SDS 387

Linear Models

Fall 2025

Lecture 23 - Tue, Nov 18, 2025

Instructor: Prof. Ale Rinaldo

- HW 4, Q8 (b): $X \sim N_d(\mu, I_d)$ (or $X \sim N_d(\mu, \sigma^2 I_d)$)

- Today: statistical inference for β^*, assuming a linear model and fixed design:

$$\underset{n \times 1}{Y} = \underset{n \times d}{\Phi} \underset{d \times 1}{\beta^*} + \underset{n \times 1}{\varepsilon}$$

Φ: fixed

- Last time we saw that

$$\hat{\beta} \xrightarrow{P} \beta^*$$

$\hat{\beta}$: OLS estimator

i.e. $\hat{\beta}$ is consistent (if $\hat{\Sigma} \to \Sigma$)

is $\hat{\beta}_\lambda$ for ridge consistent? $\to$ HW

- Today we will show that

$$(*) \quad \sqrt{n}(\hat{\beta} - \beta^*) \xrightarrow{d} N_d(0, \sigma^2 \Sigma^{-1})$$

provided that $\hat{\Sigma} = \frac{\Phi^\top \Phi}{n} \to \Sigma$

- In fact, both claims are true under the random Φ settings, provided that

$$\frac{\Phi^T\Phi}{n} = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\Phi_i}_{\text{transpose of the } i\text{th row of } \Phi} \Phi_i^T \xrightarrow{P} \underset{d\times d}{\Sigma}, \text{ pd}$$

- To prove ★ notice that

$$\sqrt{n}(\hat{\beta}-\beta^*) = \sqrt{n}\left(\hat{\Sigma}^{-1}\frac{\Phi^T Y}{n} - \beta^*\right) = \sqrt{n}\left(\cancel{\underbrace{\hat{\Sigma}^{-1}\frac{\Phi^T\Phi}{n}}_{=I_d}\beta^*} + \hat{\Sigma}^{-1}\frac{\Phi^T\varepsilon}{n} - \cancel{\beta^*}\right)$$

$$= \sqrt{n}\,\hat{\Sigma}^{-1}\frac{\Phi^T\varepsilon}{n}$$

Next $\hat{\Sigma}^{-1} \to \Sigma^{-1}$. So we need to show that

$$\sqrt{n}\,\frac{\Phi^T\varepsilon}{n} \xrightarrow{d} N_d(0, \sigma^2\Sigma)$$

because the claim then follows by Slutsky's.

But we can write $\frac{\Phi^T\varepsilon}{n} = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\Phi_i}_{\text{transpose of } i\text{th row of } \Phi}\varepsilon_i$, $\begin{bmatrix}\varepsilon_1\\ \vdots \\ \varepsilon_n\end{bmatrix} \sim (0, \sigma^2 I_n)$

We need to check the LF conditions.

Notice that $\mathbb{E}[\Phi_i\varepsilon_i] = 0$ and

$\mathrm{Var}[\Phi_i\varepsilon_i] = \sigma^2\Phi_i\Phi_i^T$

So that

$$\sum_{i=1}^{n} \text{Var}\left[\frac{\Phi_i \varepsilon_i}{n}\right] = \sigma^2 \frac{\Phi^T \Phi}{n} \to \sigma^2 \Sigma$$

The LF conditions for this problem are:

$$\underbrace{\sum_{i=1}^{n} \mathbb{E}\left[\frac{\|\Phi_i \varepsilon_i\|^2}{n} \mathbb{1}\left\{\frac{\|\Phi_i \varepsilon_i\|}{\sqrt{n}} > \eta\right\}\right]} \to 0$$

as $n \to \infty$ and for each $\eta > 0$.

$$\leq \underbrace{\left[\sum_{i=1}^{n} \frac{\|\Phi_i\|^2}{n}\right]}_{\frac{tr(\Phi^T \Phi)}{n} \to tr(\Sigma_0)} \underbrace{\max_{i=1,\dots,n} \mathbb{E}\left[\varepsilon_i^2 \mathbb{1}\left\{\frac{\|\Phi_i \varepsilon_i\|}{\sqrt{n}} > \eta\right\}\right]}_{T}$$

We have that $T \to 0$ as $n \to \infty$ if

$$\max_i \frac{\|\Phi_i\|}{\sqrt{n}} \to 0 \quad \text{as} \quad n \to \infty$$ check this!

So if $\frac{\Phi^T \Phi}{n} \to \Sigma_0$ and $\max_i \frac{\|\Phi_i\|}{\sqrt{n}} \to 0$

as $n \to \infty$ we have that

$$\sqrt{n}\,(\hat{\beta} - \beta^*) = \hat{\Sigma}^{-1} \sqrt{n}\, \frac{\Phi^T \varepsilon}{n} \xrightarrow{d} N_d(0, \sigma^2 \Sigma^{-1})$$

$$\to \Sigma^{-1} \qquad \underbrace{\sqrt{n}\, \frac{\Phi^T \varepsilon}{n}}_{\xrightarrow{d} N_d(0, \sigma^2 \Sigma_0)}$$

- From van der Vaart's book, Chapter 2: we could have assumed instead that for some $k \geq 1$

a) $$\mathbb{E}\left[|\varepsilon_i|^{2+k}\right] < \infty$$

b) $$\sum_{i=1}^{n} \frac{\|\Phi_i\|^k}{n^{k/2}} \to 0$$

because
$$\mathbb{E}\left[\varepsilon_i^2 \, \mathbb{1}\{|\varepsilon_i| a > \eta\}\right] \leq \mathbb{E}\left[|\varepsilon_i|^{2+k}\right] a^k \eta^{-k}$$

where
$$a = \frac{\|\Phi_i\|}{\sqrt{n}}$$

- Statistical inference: now that we have established asymptotic normality we should be carry out stat. inference (hypothesis testing, confidence intervals)

<u>Problem</u>: we do not know σ^2!

- Let's assume that $\varepsilon \sim N_n(0, \sigma^2 I_n)$. Therefore

$$Y \sim N_n(\Phi\beta^*, \sigma^2 I_n)$$

and

$$\sqrt{n}(\hat{\beta} - \beta^*) \sim N_d(0, \sigma^2 \hat{\Sigma}^{-1})$$

To estimate σ^2 we could use the residuals:

$$\underset{n \times 1}{e} = Y - \underbrace{\hat{Y}}_{\Phi\hat{\beta}} = Y - \underbrace{H}_{\text{hat matrix } \Phi(\Phi^T\Phi)^{-1}\Phi^T} Y = (I - H)Y$$

where H and $(I-H)$ are orthogonal projection matrices

with H projecting onto $C(\Phi)$
$\downarrow$ column space

Next,

$$e \sim N_n\left(0, \sigma^2 (I-H)\right) \quad \text{Exercise}$$

$$\left(\text{Of course } \hat{Y} \sim N_n\left(\Phi\beta^*, \sigma^2 H\right)\right)$$

We can think of e as an "estimator" of ε but the residuals are correlated and have different variances!

- Nonetheless $\|e\|^2 \sim \sigma^2 \chi^2_{n-d}$ HW problem

$$\hookrightarrow \mathbb{E}\left[\frac{\|e\|^2}{n-d}\right] = \sigma^2$$

$$\hookrightarrow \hat{\sigma}^2 = \frac{\|e\|^2}{n-d} \text{ is an unbiased estimator of } \sigma^2$$

$\downarrow$ degrees of freedom

- Furthermore $\hat{\sigma}^2 \perp\!\!\!\perp \hat{\beta}$ because $\hat{\sigma}^2$ is a function of $(I-H)Y$ and $\mathbb{E}[\hat{\beta} e^T] = 0_{d \times n}$

- At the end of the day:

$$j = 1, \dots, d \qquad \frac{\hat\beta_j - \beta^*_j}{se(\hat\beta_j)} = \frac{\dfrac{\hat\beta_j - \beta_j}{\sigma\sqrt{(\Phi^T\Phi)^{-1}_{j,j}}}}{\sqrt{\dfrac{\hat\sigma^2}{\sigma^2}}} \sim t_{n-d}$$

$$se(\hat\beta_j) = \sqrt{\hat\sigma^2 (\Phi^T\Phi)^{-1}_{j,j}}$$

ratio of a $N(0,1)$ and squared root of indep χ^2 divided by its degrees of freedom

- If the errors are not Gaussian, we can use the CLT and Slutsky's theorem to conclude that

$$\frac{\hat\beta_j - \beta^*_j}{se(\hat\beta_j)} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty$$

- Testing a submodel: Suppose Φ_0 is a column submatrix of Φ (of full column rank of course).

We want to test

$$\mathcal{H}_0 : \mathbb{E}[Y] = \Phi_0 \beta^*_0$$

(as opposed to $\mathbb{E}[Y] = \Phi\beta^*$)

Let $\underset{n \times n}{H_0}$ be the hat matrix for Φ_0 (orthogonal projection matrix onto $C(\Phi_0)$)

To test our null hypothesis we can consider the test statistic:

$$0 \leq \|e_0\|^2 - \|e\|^2 = Y^T(I - H_0)Y - Y^T(I - H)Y$$

$$= Y^T(H - H_0)Y$$

$$e_0 = (I - H_0)Y$$

if $\mathbb{E}[Y] \in C(\Phi_0)$ then

$$Y^T(H - H_0)Y \sim \sigma^2 \chi^2_{\text{rank}(H - H_0)}$$

HW

We still need to estimate σ^2 which we do using the full model. Our final test statistic for testing the null hypothesis is

$$\frac{\dfrac{Y^T(H - H_0)Y}{\text{rank}(H - H_0)}}{\dfrac{Y^T(I - H)Y}{\text{rank}(I - H)}}$$

→ ratio of 2 independent χ^2 divided by their degrees of freedom

$\text{rank}(I - H) = n - d$

$$\sim F_{\text{rank}(H - H_0),\, \text{rank}(I - H)}$$

↓

ANOVA decomposition and F testing

- Make sure you always have an intercept