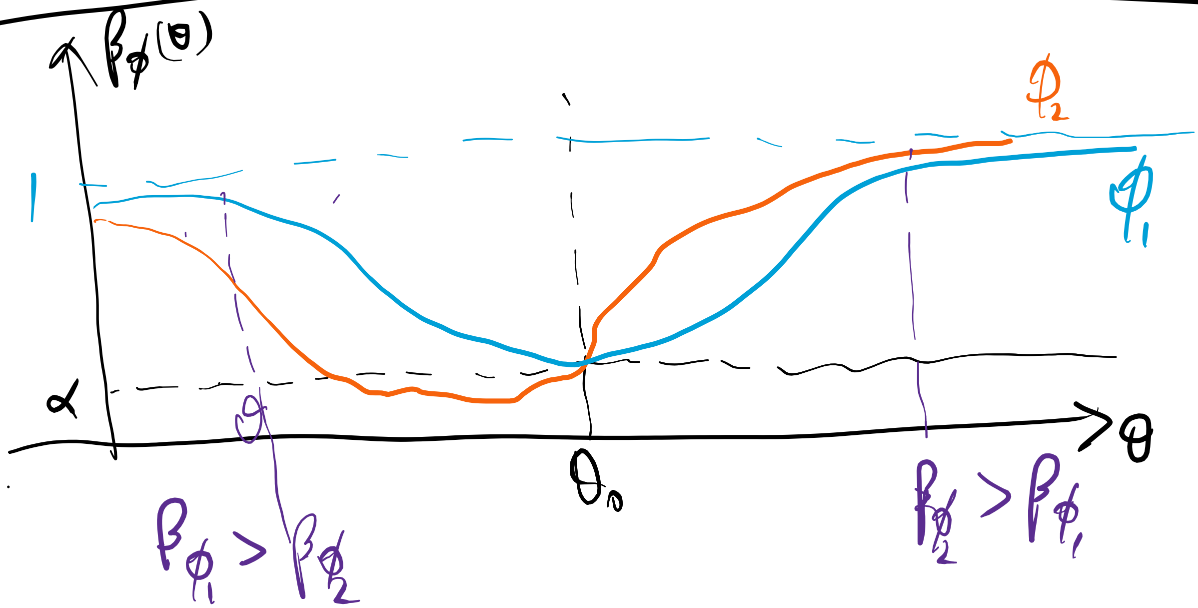
STA3000 F   Lecture 5

Homework 1 due 10/10 11:59pm

Recall



Unbiased test: $\beta_\phi(\theta) \geq \alpha \quad (\forall\, \theta \in \Theta)$.

(blue curve)

Analogous to unbiased estimation

$$\mathbb{E}_\theta\left[|\hat\theta - \theta|^2\right] \leq \mathbb{E}_\theta\left[|\hat\theta - \theta'|^2\right] \quad (\forall\, \theta' \in \Theta)$$

UMPU test: uniformly most powerful among unbiased tests

We restrict our attention to exponential families.

$$p_\eta(x) = \exp(\eta \cdot T(x) - A(\eta)) \cdot h(x).$$

Test : $H_0: \eta = \eta_0$ vs. $H_1: \eta \neq \eta_0$.

Unbiasedness $\iff$ $\beta_\phi(\cdot)$ minimized at η_0

$$\iff \frac{d}{d\eta} \mathbb{E}_\eta[\phi(X)] \Big|_{\eta=\eta_0} = 0$$

$$\parallel$$

$$\mathrm{cov}_{\eta_0}\big(T(X), \phi(X)\big) = \mathbb{E}_{\eta_0}\Big[\big(T(X) - \mathbb{E}(T(X))\big)\phi(X)\Big]$$

(Intuitively you need two equations to solve out the two thresholds.)

Want to construct

$$\phi^*(X) = \begin{cases} 0 & T(X) \in (C_1, C_2) \\ 1 & T(X) > C_2 \quad \text{or} \quad T(X) < C_1 \\ \gamma_i & T(X) = C_i \quad \text{for } i = 1, 2. \end{cases}$$

by solving

$$\begin{cases} \mathbb{E}_{\eta_0}[\phi^*(X)] = \alpha \\ \mathbb{E}_{\eta_0}\big[T(X)\,(\phi^*(X) - \alpha)\big] = 0. \end{cases}$$

Thm: Suppose $\eta_0 \in \text{int}\, \eta(\Theta)$

$\forall \alpha \in (0,1)$, $\exists\, \phi^*$ satisfying the equations.

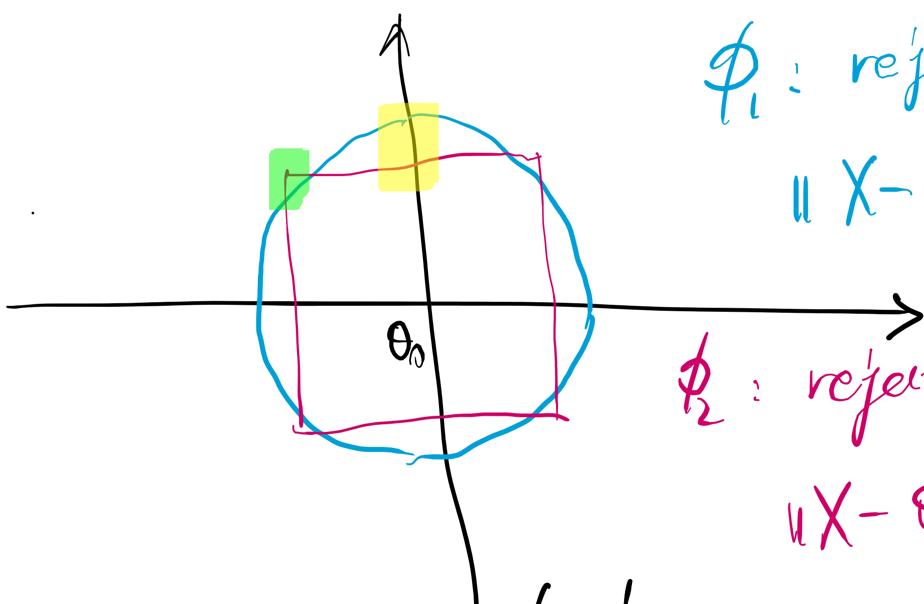
and ϕ^* is UMPU.

(Proof based on generalised NP lemma).

Multivariate testing. $X \sim N(\theta, \sigma^2 I_d)$ $(d \geq 2)$

$H_0: \theta = \theta_0$ $H_1: \theta \neq \theta_0$.

e.g. $d=2$



ϕ_1: reject when $\|X - \theta_0\|_2 \geq \varepsilon_1$

ϕ_2: reject when $\|X - \theta_0\| \geq \varepsilon_2$.

Cannot compare uniformly.

"minimax testing".

$H_0: \theta \in \Theta_0$

$H_1: \theta \in \Theta_1(\varepsilon) = \Theta_1 \cap \{\theta : \text{dist}(\theta, \Theta_0) \geq \varepsilon\}$

(distance up to user's choice)

$$R_\varepsilon(\phi) := \sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\phi(X)] + \sup_{\theta \in \Theta_1(\varepsilon)} \mathbb{E}_\theta[1 - \phi(X)].$$

Question: find an optimal ϕ^*

s.t. achieve $R_\varepsilon(\phi) \le \delta$ (user specified)

with smallest possible ε.

(Intuition: detect signal at minimal possible signal strength).

e.g. $X_1, X_2, \cdots X_n \overset{iid}{\sim} N(\theta, 1)$.

$H_0: \theta = 0$ v.s. $H_1: \theta \neq 0$.

$\varepsilon_n^* = \frac{c}{\sqrt{n}}$ (c is a constant depending on δ).

$X_1, \cdots, X_n \overset{iid}{\sim} N(\theta, I_d)$.

$H_0: \theta = 0$ v.s. $H_1: \theta \neq 0$.

$H_1(\varepsilon): \|\theta\|_2 \geq \varepsilon$.

First attempt: $\hat{\theta}_n := \frac{1}{n}\sum_{i}^{n} X_i$.

$$\|\hat{\theta}_n - \theta\|_2 \le c\sqrt{\frac{d}{n}} \quad \text{w.h.p.}$$

when $\|\theta\|_2 \ge 3c\sqrt{\frac{d}{n}}$, we have

$\theta \in \Theta(\varepsilon)$: $\|\hat{\theta}_n\|_2 \ge \|\theta\|_2 - \|\hat{\theta}_n - \theta\|_2 \ge 2c\sqrt{\frac{d}{n}}$. w.h.p.

$\theta \in \Theta_0$: $\|\hat{\theta}_n\|_2 \le c\sqrt{\frac{d}{n}}$ w.h.p.

Conjecture : $\varepsilon_n \asymp \sqrt{\frac{d}{n}}$ is minimax?

Thm: the minimax testing radius is

$$\varepsilon_n^* \asymp \frac{d^{1/4}}{n^{1/2}}.$$

Proof: Part I: construct a test to achieve this.

$$Y = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\theta, I_d/n),$$

$$\phi(x) = \mathbb{1}\{\|Y\|_2 \ge \text{threshold}\}.$$

$$\mathbb{E}_\theta[\|Y\|_2^2] = \frac{d}{n} + \|\theta\|_2^2$$

$$\mathrm{var}_\theta(\|Y\|_2^2) = \sum_{j=1}^{d} \mathrm{var}_\theta(Y_j^2)$$

$$= \frac{4}{n}\|\theta\|_2^2 + \frac{2d}{n^2}$$

Fluctuations in $\|Y\|_2^2 \ll$ magnitude of $\|Y\|_2^2$.

Let $C(a) = \mathbb{E}_0[\|Y\|_2^2] + a \cdot \sqrt{\mathrm{var}_0(\|Y\|_2^2)}$

(high-prob bound for $\|Y\|_2^2$ under H_0)

$$= \frac{d}{n} + a \cdot \sqrt{\frac{2d}{n^2}}.$$

By choosing $\phi(X) = \mathbb{1}\{\|Y\|_2^2 \geq C(a)\}$, we have.

$$\mathbb{E}_0[\phi(X)] = \mathbb{P}\left(\|Y\|_2^2 \geq \mathbb{E}_0[\|Y\|^2] + a \cdot \sqrt{\mathrm{var}_0(\|Y\|^2)}\right).$$

$$\leq \frac{1}{a^2} \quad \text{(by Chebyshev).}$$

For $\theta \neq 0$,

$$\mathbb{E}_\theta[1 - \phi(X)] = \mathbb{P}_\theta(\|Y\|_2^2 \leq C).$$

$$= \mathbb{P}_\theta\left(\|Y\|_2^2 - \mathbb{E}_\theta[\|Y\|_2^2] \le \frac{d}{n} + a\sqrt{\frac{2d}{n^2}} - \mathbb{E}_\theta[\|Y\|_2^2] \right)$$

$$\mathbb{E}_\theta[\|Y\|_2^2] = \frac{d}{n} + \|\theta\|_2^2$$

$$= \mathbb{P}_\theta\left(\|Y\|_2^2 - \mathbb{E}_\theta[\|Y\|_2^2] \le a\sqrt{\frac{2d}{n^2}} - \|\theta\|_2^2 \right)$$

(for $\|\theta\|_2 \ge \sqrt{a}\,\frac{(2d)^{1/4}}{n^{1/2}}$, by Chebyshev)

$$\le \frac{\frac{4}{n}\|\theta\|_2^2 + \frac{2d}{n^2}}{\left(\|\theta\|_2^2 - a\sqrt{\frac{2d}{n^2}}\right)^2}$$

(want to)

$$\le \frac{1}{a^2}.$$

$$\Longleftrightarrow \quad \|\theta\|_2^2 - a\sqrt{\frac{2d}{n^2}} \ge a\cdot\sqrt{\frac{4}{n}\|\theta\|_2^2 + \frac{2d}{n^2}}.$$

Solve for $\|\theta\|_2^2$. we require

$$\|\theta\|_2^2 \ge C'(a)\cdot\frac{\sqrt{d}}{n}.$$

So for this test,

$$\mathbb{E}_0[\phi(X)] + \sup_{\theta\in\Theta_1(\varepsilon)} \mathbb{E}_\theta[1-\phi(X)] \le \frac{2}{a^2}$$

whenever $\varepsilon \ge \sqrt{C'(a)} \cdot \frac{d^{1/4}}{n^{1/2}}$

Part II: Lower bound.

For any test ϕ, we want to show

$$\mathbb{E}_0[\phi(X)] + \sup_{\theta\in\Theta_1(\varepsilon)} \mathbb{E}_\theta[1-\phi(X)] \ge \frac{2}{3}.$$

when $\varepsilon \le c_1 \frac{d^{1/4}}{n^{1/2}}$.

First attempt: choose $\theta_1 \in \Theta_1(\varepsilon)$

$H_0: \theta = 0$ v.s. $H_1: \theta = \theta_1$.

Essentially 1-dim.

$$\phi(X) = \mathbb{1}\left\{ \frac{\bar{X}_n^T \theta_1}{\|\theta_1\|_2} \ge \text{threshold} \right\}.$$

radius $1/\sqrt{n}$.

$$\mathbb{E}_0[\phi(X)] + \sup_{\theta \in \Theta_1(\varepsilon)} \mathbb{E}_\theta[1-\phi(X)]$$

$$\geq \mathbb{E}_0[\phi(X)] + \int \mathbb{E}_\theta[1-\phi(X)]\,\pi(d\theta)$$

(for π s.t. $\mathrm{supp}(\pi) \subseteq \Theta_1(\varepsilon)$)

This corresponds to type I + type II error for

$H_0: \theta = 0$ v.s. $H_1': \theta \sim \pi$.

(under H_1', observations follow mixture of product)

Construction of π

- Uniform on $\varepsilon \cdot S^{d-1}$ $= \{\theta : \|\theta\|_2 = \varepsilon\}$.

  (this works, but calculation is complicated)

- $\pi = \mathrm{Unif}\left(\left\{\frac{\pm\varepsilon}{\sqrt{d}}, \frac{\pm\varepsilon}{\sqrt{d}}, \dots, \frac{\pm\varepsilon}{\sqrt{d}}\right\}\right)$

  i.e. each coordinate indp choosing

  $\frac{\varepsilon}{\sqrt{d}}$ or $\frac{-\varepsilon}{\sqrt{d}}$ w.p. $1/2$.

i.e. corner points of d-hypercube.

From last lecture.

$$\mathbb{E}_0[\phi(X)] + \mathbb{E}_{H_1}[1-\phi(X)]$$

$$\geq 1 - \underbrace{d_{TV}(P_0, P_1)}_{\text{want to be } < 1/3}$$

(where P_1 is the prob distr of obs under H_1).

$$d_{TV}(P_0, P_1) = \mathbb{E}_{P_0}\left[\left|\frac{dP_1}{dP_0}(X) - 1\right|\right].$$

(Cauchy-Schwarz)

$$\leq \sqrt{\mathbb{E}_{P_0}\left[\left|\frac{dP_1}{dP_0}(x) - 1\right|^2\right]}$$

$$= \sqrt{\underbrace{\mathbb{E}_{P_0}\left[\left|\frac{dP_1}{dP_0}(x)\right|^2\right] - 1}_{\chi^2(P_1 \| P_0)}}.$$

$$L(X_1^n) := \frac{dP_1}{dP_0}(X_1, \ldots, X_n)$$

$$= \frac{\mathbb{E}_{z \sim \mathrm{Unif}(\{\pm 1\}^d)}\left[\exp\left(\frac{-1}{2}\sum_{1}^{n}\| X_i - \frac{\varepsilon z}{\sqrt{d}}\|_2^2\right)\right]}{\exp\left(-\frac{1}{2}\sum_{1}^{n}\|X_i\|^2\right)}.$$

$$= \mathbb{E}_{z \sim \mathrm{Unif}(\{\pm 1\}^d)}\left[\exp\left(\frac{\varepsilon}{\sqrt{d}} \cdot n \bar{X}_n^T z - \frac{\varepsilon^2 n}{2}\right)\right]$$

(here, we treat $X_1 \dots X_n$ as deterministic)

Now we take 2nd moment

$$\mathbb{E}_0\left[L(X_1^n)^2\right] = \mathbb{E}_0\left[\left|\mathbb{E}\left[\cdots \mid X_1^n\right]\right|^2\right].$$

$$= \mathbb{E}_0 \mathbb{E}_{z,z' \overset{iid}{\sim} \mathrm{Unif}(\{\pm 1\}^d)}\left[\exp\left(\frac{\varepsilon}{\sqrt{d}} n \bar{X}_n^T (z+z') - \varepsilon^2 n\right)\right].$$

(MGF of Gaussian)

$$= \mathbb{E}_{z,z' \overset{iid}{\sim} \mathrm{Unif}(\{\pm 1\}^d)}\left[\exp\left(\frac{n\varepsilon^2}{2d}\|z+z'\|_2^2 - n\varepsilon^2\right)\right]$$

$$= \mathbb{E}_{\dots}\left[\exp\left(\frac{n\varepsilon^2}{d} z^T z'\right)\right].$$

$$= \left(\frac{1}{2}\exp\left(\frac{n\varepsilon^2}{d}\right) + \frac{1}{2}\exp\left(\frac{-n\varepsilon^2}{d}\right)\right)^d$$

$$\leq \exp\left(\frac{1}{2}\left(\frac{n\varepsilon^2}{d}\right)^2 \cdot d\right)$$

$$= \exp\left(\frac{n^2\varepsilon^4}{2d}\right) \underset{\text{to}}{\overset{\text{(want)}}{\leq}} \frac{10}{9}.$$

Solve for ε. $\quad \varepsilon \leq \frac{d^{1/4}}{10 n^{1/2}}$.

Generally. 1 v.s. mixture } Lower bound construction.

mixture v.s. mixture

Back to estimation.

Def (M-estimation / ERM / SAA)

$$\hat{\theta}_n := \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} f(\theta; X_i).$$

(sometimes w/ regularization).

Population-level loss function:

(assuming $X_1 \cdots X_n \overset{iid}{\sim} \mathbb{P}$).

$$F(\theta) := \mathbb{E}\left[f(\theta; X)\right].$$

$$\theta^* := \arg\min \{F(\theta)\}.$$

Criteria of evaluation.

- $F(\hat{\theta}_n) - F(\theta^*)$ (func value)
- $\|\hat{\theta}_n - \theta^*\|$ (param estimation)

e.g. MLE, regression, classification, ...

$$F_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} f(\theta; X_i)$$

By LLN / concentration ineq.

$$|F_n(\theta) - F(\theta)| \begin{cases} \xrightarrow{P} 0. \\ \leq O\left(\sqrt{\frac{\log(1/\delta)}{n}}\right) \end{cases}$$ (in bounded case).

(for fixed θ)

$$F(\hat{\theta}_n) - F(\theta^*)$$
$$= \underbrace{\boxed{\left(F(\hat{\theta}_n) - F_n(\hat{\theta}_n)\right)}}_{\text{hard.}} + \underbrace{\left(F_n(\hat{\theta}_n) - F_n(\theta^*)\right)}_{\leq 0} + \underbrace{\left(F_n(\theta^*) - F(\theta^*)\right)}_{\text{easy}}$$

$(F - F_n)$ evaluated are random $\hat{\theta}_n$ depending on F_n

$$\left|F(\hat{\theta}_n) - F_n(\hat{\theta}_n)\right| \leq \sup_{\theta \in \Theta} \left|F(\theta) - F_n(\theta)\right|.$$

We need ULLN

$$\sup_{\theta \in \Theta} \underbrace{\left|F(\theta) - F_n(\theta)\right|}_{\text{"empirical process"}} \xrightarrow{p} 0.$$

"empirical process theory" aims at

- uniform convergence
- Non-asymptotic bounds
- Limiting distribution

of empirical processes (and their suprema).

Assuming ULLN, we have

$$F(\hat{\theta}_n) \longrightarrow F(\theta^*).$$

For parameter estimation, need assumption

Corollary: suppose $\forall \varepsilon > 0$,

$$\inf_{\|\theta - \theta^*\| \geq \varepsilon} F(\theta) > F(\theta^*).$$

then ULLN implies $\hat{\theta}_n \xrightarrow{P} \theta^*$.

Proof: suppose $\inf\limits_{\dots} F(\theta) \geq F(\theta^*) + \delta$ for some $\delta > 0$

$$\mathbb{P}(\|\hat{\theta}_n - \theta^*\| > \varepsilon) \leq \mathbb{P}(F(\hat{\theta}_n) \geq F(\theta^*) + \delta) \xrightarrow{P} 0.$$

When does ULLN hold true?

- $|\Theta| < +\infty$.

$$\mathbb{P}\left(\sup_{\theta \in \Theta} |F_n(\theta) - F(\theta)| > \varepsilon\right)$$

$$\leq \sum_{\theta \in \Theta} \mathbb{P}(|F_n(\theta) - F(\theta)| > \varepsilon) \longrightarrow 0.$$

For infinite/cts Θ. idea: discretization.

· "covering #".

Given set K, metric ρ, $\varepsilon > 0$.

$$N(K;\rho,\varepsilon) := \min\left\{ N : \exists \{\theta_i\}_{i=1}^{N} \subseteq K, \text{ s.t. } K \subseteq \bigcup_{i=1}^{N} \mathbb{B}_\rho(\theta_i,\varepsilon) \right\}$$

· "Packing #"

$$M(K;\rho,\varepsilon) = \max\left\{ M : \exists \{\theta_i\}_{i=1}^{M} \subseteq K \text{ s.t. } \rho(\theta_i,\theta_j) \geq \varepsilon \right\}.$$

Thm (duality)

$$M(K;\rho,2\varepsilon) \leq N(K;\rho,\varepsilon) \leq M(K;\rho,\varepsilon).$$