Recap

Binary hypothesis testing: P_0, P_1 distributions on $\mathcal{X}$

$$T : \mathcal{X}^n \longrightarrow \{0,1\}$$

False positive: $\mathbb{P}_{\bar{x} \sim P_0^n}\left[T(\bar{x}) = 1\right] = \alpha(T)$

False negative: $\mathbb{P}_{\bar{x} \sim P_1^n}\left[T(\bar{x}) = 0\right] = \beta(T)$

$$\alpha(T) + \beta(T) \geq 1 - \delta_{TV}(P_0^n, P_1^n)$$

Ratio tests: $T(\bar{x}) = 1$ iff $\dfrac{P_1^n(\bar{x})}{P_0^n(\bar{x})} \geq \Delta$

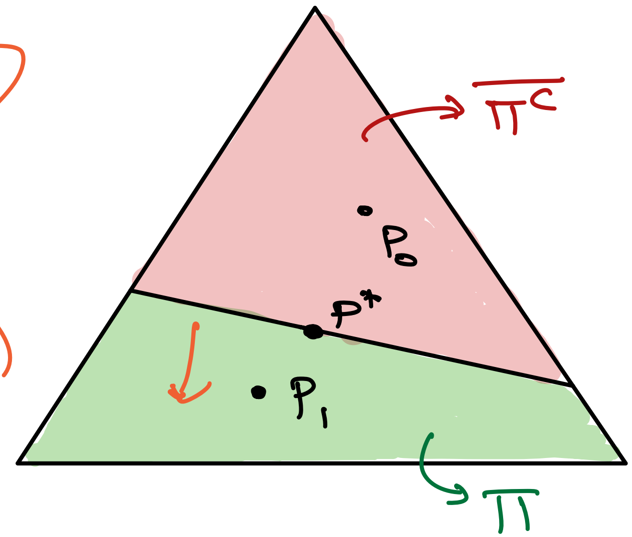
$$\Updownarrow$$

$$\underbrace{D(P_{\bar{x}} \| P_0) - D(P_{\bar{x}} \| P_1) \geq \frac{1}{n} \log \Delta}_{P_{\bar{x}} \in \Pi}$$

Behavior of exponents

depends on Δ

$$\alpha(T) \doteq 2^{-n \cdot D(p^* \| P_0)} \geq 2^{-nD(P_1 \| P_0)}$$



$$\beta(T) \doteq 2^{-n \cdot D(p^* \| P_1)} \geq 2^{-nD(P_0 \| P_1)}$$

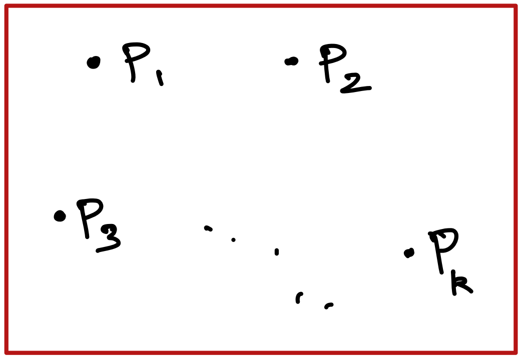
$$\frac{1}{10}\alpha(T) + \frac{9}{10}\beta(T) \gtrsim 2^{-n \cdot D(P_0 \| P_1)} + 2^{-nD(P_1 \| P_0)}$$

$$\gtrsim 2^{-n \cdot \min\{D(P_0 \| P_1),\, D(P_1 \| P_0)\}}$$

$C(P_0, P_1)$

$$\alpha(T) + \beta(T) \geq 1 - \frac{1}{2}\|P_0 - P_1\|_1 \geq 1 - \frac{1}{2}\sqrt{2\ln 2 \cdot n D(P_0 \| P_1)}$$

Multiple hypotheses: Bound via Fano's inequality



$\mathcal{V} \equiv$ Collection of k hypotheses

$$\longrightarrow \quad \bar{X} = (X_1 \ldots \ldots X_n) \quad \longrightarrow \quad \hat{V}$$

$$\bar{X} \sim P_V^n$$

$$p_e = \mathbb{P}[V \neq \hat{V}]$$

r.v. V : uniform on $\mathcal{V}$

(Fano): $H_2(p_e) + p_e \log(|\mathcal{V}|-1) \geq H(V|\bar{X})$

▷ $$p_e \geq 1 - \frac{n \cdot \mathbb{E}_{v_1, v_2 \in \mathcal{V}}\left[D(P_{v_1} \| P_{v_2})\right] + 1}{\log |\mathcal{V}|}$$

Proof:

$$1 + p_e \cdot \log|\mathcal{V}| \geq \underbrace{H_2(p_e)}_{\leq 1} + p_e \cdot \underbrace{\log(|\mathcal{V}|-1)}_{\leq \log|\mathcal{V}|} \geq H(V \mid \overline{X}) = \underbrace{H(V)}_{\log|\mathcal{V}|} - I(V;X)$$

$$I(V;\overline{X}) = D\left(P(V,\overline{X}) \,\|\, P(V)\,P(\overline{X})\right)$$

$$= \underbrace{D\left(P(V) \,\|\, P(V)\right)}_{=0} + D\left(P(\overline{X} \mid V) \,\|\, \underbrace{P(\overline{X})}_{\mathbb{E}_V P_V^n}\right)$$

$$\mathbb{E}_{V_1} D\left(P_{V_1}^n \,\|\, \mathbb{E}_{V_2} P_{V_2}^n\right)$$

$$\leq \left(\mathbb{E}_{V_1,V_2} D\left(P_{V_1} \,\|\, P_{V_2}\right)\right) \cdot n$$

Minimax rates

How many samples do we need to learn properties of distributions

$$\theta : \underbrace{\Pi}_{\text{set of distributions}} \longrightarrow \overbrace{\Theta}^{\text{set of properties}}$$

$$\underbrace{\hat{\theta}}_{\text{estimator for } \theta} : \mathcal{X}^n \longrightarrow \Theta$$

$$\ell : \Theta \times \Theta \longrightarrow \mathbb{R}_+$$

loss function

$$\ell\left(\theta(P), \hat{\theta}(X)\right)$$

Minimax risk

worst distribution

$$\mathcal{M}_n(\Pi, \ell) := \underbrace{\inf_{\hat\theta}}_{\text{best estimator}} \sup_{P \in \Pi} \mathbb{E}_{\bar{X} \sim P^n} \, \ell\left(\hat\theta(\bar{X}), \theta(P)\right)$$

e.g. $\quad \Theta = \mathcal{V} \qquad \Pi = \{P_v\}_{v \in \mathcal{V}} \qquad \ell(\theta, \hat\theta) = \begin{cases} 1 & \theta \neq \hat\theta \\ 0 & \theta = \hat\theta \end{cases}$

$\theta(P_v) = v$

$$\mathcal{M}_n(\Pi, \ell) = \inf_{\hat\theta} \sup_{P_v} \mathbb{E}_{\bar{X} \sim P_v^n} \mathbb{P}\left[\hat\theta(\bar{X}) \neq v\right]$$

$$\geq \inf_{\hat\theta} \underbrace{\mathbb{E}_{v \in \mathcal{V}} \mathbb{E}_{\bar{X} \sim P_v^n} \mathbb{P}\left[\hat\theta(\bar{X}) \neq v\right]}_{P_e}$$

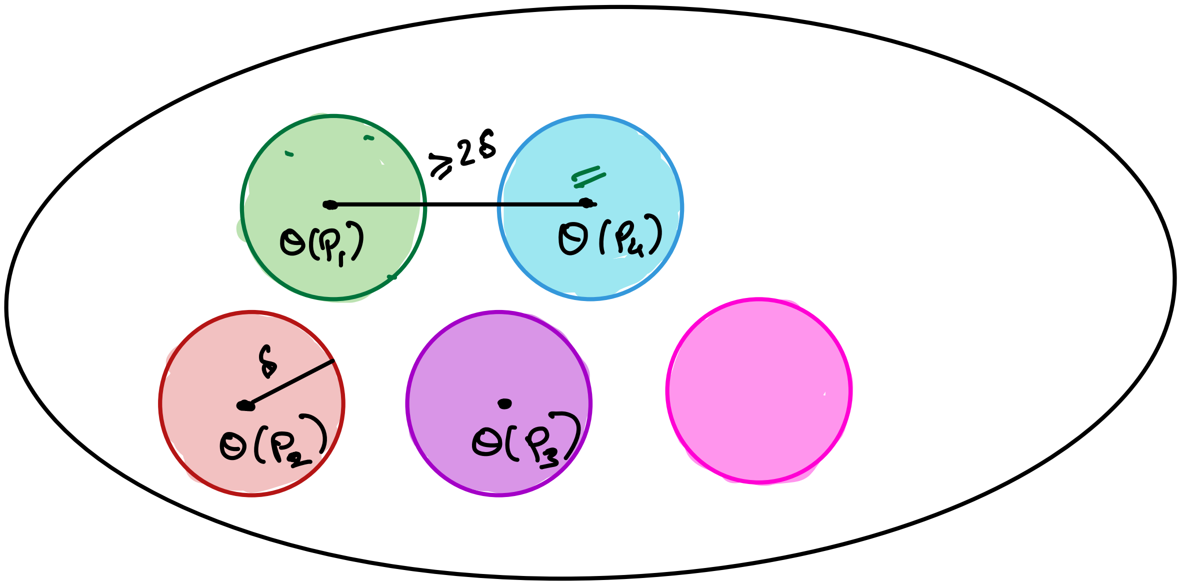
for Fano

Minimax lower bounds via hypothesis testing

$$\ell(\theta, \hat{\theta}) = \Phi(\rho(\theta, \hat{\theta}))$$

distance metric

increasing function



Cost of a mistake $\geq \Phi(\delta)$

- Let $\{P_v\}_{v \in \mathcal{V}} \subseteq \Pi$ be a finite set of distributions s.t.

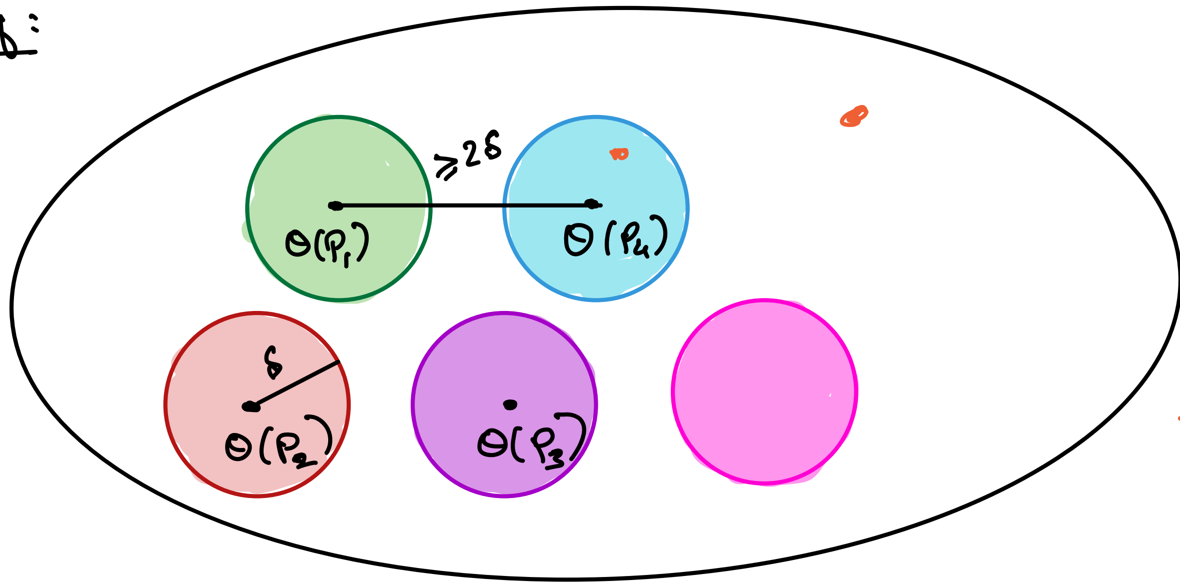
$$\rho(\theta(P_{v_1}), \theta(P_{v_2})) \geq 2\delta \quad \forall v_1 \neq v_2$$

$(\ell = \Phi(\rho(\theta, \hat{\theta})))$

Then, $\mathcal{M}_n(\Pi, \ell) \geq \Phi(\delta) \cdot \inf_T \left\{ \mathbb{E}_{v \in \mathcal{V}} \mathbb{P}_{\bar{x} \sim P_v^n} [T(\bar{x}) \neq v] \right\}$

$T : \mathcal{X}^n \to \mathcal{V}$

<u>Proof:</u>



Estimator $\hat{\theta}$ can be converted to predictor T

$$T = \underset{v \in \mathcal{V}}{\operatorname{argmin}} \ \rho(\theta(P_v), \hat{\theta})$$

$$M_n(\Pi, \ell) \geq \inf_{\hat{\theta}} \sup_{P_v} \mathbb{E}_{\bar{X} \sim P_v^n} \left[\Phi\left(\rho\left(\theta(P_v), \hat{\theta}(\bar{X}) \right) \right) \right]$$

$$\geq \inf_{\hat{\theta}} \mathbb{E}_{v \in \mathcal{V}} \mathbb{E}_{\bar{X} \sim P_v} \left[\underbrace{\Phi\left(\rho\left(\theta(P_v), \hat{\theta}(\bar{X}) \right) \right)}_{\geq \Phi(\delta) \cdot \mathbb{1}_{\{T(\bar{X}) \neq v\}}} \right]$$

One dimensional mean estimation

$\mathcal{X} = \{0, 1\}$ $\quad\quad \Pi$ = all distributions on $\mathcal{X}$

$$\theta(P) = \mathbb{E}_{x \sim P}[x] = P(1)$$

$$\ell(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2 = \left(\underbrace{|\theta - \hat{\theta}|}_{\rho}\right)^2$$

Empirical estimator: $\hat{\theta}(\bar{x}) = \hat{\theta}(x_1 \ldots x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$

Say $\theta(P) = \mu$

$$\text{error} = \mathbb{E}_{\bar{x} \sim P^n}\left[(\hat{\theta}(\bar{x}) - \mu)^2\right] = \frac{\mu(1-\mu)}{n}$$

Ex:

Lower bound (Le Cam's method)

$$P_0 = \begin{cases} 1 & \text{w.p. } 1/2 \\ 0 & \text{w.p. } 1/2 \end{cases} \qquad P_1 = \begin{cases} 1 & \text{w.p. } 1/2 + 2\delta \\ 0 & \text{w.p. } 1/2 - 2\delta \end{cases} \qquad (\mathcal{V} = \{0,1\})$$

$$\mathcal{M}_n(\Pi, \ell) \geq \Phi(\delta) \cdot \inf_{T} \mathbb{E}_{v \in \mathcal{V}} \mathbb{P}_{\bar{x} \sim P_v^n} [T(\bar{x}) \neq v]$$

$$\geq \delta^2 \cdot \inf_{T} \left\{ \frac{1}{2} \mathbb{P}_{\bar{x} \sim P_0^n} [T(\bar{x}) = 1] + \frac{1}{2} \cdot \mathbb{P}_{\bar{x} \sim P_1^n} [T(\bar{x}) = 0] \right\}$$

$$\geq \delta^2 \cdot \frac{1}{2} \left(1 - \frac{1}{2} \sqrt{2 \ln 2 \cdot n \cdot \underbrace{D(P_0 \| P_1)}_{C \cdot \delta^2}} \right)$$

Choosing δ

$$\mathcal{M}_n(\Pi, \ell) \geq \frac{\delta^2}{2} \cdot \left(1 - \frac{1}{2}\|P_0^n - P_1^n\|\right)$$

$$\geq \underbrace{\frac{\delta^2}{2}}_{\frac{(c')^2}{2}\frac{1}{n}} \left(1 - \frac{1}{2} \cdot \sqrt{2\ln 2 \cdot n \cdot C \cdot \delta^2}\right)$$

$\leq 1/2$

$\delta = c' \cdot \frac{1}{\sqrt{n}}$

High-dimensional mean estimation

$$\Pi = \{ N(\underbrace{\mu}_{\theta}, I_d) \mid \mu \in \mathbb{R}^d \} \qquad \ell(\theta, \hat{\theta}) = \| \theta - \hat{\theta} \|^2$$

Empirical estimator: $\hat{\theta}(\bar{x}) = \hat{\theta}(x_1 \ldots x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$

$$\text{error} = \mathbb{E}_{\bar{x} \sim (N(\mu, I_d))^n} \left[\| \hat{\theta}(\bar{x}) - \mu \|^2 \right]$$

$$= \frac{d}{n}$$