbb{P}\left(|\hat{\theta}_i - \theta_i| > \sqrt{\frac{2\sigma^2 \log(4/\delta)}{\tau}} \right) \\
&\leq 2\exp\left(-\frac{\tau}{2\sigma^2} \cdot \left(\sqrt{\frac{2\sigma^2 \log(4/\delta)}{\tau}} \right)^2 \right) \\
&= 2\exp(-\log(4/\delta)) \\
&= 2\exp(\log(\delta/4)) \\
&= 2 \cdot \frac{\delta}{4} = \delta/2
\end{aligned}$$

Strategy: apply treatments 1+2 each τ times

then output $\underset{i \in \{1,2\}}{\operatorname{argmax}} \hat{\theta}_{i,\tau}$

On $\mathcal{E}_1 \cap \mathcal{E}_2$:

$$\begin{aligned}
\hat{\theta}_{1,\tau} &> \theta_1 - \sqrt{\frac{2\sigma^2 \log(4/\delta)}{\tau}} \qquad (\mathcal{E}_1) \\
&> \theta_1 - \sqrt{\frac{2\sigma^2 \log(4/\delta)}{8\sigma^2 \Delta^{-2} \log(4/\delta)}} \\
&= \theta_1 - \Delta/2
\end{aligned}$$

$$= \theta_2 + \Delta/2 \qquad (\Delta = \theta_1 - \theta_2)$$

$$> \hat{\theta}_{2,s} \underbrace{- \sqrt{\frac{2\sigma^2 \log(4/\delta)}{s}} + \Delta/2}_{\geq 0} \qquad (\mathcal{E}_2)$$

$$> \hat{\theta}_{2,s}$$

$$\Longrightarrow \quad \hat{\theta}_{1,s} > \hat{\theta}_{2,s} \Longrightarrow \text{We will output } \arg\max_i \theta_i$$

Instead identifying the "best", what if we just want to minimize <u>regret</u>.

Suppose I have $\underline{n}$ treatments.

for $t = 1, 2, \ldots, T$

- Patient arrives
- Doctor prescribes action $I_t \in [n] = \{1, \ldots, n\}$
- Observe outcome $X_{I_t, t} \in \{0, 1\}$ (patient survives if $X_t = 1$)

$\underline{X \in \mathbb{R}^{n \times T}}$

$$\text{Lives saved} = \sum_{t=1}^{T} X_{I_t, t}$$

Compare to the best single action in hindsight: $\max_{i \in [n]} \sum_{t=1}^{T} X_{i,t}$

Assume $\mathbb{E}[X_{i,t}] = \theta_i$, $\{X_{i,t}\}_{t=1}^{\infty}$ are iid for all $i \in [n]$

$$\text{Regret}(T) = R_T = \max_{i=1,\dots,n} \mathbb{E}\left[\sum_{t=1}^{T} X_{i,t} - \sum_{t=1}^{T} X_{I_t,t} \right]$$

Want $R_T = o(T)$ (i.e. $\frac{R_T}{T} \to 0$ as $T \to \infty$)

We say an algorithm has "no regret" or "sub-linear regret"

$$\begin{aligned}
R_T &= \max_{i=1,\dots,n} \mathbb{E}\left[\sum_{t=1}^{T} X_{i,t} - \sum_{t=1}^{T} X_{I_t,t} \right] \\
&= \max_i \sum_{t=1}^{T} \mathbb{E}[X_{i,t}] - \sum_{t=1}^{T} \mathbb{E}[X_{I_t,t}] \\
&= \max_i T\theta_i - \sum_{t=1}^{T} \mathbb{E}\left[\sum_{j=1}^{n} X_{j,t} \mathbb{1}\{I_t = j\} \right] \\
&= \max_i T\theta_i - \sum_{t=1}^{T} \sum_{j=1}^{n} \mathbb{E}[X_{j,t} \mathbb{1}\{I_t = j\}] \\
&= \max_i T\theta_i - \sum_{t=1}^{T} \sum_{j=1}^{n} \mathbb{E}[X_{j,t}]\mathbb{E}[\mathbb{1}\{I_t = j\}] \\
&= \max_i T\theta_i - \sum_{j=1}^{n} \theta_j \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\{I_t = j\} \right]
\end{aligned}$$

$$= \max_i \sum_{j=1}^{n} \mathbb{E}[T_j]\,\theta_i - \sum_{j=1}^{n} \mathbb{E}[T_j]\,\theta_j$$

$$= \max_i \sum_{j=1}^{n} (\theta_i - \theta_j)\,\mathbb{E}[T_j]$$

$$= \sum_{j=1}^{n} \Delta_j\,\mathbb{E}[T_j] \quad ,$$

$$T_i := \sum_{t=1}^{T} \mathbb{1}\{I_t = i\}, \qquad T = \sum_{j=1}^{n} T_j = \sum_{j=1}^{n} \mathbb{E}[T_j]$$

$$\Delta_j = \begin{cases} 0 & \text{if } \theta_j = \max_i \theta_i \\ \max_i \theta_i - \theta_j & \text{o.w.} \end{cases}$$

$$R_T = \sum_{j=1}^{n} \Delta_j\,\mathbb{E}[T_j]$$

Conclude: If $R_T = o(T)$ then $\overset{\theta_j < \max_i \theta_i}{\Downarrow}$ $\frac{\mathbb{E}[T_j]}{T} \to 0$

Suppose $\Delta_{\hat{j}} > 0$ for some $\hat{j}$ and $\mathbb{E}[T_{\hat{j}}] \geq cT$ for some constant c.

Then $R_T = \sum_{j=1}^{n} \Delta_j\,\mathbb{E}[T_j]$

$$\geq \Delta_{\hat{j}} \mathbb{E}[T_{\hat{j}}]$$

$$\geq \Delta_{\hat{j}} cT \qquad \text{(by assump)}$$

$$\neq o(T)$$

Suppose we played every action τ times and then played $\underset{i \in [n]}{\operatorname{argmax}} \hat{\theta}_{i,\tau}$ for the remaining $T - n\tau$ times. What is the regret?