Summary

linear models $\quad Y = X\theta^* + \varepsilon \;;\; Y, \varepsilon \in \mathbb{R}^n,\ X \in \mathbb{R}^{n\times p+1},\ \theta^* \in \mathbb{R}^{p+1}$

- fixed design $\quad X$ is deterministic
  - $\mathbb{E}[\varepsilon] = 0 \qquad Var(\varepsilon) = \sigma^2$
  - $\varepsilon \sim \mathcal{N}(0, \Sigma) \leftarrow$
- Random design $\quad X$: random

- Goal : get a <u>point</u> estimator of θ^*, $\hat{\theta}$
  - MLE
  - OLS

$(X^T X)\hat{\theta} = XY \leftarrow$ a solution

$Ker(X) = Ker(X^T X) \neq \{0\}$, i.e $(X^T X)$ is invertible

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

Today : Confidence interval estimation

<u>Rem</u> $\quad$ Cochran's Lemma (under the Gaussian model)

1) $\hat{\theta} \sim \mathcal{N}(\theta^*, \sigma^2 (X^T X)^{-1})$

2) $n-p-1 \dfrac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p-1}$

3) $\hat{\theta}$ and $\hat{\sigma}$ are independent

4) $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$

5) relation to T-student distribution

Notation

$\hat{\varepsilon}_i = y_i - \hat{y}_i$

$\hat{\sigma}^2 = \dfrac{1}{n-p-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2$

Slide 5 - 'Method 1 for CI

Random coin : $Be(p) = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{else (w.p. } 1-p) \end{cases}$

$X_1, \dots, X_n$, $n = 1000$

goals: 1) $\hat{p}$, give CI $[\hat{p} - \delta, p + \delta]$

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

▸ We built C.I using a concentration inequality Tchebyschev

r.v. Z, $\forall \delta > 0$ $\quad P(|Z - E[Z]| > \delta) \le \frac{Var(Z)}{\delta^2}$

$\hat{p}$ is a r.v $\quad E[\hat{p}] = p$

$E[X_1] = P(X=1)\,1 + P(X=0)\,0 = p$

$E[\hat{p}] = E[\frac{1}{n}\sum X_i] = p$ → unbiased

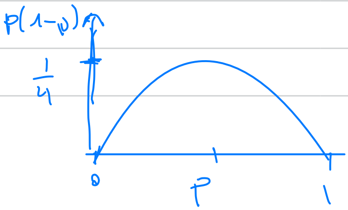
$Var(\hat{p}) = Var(\frac{1}{n}\sum X_i) = \frac{1}{n^2}\sum Var(X_i) = \frac{1}{n^2} n (p - p^2) = \frac{1}{n}(p - p^2)$

↳ using $Var(X_1) = E[X_1^2] - E[X_1]^2$

$E[X_1^2] = P(X=1)\,1^2 + P(X=0)\,0^2 = p$

$$P(|\hat{p} - p| > \delta) \le \frac{p - p^2}{n\delta^2} = \frac{1}{4n\delta^2}$$

$p - p^2 = p(1-p) \qquad 0 \le p \le 1$



We want a C.I. $P(p \in [\hat{p}-\delta, \hat{p}+\delta]) \geq 0.95$

$\Rightarrow P(|\hat{p}-p| > \delta) \leq 0.05$

$P(|\hat{p}-p| > \delta) \leq \frac{1}{4n\delta^2} = 0.05$, for $n = 1000$

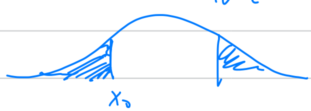
$$\delta = \frac{1}{\sqrt{200}} \simeq 0.07$$

→ Different inequalities. Markov, Bernstein, Hoeffding, ... --

Slide 7

cumulative distributions $P(X \leq x_0)$

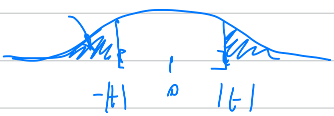
$N(0,1)$



$P(X \geq x_1) = 1 - P(X \leq x_1)$

$cdf(x_1)$ ← notation   norm.cdf(__)

for any $t \in \mathbb{R}$



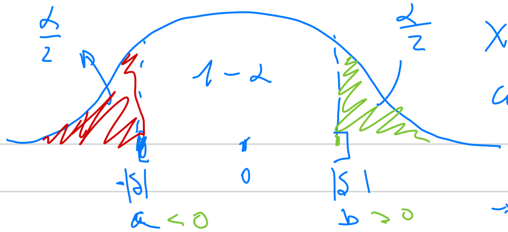
$2 \cdot cdf(-|t|)$

→ Quantiles: inverse of the cumulative distribution function
↳ t_α α-th quantile

$P(X \leq x_0) = \alpha \iff t_\alpha = X$,

The α-th quantile of P is x_0

We do not give analytical expressions for cdf., t_α. But we can compute them



$X \sim N(0,1)$

compute a CI around 0 of level $1-\alpha$

→ Since its symmetric, $a = -b$

δ ?

~~a~~ : $P(X \le a) = \frac{\alpha}{2} \Rightarrow a = t_{\frac{\alpha}{2}} \le 0$

$P(X \ge b) = \frac{\alpha}{2}$

$P(X \le b) = 1 - \frac{\alpha}{2} \Leftrightarrow b = t_{1-\frac{\alpha}{2}} > 0$

→ To conclude: A confidence interval (CI) for a normally distributed r.v. : $P(-\delta \le X \le \delta) \ge 1-\alpha$

$$[t_{\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}}] = [-t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}}]$$

$x_0 \in \mathbb{R}^{p+1}$ $X \in \mathbb{R}^{n\times(p+1)}$

Slide 22. CI for the mean response

Goal : give a CI for the prediction at x_0 ⇒ $\hat{y}_0 = x_0^T \hat{\theta}$

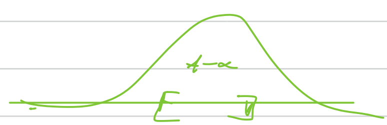
$$\mathbb{E}[\hat{y}_0] = \mathbb{E}[x_0^T \hat{\theta}] = x_0^T \theta^*$$
$$\mathrm{Var}(\hat{y}_0) = \mathrm{Var}(x_0^T \hat{\theta}) = x_0^T \mathrm{Cov}(\hat{\theta})\, x_0 = \sigma^2 x_0^T (X^T X)^{-1} x_0$$
let $y_0^* = x_0^T \theta^*$

$$\frac{\dfrac{y_0^* - \hat{y}_0}{\sqrt{\cancel{\sigma^2}\, x_0^T (X^T X)^{-1} x_0}}}{\sqrt{\dfrac{\hat{\sigma}^2}{\cancel{\sigma^2}} \dfrac{(n-\cancel{p}-1)}{(n-p-1)}}} = T$$

$\sim N(0,1)$

$\sim \chi^2_{n-p-1}$

$$\frac{U}{\sqrt{Z_\nu / \nu}} \sim T_{n-p-1}$$



(i) $T \sim T_{n-p-1}$

(ii) $$T = \frac{y_0^* - \hat{y}_0}{\sqrt{\hat{\sigma}^2 x_0^T (X^T X)^{-1} x_0}}$$

(iii) . We can evaluate every term in T but y_0^*

Slide 23 Prediction interval

$$y_0 = x_0^T \theta^* + \varepsilon \,, \qquad \overbrace{y_0 - \hat{y}_0}$$

I will predict y_0 using $\hat{y}_0 = x_0^T \hat{\theta}$

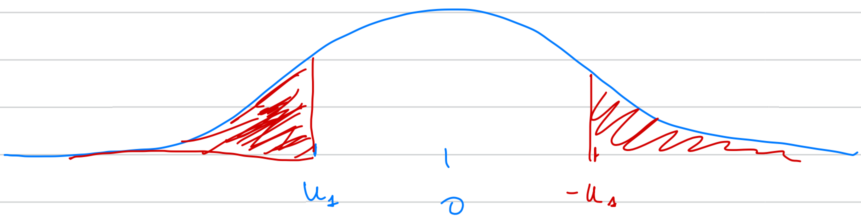
$$\mathbb{E}[y_0] = \mathbb{E}[x_0^T \theta^* + \varepsilon] = x_0^T \theta^*$$

$$\mathrm{Var}(y_0) = \mathrm{Var}(x_0^T \theta^* + \varepsilon) = \sigma^2$$

$$\mathrm{Var}(y_0 - \hat{y}_0) = \mathrm{Var}(x_0^T \theta^* + \varepsilon - x_0^T \hat{\theta}) = \\ = \sigma^2 (1 + x_0^T (X^T X)^{-1} x_0)$$

$$\frac{\dfrac{y_0 - \hat{y}_0}{\sqrt{\sigma^2 (1 + x_0^T (X^T X)^{-1} x_0)}}}{\sqrt{\dfrac{\hat{\sigma}^2}{\sigma^2} \dfrac{(n-p-1)}{n-p-1}}} = \frac{y_0 - \hat{y}_0}{\sqrt{\sigma^2 (1 + x_0^T (X^T X)^{-1} x_0)}} \sim T_{n-p-1}$$

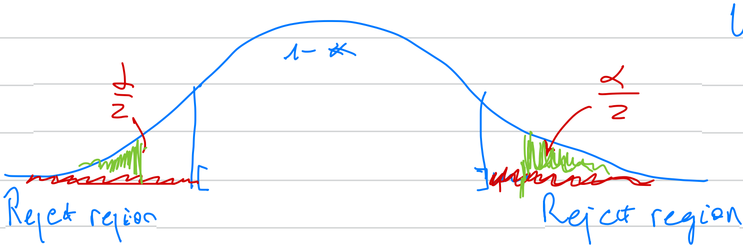
p-value : probability of observing an event more extreme than the one observed (u)



pvalue $2 \cdot \mathrm{cdf}(-|u_1|)$

Type I error : probability of a wrong reject : α

$u \sim N(0,1)$



- For a test of level $1-\alpha$, p-value

if rejecting H_0 | p-value $< \alpha \rightarrow$ reject H_0