STA355 Lecture 8

General testing problem $(\theta \in \Theta)$

$H_0: \theta \in \Theta_0$ v.s. $H_1: \theta \notin \Theta_0$.

$(\Theta_0 \subseteq \Theta)$.

- Analogy to estimation: MLE
- For testing: Likelihood Ratio Test.

Idea: compare MLE under Θ_0 and Θ.

the likelihood values of

If H_0 is false and θ's bounded away from Θ_0

Then $$\max_{\theta \in \Theta} \sum_{i=1}^{n} \log p_\theta(X_i) \gg \max_{\theta \in \Theta_0} \sum_{i=1}^{n} \log p_\theta(X_i).$$

- LHS $\geq$ RHS always true

- For $\theta \notin \Theta_0$

$$\text{LHS} \approx n \cdot \max_{\theta \in \Theta} \mathbb{E}[\log p_\theta(X)]$$

$$\text{RHS} \approx n \cdot \max_{\theta \in \Theta_0} \mathbb{E}[\log p_\theta(x)].$$

$$\text{LHS} - \text{RHS} \asymp n.$$

- For $\theta \in \Theta_0$.

$$\text{LHS} - \text{RHS} = \Theta(1).$$

$$\lambda := 2\log\left(\sup_{\theta \in \Theta} p_\theta(X_1^n)\right) - 2\log\left(\sup_{\theta \in \Theta_0} p_\theta(X_1^n)\right)$$

Reject when $\lambda \geq \lambda_\alpha$.

Thm: $\Theta = \mathbb{R}^d$, Θ_0 $(d-k)$- dim linear subspace.

(e.g. Θ_0: $\theta_1 = z_1$, $\theta_2 = z_2$, $\cdots$, $\theta_k = z_k$
for some fixed $z_1, z_2, \cdots z_k$).

under the null, assuming some regularity conditions (non-degenerate Fisher info, 2nd order smoothness)

(Basically conditions needed for asymp norm of MLE).

We have $\lambda \xrightarrow{d} \chi^2_k$.

(a χ^2_k distn is the prob. distr. of $\sum_{i=1}^{k} g_i^2$ where $g_i \sim N(0,1)$).

Operationally.

- Compute unconstrained $\hat{\theta}_n$ and constrained $\hat{\theta}_{n,0}$
- Compute λ by comparing the likelihood values.
- Reject when $\lambda \geq (1-\alpha)$ quantile of χ^2_k.

Proof idea: Taylor expansion (for $|\theta^* - \theta| = O_p(\frac{1}{\sqrt{n}})$).

want to use in testing

$$\sum_{i=1}^n \log p_{\theta^*}(X_i) = \sum_{i=1}^n \log p_\theta(X_i) + \left(\sum_{i=1}^n \nabla \log p_\theta(X_i)\right)^T (\theta^* - \theta) + \frac{1}{2}(\theta^* - \theta)^T \left(\sum_1^n \nabla^2 \log p_\theta(X_i)\right)(\theta^* - \theta) + o_p(1).$$

Fixed. indep of θ.

For unconstrained MLE, $\sum_1^n \nabla \log p_{\hat\theta_n}(X_i) = 0$

and $\sqrt{n}(\hat\theta_n - \theta^*) \xrightarrow{d} N(0, I(\theta^*)^{-1})$.

We also have

$$\frac{1}{n}\sum_{i=1}^n \nabla^2 \log p_{\hat\theta_n}(X_i) \xrightarrow{P} \mathbb{E}_{\theta^*}[\nabla^2 \log p_{\theta^*}(X)]$$
$$\mathbb{E}_{\theta^*}[\nabla^2 \log p_{\theta^*}(X)] = -I(\theta^*).$$

$$\sum_1^n \log p_{\hat\theta_n}(X_i) = \sum_1^n \log p_{\theta^*}(X_i) + \frac{n}{2}(\hat\theta_n - \theta^*)^T I(\theta^*)(\hat\theta_n - \theta^*) + o_p(1)$$

For constrained case, $\hat{\theta}_{n,0} = \underset{\theta \in \Theta_0}{\arg\max} \{ \cdots \}$.

$$\left.\begin{array}{l} \sum_{i}^{n} \nabla \log p_{\hat{\theta}_{n,0}}(X_i) \perp \Theta_0 \\ \hat{\theta}_{n,0},\ \theta^* \in \Theta_0 \end{array}\right\} \to \text{inner product} = 0.$$

$$\sum_{i}^{n} \log p_{\hat{\theta}_{n,0}}(X_i) = \sum_{i}^{n} \log p_{\theta^*}(X_i)$$

$$+ \frac{n}{2} (\hat{\theta}_{n,0} - \theta^*)^T I_\theta(\theta^*) \cdot (\hat{\theta}_{n,0} - \theta^*).$$

$$\sqrt{n}(\hat{\theta}_{n,0} - \theta^*) \xrightarrow{d} N(0, I_0(\theta^*)^{+})$$

(Pseudo-inverse).

$$\lambda \approx n \cdot (\hat{\theta}_n - \theta^*)^T I(\theta^*) (\hat{\theta}_n - \theta^*)$$

$$- n (\hat{\theta}_{n,0} - \theta^*)^T I_0(\theta^*) (\hat{\theta}_{n,0} - \theta^*).$$

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \approx I(\theta^*)^{-1/2} \cdot \xi \qquad \xi \sim N(0, I_d)$$

$$\sqrt{n}(\hat{\theta}_{n,0} - \theta^*) \approx I(\theta^*)^{-1/2} \cdot P_{\Theta_0} \xi$$

where P_{Θ_0} is orthogonal projection onto Θ_0

So we get

$$\lambda \approx \|\xi\|_2^2 - \|P_{\Theta_0}\xi\|_2^2 = \|P_\perp \xi\|_2^2 \xrightarrow{d} \chi_k^2.$$

Multivariate testing radius.

Detection problem:

$$H_0: \theta = \theta_0$$

$$H_1: \|\theta - \theta_0\| > \varepsilon$$

Want to find smallest possible $\varepsilon > 0$ s.t. we can make both types of errors small.

$X_1, X_2, \cdots, X_n \overset{iid}{\sim} N(\theta, I_d)$. $H_0: \theta = 0$.

Test statistics: $T = \|\frac{1}{n}\sum_{i=1}^{n} X_i\|_2^2$

$n \cdot T \sim \chi_d^2$

Reject when $n \cdot T > (1-\alpha)$ quantile of χ_d^2.

Question: what is the order of magnitude of ε s.t. we can make type-II err small as well?

(as a function of n and d)

First guess: $\sqrt{\frac{d}{n}}$?

Since $\mathbb{E}\left[\|\bar{X}_n - \theta\|_2^2\right] = \frac{d}{n}$, we have $\|\bar{X}_n - \theta\|_2 \le C\sqrt{\frac{d}{n}}$ w.h.p.

When $\varepsilon > 2C\cdot\sqrt{\frac{d}{n}}$,

under H_1, $\|\bar{X}_n\|_2 \ge \|\theta\|_2 - \|\bar{X}_n - \theta\|_2$

$\ge C\cdot\sqrt{\frac{d}{n}}$ (w.h.p.).

So we can show that

$$\mathbb{P}_\theta(\text{reject}) \geq \frac{3}{4} \quad \text{when } \|\theta\|_2 > \varepsilon.$$

For testing, we can achieve $\varepsilon = c \cdot \frac{d^{1/4}}{n^{1/2}}$.

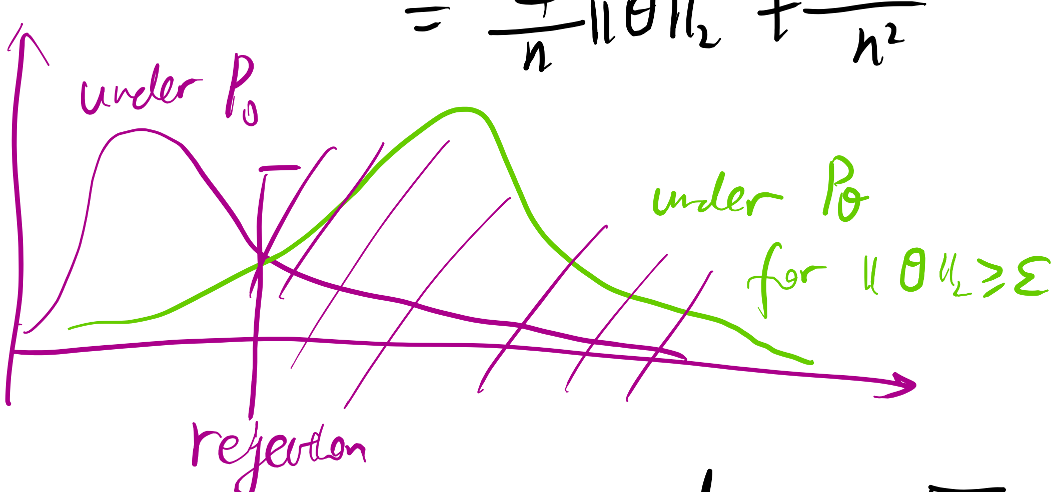
(Detecting a signal is easier than estimating a signal).

Proof idea: $T = \|\frac{1}{n}\sum_{i=1}^{n} X_i\|_2^2$ $\qquad \frac{1}{n}\sum_{1}^{n} X_i \sim \mathcal{N}(\theta, \frac{I_d}{n})$

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

$$\mathrm{Var}_\theta(T) = \sum_{j=1}^{d} \mathrm{var}\left(\left(\frac{1}{n}\sum_{1} X_{ij}\right)^2\right)$$

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



$$\text{Rejection threshold} = \frac{d}{n} + a\frac{\sqrt{2d}}{n}$$

Under null. by Chebyshev ineq,

$$\mathbb{P}_0(\text{reject}) = \mathbb{P}_0\left(T > \frac{d}{n} + a\frac{\sqrt{2d}}{n}\right)$$

$$\leq \frac{\mathrm{Var}_0(T)}{a\cdot\left(\frac{\sqrt{2d}}{n}\right)^2} = \frac{1}{a^2}.$$

Under alternative:

$$\mathbb{P}_\theta(\text{not reject}) = \mathbb{P}_\theta\left(T \leq \frac{d + a\sqrt{2d}}{n}\right).$$

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

assuming $\varepsilon > \left(\frac{a\sqrt{2d}}{n}\right)^{1/2}$

$$\leq \frac{\mathrm{var}_\theta(T)}{\left(\|\theta\|_2^2 - \frac{a\sqrt{2d}}{n}\right)^2} = \frac{\frac{2d}{n^2} + \frac{4\|\theta\|^2}{n}}{\left(\|\theta\|_2^2 - \frac{a\sqrt{2d}}{n}\right)^2}$$

(want to show $\leq \frac{1}{a^2}$)

Need to have:

$$\|\theta\|_2^2 - \frac{a\sqrt{2d}}{n} \geq a\sqrt{\frac{4}{n}\|\theta\|_2^2 + \frac{2d}{n^2}}$$

This is satisfied when

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

(where $C(a)$ is a const depending only on a).

Remark: testing is useful only when it detects something. We should care about type-II err.

Decision theory:

Motivation: How to compare estimators.

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

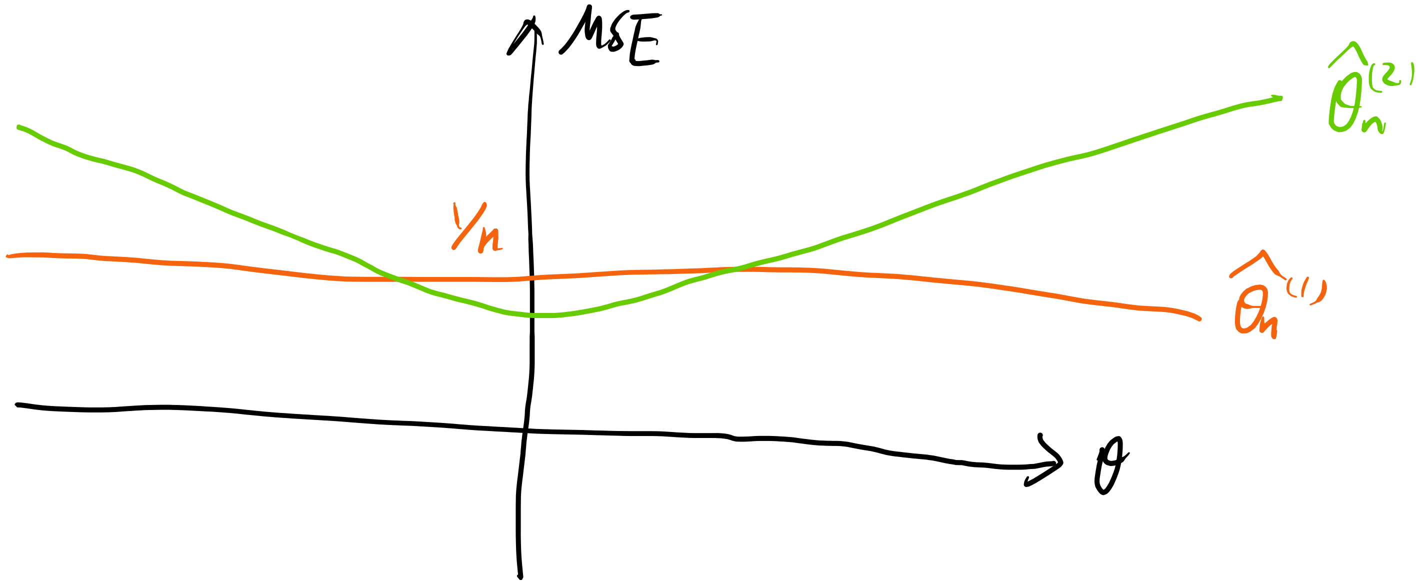
$$\hat{\theta}_n^{(1)} = \bar{X}_n \qquad \mathbb{E}[|\bar{X}_n - \theta|^2] = \frac{1}{n}.$$

$$\hat{\theta}_n^{(2)} = \frac{\sum_1^n X_i}{n+1}$$

$$\mathbb{E}\left[|\hat{\theta}_n^{(2)} - \theta|^2\right]$$

$$= \mathrm{bias}_\theta(\hat{\theta}_n^{(2)})^2 + \mathrm{var}_\theta(\hat{\theta}_n^{(2)})$$

$$= \left(\frac{\theta}{n+1}\right)^2 + \frac{n}{(n+1)^2} = \frac{n+\theta^2}{(n+1)^2}$$



Unable to uniformly compare.

Notations:

- Loss function $L(\theta, \hat{\theta}_n)$ (e.g. $\|\theta - \hat{\theta}_n\|^2$).
- Risk $R(\theta, \hat{\theta}_n) = \mathbb{E}_\theta[L(\theta, \hat{\theta}_n)]$. (e.g. MSE).

Def. "Bayes risk"

π "prior": a prob distribution on Θ

$$r_\pi(\hat{\theta}_n) = \int R(\theta, \hat{\theta}_n)\, \pi(\theta)\, d\theta.$$

"Bayes estimator":

$$\operatorname{argmin} \{ r_\pi(\hat{\theta}_n) \}.$$

Def. "minimax risk".

$$\bar{R}(\hat{\theta}_n) := \sup_\theta R(\theta, \hat{\theta}_n).$$

$$\hat{\theta}_n := \operatorname{argmin} \bar{R}(\hat{\theta}_n)$$

"minimax estimator".

How to find Bayes estimator?

Suppose $p_\theta(x)$ is the pdf (or pmf) of x under $\mathbb{P}_\theta$.

$$r_\pi(\hat{\theta}_n) = \int_\Theta \int_{\mathcal{X}} L(\theta, \hat{\theta}_n(X)) \cdot p_\theta(x) \cdot \pi(\theta)\, dx\, d\theta$$

$$= \int_{\mathcal{X}} \boxed{\int_\Theta L(\theta; \hat{\theta}_n(x)) \cdot p_\theta(x) \cdot \pi(\theta)\, d\theta}\, dx$$

For fixed $x \in \mathcal{X}$, we can choose $\hat{\theta}_n(x)$ to minimize it

$$\hat{\theta}_n(x) = \underset{\psi \in \Theta}{\arg\min} \int_\Theta L(\theta; \psi) \cdot p_\theta(x)\, \pi(\theta)\, d\theta$$

$$= \underset{\psi \in \Theta}{\arg\min} \frac{\int_\Theta L(\theta; \psi) \cdot p_\theta(x)\, \pi(\theta)\, d\theta}{\int_\Theta p_\theta(x) \cdot \pi(\theta)\, d\theta}$$

$$= \underset{\psi \in \Theta}{\arg\min} \int_\Theta L(\theta; \psi)\, \pi(\theta | x_1, \dots, x_n)\, d\theta$$

$$\pi(\theta|x_1 \cdots x_n) := \frac{\pi(\theta)\cdot p_\theta(x)}{\int_\Theta p_{\theta'}(x)\cdot \pi(\theta')\, d\theta'}.$$

"Posterior distribution".

Bayesian interpretation:

Assume $\theta \sim \pi$, $X|\theta \sim p_\theta$

By Bayes formula, we can conclude

$$p(\theta|x) = \frac{\pi(\theta)\cdot p_\theta(x)}{\int p_{\theta'}(x)\cdot \pi(\theta')\, d\theta'} = \pi(\theta|x).$$

Bayes estimator = Posterior risk minimizer.

For MSE

$$\hat{\theta} = \underset{\psi \in \Theta}{\arg\min} \int_\Theta \|\theta - \psi\|_2^2 \cdot \pi(\theta|x)\, d\theta$$

$$= \int_\Theta \theta \cdot \pi(\theta|x)\, d\theta.$$

Posterior mean.

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

$\pi = N(0,1)$.

$$\Pi(\cdot \mid X_1, \cdots, X_n) = N\left(\frac{\sum_1^n X_i}{n+1}, \frac{1}{n+1}\right).$$

So $\hat{\theta}_n^{(2)}$ at the beginning is the Bayes estimator for $\pi = N(0,1)$.

How to find minimax estimators?

Fact. const risk Bayes estimators are minimax.

$\forall \theta_1, \theta_2 \in \Theta \qquad R(\theta_1, \hat{\theta}_n) = R(\theta_2, \hat{\theta}_n) \left(= r_\pi(\hat{\theta}_n)\right).$

e.g. $X \sim Binom(n, p)$. known n, estimate p.

"natural choice": $\hat{p} = \frac{X}{n}$.

$$R(p; \hat{p}) = \frac{p(1-p)}{n}.$$