DISTRIBUTION OF THE ESTIMATORS

Given $\underline{Y} = (Y_1, \dots, Y_n)$ and X $(n \times p)$

we have seen that the estimators are

$$\hat{\underline{\beta}} = (X^T X)^{-1} X^T \underline{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n} \underline{\varepsilon}^T \underline{\varepsilon} \qquad \text{with } \underline{\varepsilon} = \underline{Y} - \hat{\underline{Y}} = \underline{Y} - X\hat{\underline{\beta}}$$

$$S^2 = \frac{1}{n-p} \underline{\varepsilon}^T \underline{\varepsilon}$$

we want to derive their exact distribution to perform inference.

• DISTRIBUTION OF $\hat{\underline{\beta}}$

estimator $\hat{\underline{\beta}} = (X^T X)^{-1} X^T \underline{Y}$ linear in $\underline{Y}$ (i.e. $\hat{\underline{\beta}} = A \cdot \underline{Y}$) with $\underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I_n)$

• LINEAR TRANSFORMATIONS of GAUSSIAN RANDOM VECTORS

$\underline{z} \sim N_d(\underline{\mu}, \Sigma)$, A $(k \times d)$ matrix, $\underline{b} \in \mathbb{R}^k$

$$\Rightarrow \quad \underline{T} = \underbrace{A}_{k \times d} \cdot \underbrace{\underline{z}}_{d \times 1} + \underbrace{\underline{b}}_{k \times 1} \sim N_k(A\underline{\mu} + \underline{b}, \; A \Sigma A^T)$$

$$\mathbb{E}[\underline{T}] = \mathbb{E}[A\underline{z} + \underline{b}] = A\,\mathbb{E}[\underline{z}] + \underline{b} = A\underline{\mu} + \underline{b}$$

$$\text{var}(\underline{T}) = \text{var}(A\underline{z} + \underline{b}) = \text{var}(A\underline{z}) = \mathbb{E}[(A\underline{z} - \mathbb{E}[A\underline{z}])(A\underline{z} - \mathbb{E}[A\underline{z}])^T]$$
$$= \mathbb{E}[(A\underline{z} - A\underline{\mu})(A\underline{z} - A\underline{\mu})^T] = \mathbb{E}[A(\underline{z} - \underline{\mu})(\underline{z} - \underline{\mu})^T A^T] = A \Sigma A^T$$

$$\Rightarrow \quad \hat{\underline{\beta}} = A\underline{Y} \quad \text{with } A = (X^T X)^{-1} X^T \qquad p \times n \quad \text{rank} = p$$

$$\mathbb{E}[\hat{\underline{\beta}}] = A\,\mathbb{E}[\underline{Y}] = (X^T X)^{-1} X^T \mathbb{E}[\underline{Y}] = (X^T X)^{-1} X^T X \underline{\beta} = \underline{\beta}$$

$$\text{var}(\hat{\underline{\beta}}) = \text{var}(A\underline{Y}) = A\,\text{var}(\underline{Y})\,A^T = A(\sigma^2 I_n)A^T = \sigma^2 AA^T = \sigma^2 (X^T X)^{-1} X^T X ((X^T X)^{-1})^T = \sigma^2 (X^T X)^{-1}$$

$$\Rightarrow \quad \hat{\underline{\beta}} \sim N_p(\underline{\beta}, \; \sigma^2 (X^T X)^{-1})$$

the marginal distribution of the j-th component is $\hat{\beta}_j \sim N_1(\beta_j, \; \sigma^2 [(X^T X)^{-1}]_{(j,j)})$ $\qquad j = 1, \dots, p$

the covariance between two components j and s $\text{cov}(\hat{\beta}_j, \hat{\beta}_s) = \sigma^2 [(X^T X)^{-1}]_{(j,s)}$ $\qquad j, s = 1, \dots, p$

Notice that the variance is $\sigma^2 (X^T X)^{-1}$: once again, we need $(\underline{x}_1, \dots, \underline{x}_p)$ to be linearly independent.

indeed $(X^T X)^{-1} = \frac{1}{\det(X^T X)} \cdot [\dots]$

If they are linearly dependent, $\det(X^T X) = 0$ and $(X^T X)$ is not invertible

Remark: if they are "almost collinear" ($\det(X^T X) \approx 0$) the variance of the estimator explodes

$$\text{var}(\hat{\underline{\beta}}) = \sigma^2 \cdot \underbrace{\frac{1}{\det(X^T X)}}_{\to +\infty} [\dots]$$

• DISTRIBUTION OF THE RESIDUALS

$$\underline{e} = \underline{y} - \hat{\underline{y}} = \underline{y} - X\hat{\underline{\beta}} = \underline{y} - X \cdot (X^T X)^{-1} X^T \underline{y} = (I_n - \underbrace{X(X^T X)^{-1} X^T}) \underline{y} = (I_n - P)\underline{y}$$

we called it P:
$P = X(X^T X)^{-1} X^T$ $(n \times n)$ matrix

Let us study the corresponding random quantity $\underline{E}$

$$\underline{E} = (I_n - P)\underline{Y}$$
$$= (I_n - P)(X\underline{\beta} + \underline{\varepsilon})$$
$$= \underbrace{(I_n - P)X\underline{\beta}}_{=0} + (I_n - P)\underline{\varepsilon} = (I_n - P)\underline{\varepsilon} \qquad \underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I_n)$$

$$\downarrow$$

$$(I_n - P)X\underline{\beta} = X\underline{\beta} - PX\underline{\beta} = X\underline{\beta} - \underbrace{X(X^T X)^{-1} X^T}_{P} X\underline{\beta} = X\underline{\beta} - X\underline{\beta} = 0$$

Hence, $\underline{E} = (I_n - P)\underline{\varepsilon}$ $\qquad \underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I_n)$

linear combination of a Gaussian is Gaussian $\rightarrow$ $\underline{E} \sim N_n(\dots, \dots)$

$$\mathbb{E}[\underline{E}] = (I_n - P)\underline{0} = \underline{0}$$

$$\text{var}(\underline{E}) = (I_n - P)\,\text{var}(\underline{\varepsilon})\,(I_n - P)^T = \sigma^2 (I_n - P)(I_n - P)^T = \sigma^2 (I_n - P)$$

$$\Rightarrow \quad \underline{E} \sim N_n(\underline{0}, (I_n - P)\sigma^2)$$

(i.e. $\text{var}(E_i) = \sigma^2 (I_n - P)_{(ii)}$ $\rightarrow$ not homoscedastic)

• DISTRIBUTION OF $\hat{\sigma}^2$

Preliminary result

If $\underline{z} \sim N_d(\underline{0}, \sigma^2 I_n)$; B a $(d \times d)$ matrix, symmetric, idempotent, with rank k $(1 \le k \le d)$

$$\Rightarrow \quad Q = \frac{1}{\sigma^2} \underline{z}^T B \underline{z} \qquad Q \sim \chi^2_k$$

$$\hat{\sigma}^2 = \frac{1}{n} \underline{\varepsilon}^T \underline{\varepsilon}$$

$$\Rightarrow \quad n\hat{\sigma}^2 = \underline{\varepsilon}^T \underline{\varepsilon} = ((I_n - P)\underline{\varepsilon})^T ((I_n - P)\underline{\varepsilon})$$
$$= \underline{\varepsilon}^T (I_n - P)\underline{\varepsilon}$$

We have seen that $(I_n - P)$ is symmetric and idempotent.

Moreover, the rank of $(I_n - P)$ is equal to $(n - p)$.

$\Rightarrow$ applying the result we get

$$\frac{1}{\sigma^2} \cdot \underline{\varepsilon}^T (I_n - P)\underline{\varepsilon} \sim \chi^2_{n-p} \qquad \Rightarrow \qquad \frac{\underline{\varepsilon}^T \underline{\varepsilon}}{\sigma^2} = \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$$

$n - p = \text{rank}(I_n - P)$
$= \#\text{units} - \#\text{covariates}$

Hence, we also find that $\mathbb{E}\left[\frac{n\hat{\sigma}^2}{\sigma^2}\right] = n - p \quad \Rightarrow \quad \mathbb{E}[\hat{\sigma}^2] = \frac{\sigma^2}{n}(n - p)$ biased

As usual, we obtain an unbiased estimator as

$$S^2 = \frac{\underline{\varepsilon}^T \underline{\varepsilon}}{n - p} = \frac{n\hat{\sigma}^2}{n - p} \quad \text{with} \quad \frac{(n - p)S^2}{\sigma^2} \sim \chi^2_{n-p} \quad \Rightarrow \quad \mathbb{E}[S^2] = \sigma^2$$

Moreover, $\hat{\underline{\beta}} \perp\!\!\!\perp \hat{\sigma}^2$ and $\hat{\underline{\beta}} \perp\!\!\!\perp S^2$