Multi-armed Bandit Bounds

n arms, $\theta_1^* > \theta_2^* \geq \ldots \geq \theta_n^*$, $\quad \Delta_i = \theta_1^* - \theta_i^* \quad \forall i$

Best arm identification

for $t = 1, 2, \ldots$

Player chooses $I_t \in [n]$

Nature reveals $\theta^*_{I_t, t}$

If player chooses to stop:

exit and output $\hat{i}_t \in [n]$.

Def] We say an alg is δ-PAC for $\theta^* \in [0,1]^n$ if

when the algorithm exits at time $\mathcal{T}$ and outputs $\hat{i}_{\mathcal{T}} \in [n]$

if $\mathbb{P}_{\theta^*}\left(\hat{i}_{\mathcal{T}} = \arg\max_{i \in [n]} \theta_i^* \right) \geq 1 - \delta$.

W.p. $\geq 1-\delta$ Elimination alg exited at time $\mathcal{T}$ and output best arm

and $\mathcal{T} \leq c \sum_{i=2}^{n} \Delta_i^{-2} \log\left(\frac{n \log(\Delta_i^{-2})}{\delta} \right)$

Hypothesis testing:

Suppose we observe IID R.V. $X_1, X_2, \ldots, X_n \in \mathbb{R}$

$H_0 : X_i \sim P_0 = \mathcal{N}(0,1)$

$H_1 : X_i \sim P_1 = \mathcal{N}(\Delta, 1)$

Given $\{X_i\}_{i=1}^n$ can we determine H_0 vs H_1?

Recall: $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\quad \mathbb{P}\left(|\hat{\mu}_n - \mathbb{E}[\hat{\mu}_n]| \geq \underbrace{\sqrt{\frac{2\log(2/\delta)}{n}}}_{\leq \frac{\Delta}{2}}\right) \leq \delta$

So, if $\boxed{n \geq \frac{8\log(2/\delta)}{\Delta^2}}$ then

Output $\begin{cases} 1 & \text{if } \hat{\mu}_n > \frac{\Delta}{2} \\ 0 & \text{if } \hat{\mu}_n \leq \frac{\Delta}{2} \end{cases}$

$\rightarrow$ sufficient condition.

What is a necessary condition on n so that hypothesis test is decided correctly w.p. $\geq 1-\delta$?

Let $\phi : \mathbb{R}^n \to \{0,1\}$

$\mathbb{P}_i(\cdot)$ is the probability under H_i.

$$\inf_{\phi} \max\{\mathbb{P}_0(\phi = 1), \mathbb{P}_1(\phi = 0)\}$$

$$\geq \frac{1}{4}\exp\left(-KL(\mathbb{P}_1 \| \mathbb{P}_0)\right)$$

Fix p, q distributions

$$KL(p \| q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx$$

$$KL(\mathbb{P}_1^{(n)} \| \mathbb{P}_0^{(n)}) = \int_{x_1, \dots, x_n} \underbrace{\mathbb{P}_1(x_1, \dots, x_n)}_{= \mathbb{P}_1(x_1)\mathbb{P}_1(x_2) \cdots \mathbb{P}(x_n)} \log\left(\frac{\mathbb{P}_1(x_1, \dots, x_n)}{\mathbb{P}_0(x_1, \dots, x_n)}\right) dx$$

$$= \int \mathbb{P}_1(x_1)\mathbb{P}_1(x_2) \cdots \mathbb{P}(x_n) \sum_{i=1}^{n} \log\left(\frac{\mathbb{P}_1(x_i)}{\mathbb{P}_0(x_i)}\right) dx$$

$$= \sum_{i=1}^{n} \int \mathbb{P}_1(x_1) \cdots \mathbb{P}_1(x_n) \log\left(\frac{\mathbb{P}_1(x_i)}{\mathbb{P}_0(x_i)}\right) dx$$

$$= \sum_{i=1}^{n} \int \mathbb{P}_1(x_i) \log\left(\frac{\mathbb{P}_1(x_i)}{\mathbb{P}_0(x_i)}\right) dx_i$$

$$= n \int \mathbb{P}_1(x_1) \log\left(\frac{\mathbb{P}_1(x_1)}{\mathbb{P}_0(x_1)}\right) dx_1$$

If $\mathbb{P}_1(x_1) = \mathcal{N}(\Delta, 1)$, $\mathbb{P}_0(x_1) = \mathcal{N}(0,1)$

then $\int \mathbb{P}_1(x_1) \log\left(\frac{\mathbb{P}_1(x_1)}{\mathbb{P}_0(x_1)}\right) dx_1$

$$= \Delta^2/2$$

$$\Longrightarrow \inf_{\phi} \max\{\mathbb{P}_0(\phi=1), \mathbb{P}_1(\phi=0)\}$$

$$\geq \frac{1}{4}\exp(-n\Delta^2/2)$$

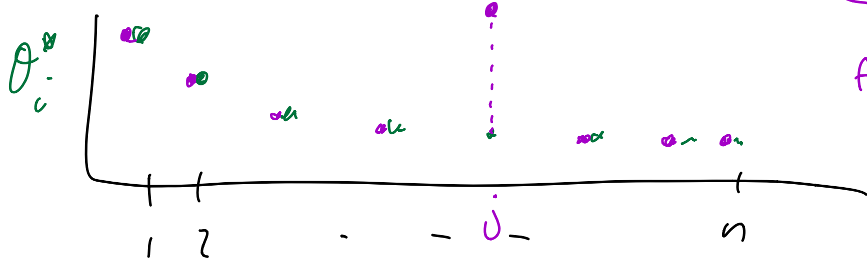
$$\overset{\text{(want)}}{=} \delta \quad \text{if} \quad n = \frac{2\log\left(\frac{1}{4\delta}\right)}{\Delta^2}$$

Take away: To determine whether n iid samples have mean 0 or Δ w.p. $\geq 1-\delta$ it is necessary and sufficient to have $n = \Theta\left(\frac{\log(1/\delta)}{\Delta^2}\right)$

Consider best arm identification

Fix $\theta^* \in [0,1]^n$, $\quad \theta_i^{*(j)} := \begin{cases} \theta_i^* & \text{if } i \neq j \\ \theta_1^* + \varepsilon & \text{if } i = j \end{cases}$

for arbitrarily small $\varepsilon > 0$



If alg is δ-PAC, on θ^* alg outputs arm 1 w.p. $\geq 1-\delta$, and on $\theta^{*(j)}$ alg outputs arm j w.p. $\geq 1-\delta$.

But by lower bound argument, to determine whether player is playing θ^* vs $\theta^{*(j)}$, player must determine whether arm j has mean

$$\theta_j^* \text{ or } \theta_j^{*(j)} = \theta_1^* + \varepsilon$$

Equivalent to determining whether arm j has mean 0 or $(\theta_1^* - \theta_j^*) + \varepsilon$

$$= \Delta_j + \varepsilon$$

By above $T_j \geq \dfrac{2\log\left(\frac{1}{4\delta}\right)}{(\Delta_j + \varepsilon)^2}$

$\Longrightarrow$ Any δ-PAC algorithm requires

$$\boxed{c \sum_{i=2}^{n} \Delta_i^{-2} \log(1/4\delta)}$$

Proved Mannor Tsitsiklis '04.

$$LB \geq \Delta_2^{-2} \log\left(\frac{\log(\Delta_2^{-2})}{\delta}\right)$$

Appeal to law of iterated logarithm.

Regret

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

Player chooses $I_t \in [n]$

and recieves reward $X_{I_t, t} \sim N(\theta^*_{I_t}, 1)$

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

$$\theta^* = (\Delta, 0, 0, \ldots, 0) \in [0,1]^n$$

Any algorithm that plays against θ^*,

$$\exists \hat{i} : \mathbb{E}_{\theta^*}[T_{\hat{i}}] \leq \frac{T}{n}$$

$$\varphi^* = (\Delta, 0, 0, \ldots, \underset{\hat{i}}{2\Delta}, 0, \ldots, 0)$$

$$\Rightarrow \text{ if } T_{\hat{i}} \leq \frac{T}{n} < \frac{2\log\left(\frac{1}{4\delta}\right)}{(2\Delta)^2}$$

Happens when $\Delta \approx \sqrt{\frac{n}{T}}$

then by above LB's algorithm cannot tell whether its playing against θ^* or φ^*,

If $\mathbb{E}_{\theta^*}[T_1] \le \frac{T}{2}$ then

$$R_T = \sum_{i \ge 2}^{n} \mathbb{E}_{\theta^*}[T_i] \Delta_i \ge \Delta T/2$$

If $\mathbb{E}_{\varphi^*}[T_1] > \frac{T}{2}$ then

$$R_T \ge \Delta T/2$$

But by above argument, w/ const prob, Player cannot distinguish between playing against θ^* or $\varphi^* \Longrightarrow R^T \ge \underbrace{\Delta T/2}_{\approx c\sqrt{nT}}$

(Not a proof)

More formally $R^T \geq \sqrt{(n-1)T/27}$

Recall, we showed elimination alg achieves $R_T \leq \sqrt{nT\log(nT)}$

$\exists$ algs that achieves $\sqrt{nT}$.

Gap-dependent Regret LB:

any algorithm that satisfies

$\mathbb{E}[T_i] = o(T^\alpha)$ for $\alpha \in (0,1)$

and $\Delta_i > 0$, then

$$\lim_{T \to \infty} \frac{R_T}{\log(T)} \geq \sum_{i=2}^{n} \Delta_i^{-1}$$

Recall, elimination alg satisfies

$$R_T \le \sum_{i=2}^{n} \Delta_i^{-1} \log(nT)$$

Reducing regression to hypothesis testing.

Identify 2 hypotheses about the mean of $X_1, \dots, X_n$ iid.

$$H_0 : X_i \sim N(0,1)$$

$$H_1 : X_i \sim N(\Delta,1)$$

$$\min_{\phi} \max\{\mathbb{P}_0(\phi=1), \mathbb{P}_1(\phi=0)\} \ge \frac{1}{4}e^{-n\Delta^2/2}$$

Goal: Output $\hat{\mu}$ to estimate $\mathbb{E}[X_1]$

$$\max_{\mu} \mathbb{E}_{X_i \sim N(\mu,1)}\left[\left(\hat{\mu}(X_1,\dots,X_n) - \mu \right)^2 \right]$$

$$\geq \max_{i \in \{1,2\}} \mathbb{E}_i\left[\left(\hat{\mu} - \mathbb{E}_i[X_1] \right)^2 \right]$$

$$\geq \max_{i \in \{1,2\}} \left(\frac{\Delta}{2}\right)^2 \mathbb{P}_i\left(\left(\hat{\mu} - \mathbb{E}[X_1] \right)^2 > \left(\frac{\Delta}{2}\right)^2 \right)$$

$$\geq \left(\frac{\Delta}{2}\right)^2 \inf_{\phi} \max_{i \in \{1,2\}} \max\{\mathbb{P}_1(\phi = 0), \mathbb{P}_{..})$$

$$\geq \left(\frac{\Delta}{2}\right)^2 \frac{1}{\cancel{e}\,4} e^{-n\Delta^2/2}$$

$$\gtrsim \frac{c}{n} \qquad \text{for } \Delta = \frac{1}{\sqrt{n}}$$