SDS 387
Linear Models

Fall 2024

Lecture 26 - Thu, Dec 5, 2024

Instructor: Prof. Ale Rinaldo

- <u>Announcement</u>: final project due by Sunday Dec 15 midnight

- Last time we saw that if we fit the OLS estimator $\hat{\beta}$ when the model is not necessarily linear and the covariates are random then, for fixed d,

$$\sqrt{n}\,(\underbrace{\hat{\beta} - \beta^*}_{\text{projection parameter}}) \xrightarrow{d} N_d\left(0, \Sigma^{-1} V \Sigma^{-1}\right)$$

where $\Sigma = \mathbb{E}\left[\Phi \Phi^\top\right]$ and

$$V = \operatorname{Var}\Big[\underbrace{\Phi_i\,(Y_i - \Phi_i^\top \beta^*)}_{\text{last time we called this } \psi_i}\Big]$$

That is $\sqrt{n}\,(\hat{\beta} - \beta^*)$ has the same asymptotic distribution as $\sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n} \psi_i$

- To carry out formal statistical inference, we need a consistent estimator of the sandwich covariance matrix!

- How do we estimate $\Sigma^{-1} V \Sigma^{-1}$?

- A natural estimator is the plug-in estimator:

$$\hat{\Sigma}^{-1} \hat{V} \hat{\Sigma}^{-1} \quad \text{where} \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \Phi_i \Phi_i^T$$

$$\hat{V} = \frac{1}{n} \sum_{i=1}^{n} \Phi_i \Phi_i^T \left(Y_i - \Phi_i^T \hat{\beta} \right)^2$$

To show that this is a consistent estimator it is sufficient to show that $\hat{\Sigma} \xrightarrow{p} \Sigma$ → trivial: it follows from WLLN

$\hat{V} \to V$

because the result will follow from CMT.

- To show that $\hat{V} \xrightarrow{p} V$ we will let

$$\tilde{V} = \frac{1}{n} \sum_{i=1}^{n} \Phi_i \Phi_i^T \left(Y_i - \Phi_i^T \beta^* \right)$$

↳ non-computable

Then $\tilde{V} \xrightarrow{p} V$ by WLLN. So we need to show that

$$\hat{V} - \tilde{V} \xrightarrow{p} 0$$

- Let's do it:

$$\hat{V} - \tilde{V} = \frac{1}{n} \sum_{i=1}^{n} \Phi_i \Phi_i^T \left[\left(\Phi_i^T \hat{\beta} \right)^2 - \left(\Phi_i^T \beta^* \right)^2 + 2 Y_i \Phi_i^T \left(\beta^* - \hat{\beta} \right) \right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \Phi_i \Phi_i^T \left[\left(\Phi_i^T(\hat\beta - \beta^*)\right)^2 + 2\left(y_i - \Phi_i^T\beta^*\right)\Phi_i^T(\beta^* - \hat\beta)\right]$$

- Next,

$$\|\hat V - \tilde V\|_{op} \le \frac{1}{n}\sum_{i=1}^{n} \|\Phi_i\|^2 \left| \left(\Phi_i^T(\hat\beta - \beta^*)\right)^2 + 2\left(y_i - \Phi_i^T\beta^*\right)\Phi_i^T(\beta^* - \hat\beta)\right|$$

$$= \frac{1}{n}\sum_{i=1}^{n} \|\Phi_i\|^2 \left(\Phi_i^T(\hat\beta - \beta^*)\right)^2 + \frac{2}{n}\sum_{i=1}^{n} \|\Phi_i\| \left|\left(y_i - \Phi_i^T\beta^*\right)\right| \|\Phi_i\| \left|\Phi_i^T(\beta^* - \hat\beta)\right|$$

$$\le \frac{1}{n}\sum_{i=1}^{n} \|\Phi_i\|^2 \left(\Phi_i^T(\hat\beta - \beta^*)\right)^2 + 2\sqrt{\frac{1}{n}\sum_{i=1}^{n} \|\Phi_i\|^2 \left(\Phi_i^T(\beta^* - \hat\beta)\right)^2} \times \sqrt{\frac{1}{n}\sum_{i=1}^{n} \|\Phi_i\|^2 \left(y_i - \Phi_i^T\beta^*\right)^2}$$

by Cauchy Schwartz

$$:= A + 2\sqrt{A}\sqrt{B}$$

Next

by Cauchy Schwartz

$$A \le \underbrace{\left[\frac{1}{n}\sum_{i=1}^{n} \|\Phi_i\|^4\right]}_{\xrightarrow{P} \mathbb{E}\left[\|\Phi_1\|^4\right] \text{ finite}} \|\hat\beta - \beta^*\|^2 \xrightarrow{P} 0$$

$$\xrightarrow{P} 0$$

As for B:

$$B \xrightarrow{P} \mathrm{tr}\left(\mathrm{Var}\left[\Phi_1\left(Y_1 - \Phi_1^T\beta^*\right)\right]\right) = \mathrm{tr}(V)$$

also finite

So

$A + 2\sqrt{A}\sqrt{B} \xrightarrow{P} 0$ by CMT, and $\hat{V} - \tilde{V} \xrightarrow{P} 0$.

Extension to non-iid data.

Lai & Wei (1982) Least squares estimates in stochastic regression models with application to identifications and control of dynamical systems

We observe sequentially observations of the form $(\Phi_i, y_i) \in \mathbb{R}^{d+1}$ $i = 1, 2, 3, \ldots$ where

$$y_i = \Phi_i^T\beta^* + \varepsilon_i$$

where Φ_i may depend on $\{(\Phi_j, \varepsilon_j), j = 1, \ldots, i-1\}$

and $\varepsilon_i \mid (\Phi_j, \varepsilon_j), j = 1, \ldots, i-1 \sim 0, \sigma^2$

with $\mathbb{E}\left[|\varepsilon_i|^\alpha \mid (\Phi_j, \varepsilon_j)\ j = 1, \ldots, i-1\right] < \infty$

some $\alpha > 2$.

Let $\Phi^{(n)}$ be the $n \times d$ matrix whose i^{th} row is Φ_i^T.

Then if

$$\text{wp}1 \quad \begin{cases} \lambda_{min}\left(\Phi^{(n)T}\Phi^{(n)}\right) \to \infty \\ \log \lambda_{max}\left(\Phi^{(n)T}\Phi^{(n)}\right) = o\ \lambda_{min}\left(\Phi^{(n)T}\Phi^{(n)}\right) \end{cases}$$

then
$$\max_{j=1,\dots,d} |\hat{\beta}_j - \beta_j^*| = O\left(\sqrt{\frac{\log \lambda_{max}}{\lambda_{min}}}\right)$$

wp 1 !

Furtermore if $\exists \{B_n\}$, a sequence of p.d. matrices, s.t.

$$B_n^{-1}\left(\Phi^{(n)T}\Phi^{(n)}\right)^{1/2} \xrightarrow[p]{} I_{d\times d} \quad \text{with}$$

$$\max_i \|B_n^{-1}\Phi_i\| \xrightarrow{p} 0$$

then
$$\left(\Phi^{(n)T}\Phi^{(n)}\right)^{1/2}\left(\hat{\beta} - \beta^*\right) \xrightarrow{d} N(0, \sigma^2 I_d)$$