SDS 387
Linear Models

Fall 2024

Lecture 21 - Tue, Nov 12, 2024

Instructor: Prof. Ale Rinaldo

ASYMPTOTIC NORMALITY OF OLS

Recall we are in the fixed-design, well-specified settings:

$$Y = \Phi \beta^* + \varepsilon$$

Φ: fixed $n \times d$ matrix

ε: n-dimensional vector, mean 0 and covariance $\sigma^2 I_n$

Then:

$$\sqrt{n}\left(\hat{\beta} - \beta^*\right) = \hat{\Sigma}^{-1} \sqrt{n} \frac{\Phi^T \varepsilon}{n}$$

$$\hat{\Sigma} = \frac{\Phi^T \Phi}{n} \rightarrow \underset{d \times d}{\Sigma} \quad \text{by assumption}$$

Last time we wrote that $\frac{\Phi^T \varepsilon}{n}$ can be expressed as

$$\sum_{i=1}^{n} \frac{\Phi_i \varepsilon_i}{n}$$ where Φ_i is the (transpose of) the ith row of Φ

✓ average

Then $\operatorname{Var}\left[\frac{\Phi_i \varepsilon_i}{\sqrt{n}}\right] = \sigma^2 \frac{\Phi_i \Phi_i^T}{n}$ so

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

$$= \sigma^2 \hat{\Sigma} \to \sigma^2 \Sigma$$

- The CLT will follow if we verify the LF condition:

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

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

Notice that, for each $i = 1, \dots, n$,

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

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

Next, because ε_i are iid,

$T \to 0$ as $n \to \infty$ if $\max_i \frac{\|\Phi_i\|}{\sqrt{n}} \to 0$

So, assuming

i) $\hat{\Sigma} \to \Sigma$ as $n \to \infty$

ii) $\max \frac{\|\Phi_i\|}{\sqrt{n}} \to 0$ as $n \to \infty$

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

So

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

by Slutsky's

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

<u>Remark</u>: we could replace condition ii) above with another condition. Notice that

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

$a > 0$, $\frac{\|\Phi_i\|}{\sqrt{n}}$ $\qquad k = 1, 2, \ldots$

So another sufficient condition for LF is that

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

b) $\sum_{i=1}^{n} \left(\frac{\|\Phi_n\|}{\sqrt{n}}\right)^{\kappa} \to 0$

See van der Vaart, chapter 2.

STATISTICAL INFERENCE

- We are interested in formal statistical inference (estimation, hypothesis testing & confidence intervals) for β^* [again, assuming a well-specified linear model and fixed design covariates]

- For now, let's make the simplifying assumption that $\varepsilon \sim N_n(0, \sigma^2 I_n)$. So

$$Y = \Phi\beta^* + \underbrace{\varepsilon}_{\hookrightarrow N_n(0,\sigma^2 I_n)}$$

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

- So the problem reduces to mean estimation of a n-dim Gaussian, under the assumption that the mean belongs to the column space of Φ.

- First notice that

$$\boxed{(\hat{\beta} - \beta^*) \sim N_d\left(0, \sigma^2 (\Phi^T \Phi)^{-1}\right)}$$

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

- In principle we are done: $\hat{\beta} \sim N_d(\beta^*, \sigma^2 (\Phi^T \Phi)^{-1})$

Issue: we do not know $\sigma^2 = \text{Var}[\varepsilon_i]$

- It is natural to use the residuals to estimate σ^2.

Recall that the residuals are

$$\underset{n \times 1}{e} = Y - \hat{Y} = Y - HY = \underbrace{(I_n - H)}_{\text{also a projection matrix}} Y$$

$\hat{Y} \to \Phi\hat{\beta} = HY$ vector of fitted values

where $H = \Phi(\Phi^T\Phi)^{-1}\Phi^T$

$H \to$ hat matrix (a projector onto $C(\Phi)$)

- Thus, since $Y \sim N(\Phi\beta^*, \sigma^2 I_n)$,

$$e = (I_n - H) Y = N(0, \sigma^2 (I_n - H))$$

Exercise

So the residuals are correlated and have different variances

- Nonetheless

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

HW!

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

is an unbiased estimator of σ^2

- Furthermore $\hat{\sigma}^2 \perp\!\!\!\perp \hat{\beta}$ because $\hat{\beta} \in C(\Phi)$ and $\hat{\sigma}^2$ is a function of $(I-H)Y = e$, so

$\perp\!\!\!\perp$ means = they are independent,

$\mathbb{E}[\hat{\beta} e^T] = 0$
$\Rightarrow \hat{\beta} \perp\!\!\!\perp e$ [by Gaussianity]
$\Rightarrow \hat{\beta} \perp\!\!\!\perp \hat{\sigma}^2$

- So

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

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

ratio of $N(0,1)$ and χ^2_{n-d} indep

- <u>Testing a submodel</u> Suppose that Φ_0 is a sub-matrix of Φ obtained by selecting a subset of columns of Φ. We want to test the null hypothesis:

$$H_0: \mathbb{E}[Y] = \Phi_0 \beta_0^*$$

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

Φ_0 identify a submodel.

Let H_0 be the hat matrix for Φ_0 (the projection matrix onto $C(\Phi_0)$)

To test H_0 we can consider:

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

$e_0 = (I-H_0)Y$ residuals for submodel

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

If $\mathbb{E}[Y] \in C(\Phi_0)$ (i.e. if H_0 is true) then

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

[Aside if $\mathbb{E}[Y] = \mu \notin C(\Phi_0)$ then the distribution

$$\sigma^2 \chi^2_{\text{rank}(H-H_0)} \left(\frac{\|(H-H_0)\mu\|^2}{2} \right)$$]

We need to estimate σ^2. We still use

$$\frac{e^T e}{n-d} = \frac{Y^T (I-H) Y}{\text{rank}(I-H)}.$$ We still use the full model to estimate σ^2

So our final test statistic is, under H_0,

$$\frac{\dfrac{Y^T (H-H_0) Y}{\text{rank}(H-H_0)}}{\dfrac{Y^T (I-H) Y}{\text{rank}(I-H)}}$$ ratio of 2 independent χ^2 each divided by their corresponding df! Exercise!

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

- ANOVA (Make sure you always have intercept in your sub-models)

- What if the errors are not Gaussian? If they are independent centered and with constant variance, use the CLT! All the above formulas hold true asymptotically as $n \to \infty$ but make sure to replace $\hat{\sigma}^2$ by $\hat{\sigma}^2/n$ and $\Phi^T\Phi$ by

$\frac{\Phi^T\Phi}{n}$

- Comments. If you use the CLT, you may replace t_{n-d} with $N(0,1)$
  Also the quantity d is # covariate + 1
  $\downarrow$ intercept