Agenda:

1. linear model with random features

2. Consistency of $\hat{\beta}$ & s^2

3. Asymptotic normality of $\hat{\beta}$

1. Linear model with random features

I. Linearity: $y_i = x_i^T \beta^* + \varepsilon_i$ for some vector of regression coeff's $\beta^* \in \mathbb{R}^p$

II. Predetermined features: $E[\overbrace{x_i \varepsilon_i}^{g_i}] = 0$

III. stationarity & LLN: (x_i, y_i) are identically distributed and

$\frac{1}{n}\sum_{i=1}^n f(x_i, y_i) \xrightarrow{P} E[f(x_1, y_1)]$ for any (reasonable) f

IV: no multicollinearity: $E[x_1 x_1^T] = \Sigma_x$ is non-singular

V: $\underset{p\times 1}{g_i} \doteq \underset{p\times 1}{x_i} \underset{1\times 1}{\varepsilon_i}$ satisfies a CLT: $\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i \xrightarrow{d} N(0, \Sigma_g)$ for some covariance matrix Σ_g

IID (x_i, y_i) pairs: $(x_i, y_i) \overset{IID}{\sim} P$ (P is some dist that satisfies the other 3 conditions)

Recall exogeneity (from classical linear model): $E[\varepsilon_i \mid \underbrace{x_1, \dots, x_n}_{X}] = 0$

$$E[x_i \varepsilon_i] = E[E[x_i \varepsilon_i \mid X]]$$
$$= E[x_i \underbrace{\cancel{E[\varepsilon_i \mid X]}}_{0}]$$

IV: $E[x_1 x_1^T] \doteq \Sigma_x$ is non-singular

$$a^T \Sigma_x a = a^T E[x_1 x_1^T] a = E[a^T x_1 x_1^T a] = E[(a^T x_1)^2] \geq 0$$

III + IV, then $\frac{1}{n}\sum_{i=1}^n x_i x_i^T \xrightarrow{P} \Sigma_x$

$$P(\|\tfrac{1}{n}\sum_{i=1}^n x_i x_i^T - \Sigma_x\|_2^2 > \epsilon) \to 0$$

This implies $\underbrace{\frac{1}{n}\sum_{i=1}^n x_i x_i^T}_{\substack{\shortparallel \\ X^T X}}$ is non-singular

$X^T X$ = $= \sum_{i=1}^n x_i x_i^T$

$$p \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} \begin{bmatrix} \vdots \\ - x_n^T - \end{bmatrix} = \sum_{i=1} \underset{p\times 1}{} \underset{1\times p}{}$$

2. Consistency of $\hat{\beta}$ & s^2

A sequence of estimators $\hat{\theta}_n$ of unknown parameters θ^* is consistent iff $\hat{\theta}_n \xrightarrow{p} \theta^*$

$$\underbrace{(x_1, Y_1), (x_2, Y_2), \ldots \text{ first } n\text{-pairs}}$$
$$\Downarrow$$
$$\hat{\beta}_n$$

Recall if $\hat{\theta}_n \xrightarrow{p} \theta^*$, then $\mathbb{P}(\|\hat{\theta}_n - \theta^*\|_2 \geq \epsilon) \to 0$ for any $\epsilon > 0$

Consistency of OLS:

$$\hat{\beta}_n = (X_n^T X_n)^{-1} X_n^T \vec{Y}_n$$
$$= (X_n^T X_n)^{-1} X_n^T (X_n \beta^* + \vec{\varepsilon}_n)$$
$$= \cancel{(X_n^T X_n)^{-1} X_n^T X_n} \beta^* + (X_n^T X_n)^{-1} X_n^T \vec{\varepsilon}_n$$
$$= \beta^* + \underbrace{(X_n^T X_n)^{-1} X_n^T \vec{\varepsilon}_n}$$

want to show this error term converges in probability to zero

$$X_{\boxed{n}} \doteq \begin{bmatrix} - x_1^T - \\ \vdots \\ - x_{\boxed{n}}^T - \end{bmatrix}, \quad \vec{Y}_{\boxed{n}} \doteq \begin{bmatrix} Y_1 \\ \vdots \\ Y_{\boxed{n}} \end{bmatrix}$$

$$\hat{\beta}_n = (X_n^T X_n)^{-1} X_n^T Y_n$$

$$\vec{\varepsilon}_n = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$X_n^T \vec{\varepsilon}_n \quad \overbrace{\begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix}}^{X_n^T} \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$(X_n^T X_n)^{-1} X_n^T \vec{\varepsilon}_n = \left(\frac{1}{n} X_n^T X_n\right)^{-1} \frac{1}{n} X_n^T \vec{\varepsilon}_n$$
$$= \left(\frac{1}{n} \sum_{i=1}^n x_i x_i^T\right)^{-1} \frac{1}{n} \sum_{i=1}^n x_i \varepsilon_i$$

$\frac{1}{n} \sum_{i=1}^n x_i x_i^T \xrightarrow{p} E[x_1 x_1^T] \doteq \Sigma_X$ (by LLN)

(i) $\left(\frac{1}{n} \sum_{i=1}^n x_i x_i^T\right)^{-1} \xrightarrow{p} \Sigma_X^{-1}$ (by CMT)

(ii) $\frac{1}{n} \sum_{i=1}^n x_i \varepsilon_i \xrightarrow{p} E[x_1 \varepsilon_1] = 0$ (by LLN + predetermined features)

By properties of convergence in probability,
(i) & (ii) imply $\left(\frac{1}{n} \sum_{i=1}^n x_i x_i^T\right)^{-1} \frac{1}{n} \sum_{i=1}^n x_i \varepsilon_i \xrightarrow{p} \Sigma_X^{-1} \cdot 0 = 0$

Claim: consistency of $s^2 \doteq \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i) = \frac{1}{n} (\vec{Y}_n - \hat{\vec{Y}}_n)^T (\vec{Y}_n - \hat{\vec{Y}}_n)$

$$= \frac{1}{n}(\vec{y}_n - X_n\hat{\beta}_n)^T(\vec{y}_n - X_n\hat{\beta}_n) \quad (\hat{\vec{y}}_n = X_n\hat{\beta}_n)$$

$$(\vec{y}_n = X_n\beta^* + \vec{\varepsilon}_n) \quad = \frac{1}{n}(X_n\beta^* + \vec{\varepsilon}_n - X_n\hat{\beta}_n)^T(X_n\beta^* + \vec{\varepsilon}_n - X_n\hat{\beta}_n)$$

$$= \frac{1}{n}(\beta^* - \hat{\beta}_n)^T X_n^T X_n(\beta^* - \hat{\beta}_n) + \frac{2}{n}\vec{\varepsilon}_n^T X_n(\beta^* - \hat{\beta}_n)$$

$$+ \frac{1}{n}\vec{\varepsilon}_n^T\vec{\varepsilon}_n$$

$$= (\beta^* - \hat{\beta}_n)^T\left(\frac{1}{n}X_n^T X_n\right)(\beta^* - \hat{\beta}_n) + \frac{2}{n}\boxed{\vec{\varepsilon}_n^T X_n}(\beta^* - \hat{\beta}_n) + \frac{1}{n}\vec{\varepsilon}_n^T\vec{\varepsilon}_n$$

$$= \underbrace{(\beta^* - \hat{\beta}_n)^T}_{\xrightarrow{p} 0}\underbrace{\left(\frac{1}{n}\sum_{i=1}^n x_i x_i^T\right)}_{\xrightarrow{p} E[x_i x_i^T] = \Sigma_x}\underbrace{(\beta^* - \hat{\beta}_n)}_{\xrightarrow{p} 0} + 2\underbrace{\left(\frac{1}{n}\sum_{i=1}^n x_i\varepsilon_i\right)^T}_{\xrightarrow{p} E[x_i\varepsilon_i] = 0}\underbrace{(\beta^* - \hat{\beta}_n)}_{\xrightarrow{p} 0} + \underbrace{\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2}_{\xrightarrow{p} \sigma^2}$$

$$\xrightarrow{p} \sigma^2$$

Asymptotic normality of OLS:

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

$$= \left(\frac{1}{n}X_n^T X_n\right)^{-1}\frac{1}{\sqrt{n}}X_n^T\vec{\varepsilon}_n$$

$$= \underbrace{\left(\frac{1}{n}\sum_{i=1}^n x_i x_i^T\right)^{-1}}_{\xrightarrow{p} \Sigma_x^{-1} \text{ (by CLN)}}\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i\varepsilon_i}_{\xrightarrow{d} N(0, \Sigma_g) \text{ (by CLT)}}$$

$$\xrightarrow{d} \Sigma_x^{-1} Z, \quad Z \sim N(0, \Sigma_g)$$

$$= N(0, \Sigma_x^{-1}\Sigma_g\Sigma_x^{-1})$$