

1 Warm-up

Let X be a random vector in \mathbb{R}^d with mean $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$ and A a symmetric matrix in $\mathbb{R}^{d \times d}$. Then,

$$\mathbb{E}[X^\top AX] = \mu^\top A\mu + \text{Trace}(A\Sigma).$$

As $X^\top AX$ is a real number, $\mathbb{E}[X^\top AX] = \mathbb{E}[\text{Trace}(X^\top AX)] = \mathbb{E}[\text{Trace}(AXX^\top)]$. By linearity, $\mathbb{E}[X^\top AX] = \text{Trace}(A\mathbb{E}[XX^\top])$ which yields,

$$\mathbb{E}[X^\top AX] = \text{Trace}(A\{\mathbb{V}[X] + \mathbb{E}[X]\mathbb{E}[X]^\top\}) = \mu^\top A\mu + \text{Trace}(A\Sigma).$$

2 Student's t-statistics

We assume that for all $1 \leq i \leq n$, $Y_i = X_i^\top \theta_\star + \varepsilon_i$ for some unknown $\theta_\star \in \mathbb{R}^d$ where the $(\varepsilon_i)_{1 \leq i \leq n}$ are i.i.d. random variables with distribution $\mathcal{N}(0, \sigma_\star^2)$. Let $\varepsilon \in \mathbb{R}^n$ be the random vector such that for all $1 \leq i \leq n$, the i -th component of ε is ε_i . The model is then written $Y = X\theta_\star + \varepsilon$. Assume that X has full rank and that $\hat{\theta}_n = (X^\top X)^{-1}X^\top Y$ and $\hat{\sigma}_n^2 = \|Y - X\hat{\theta}_n\|^2/(n-d)$.

1. For all $1 \leq j \leq d$, show that

$$\frac{\hat{\theta}_{n,j} - \theta_{\star,j}}{\hat{\sigma}_n \sqrt{(X^\top X)^{-1}_{j,j}}} \sim \mathcal{S}(n-d),$$

where $\mathcal{S}(n-d)$ is the Student's t-distribution with $n-d$ degrees of freedom, i.e. the law of $X/\sqrt{Y/(n-d)}$ where $X \sim \mathcal{N}(0, 1)$ is independent of $Y \sim \chi^2(n-d)$.

By definition, for all $1 \leq j \leq d$,

$$\frac{\hat{\theta}_{n,j} - \theta_{\star,j}}{\hat{\sigma}_n \sqrt{(X^\top X)^{-1}_{j,j}}} = \frac{\sigma_\star^{-1}(\hat{\theta}_{n,j} - \theta_{\star,j})}{\sigma_\star^{-1}\hat{\sigma}_n \sqrt{(X^\top X)^{-1}_{j,j}}} = \frac{e_j^\top (\sigma_\star^{-1}(\hat{\theta}_n - \theta_\star))}{\sigma_\star^{-1}\hat{\sigma}_n \sqrt{(X^\top X)^{-1}_{j,j}}}.$$

Note that $\sigma_\star^{-1}(\hat{\theta}_n - \theta_\star) \sim \mathcal{N}(0, (X^\top X)^{-1})$ so that $e_j^\top (\sigma_\star^{-1}(\hat{\theta}_n - \theta_\star)) \sim \mathcal{N}(0, e_j^\top (X^\top X)^{-1} e_j)$ and

$$\frac{e_j^\top (\sigma_\star^{-1}(\hat{\theta}_n - \theta_\star))}{\sqrt{(X^\top X)^{-1}_{j,j}}} \sim \mathcal{N}(0, 1).$$

In addition,

$$\sigma_\star^{-1}\hat{\sigma}_n = \sqrt{\sigma_\star^{-2}\hat{\sigma}_n^2} = \sqrt{\|\sigma_\star^{-1}(I_n - X(X^\top X)^{-1}X^\top)\varepsilon\|_2^2/(n-d)},$$

where $(n-d)\sigma_\star^{-2}\hat{\sigma}_n^2 = \|\sigma_\star^{-1}(I_n - X(X^\top X)^{-1}X^\top)\varepsilon\|_2^2 \sim \chi^2(n-d)$. The proof is concluded by noting that θ_n and $\hat{\sigma}_n^2$ are independent.

2. Provide a confidence interval with confidence level $1 - \alpha$ for $\theta_{\star,j}$.

For $\alpha \in (0, 1)$, if $s_{1-\alpha/2}^{n-d}$ denotes the quantile of order $1 - \alpha/2$ of the law $\mathcal{S}(n-d)$, then

$$\mathbb{P} \left(\left| \frac{\hat{\theta}_{n,j} - \theta_{\star,j}}{\hat{\sigma}_n \sqrt{(X^T X)_{j,j}^{-1}}} \right| \leq s_{1-\alpha/2}^{n-d} \right) = 1 - \alpha.$$

Therefore,

$$I_{n,j}^{n-d}(\theta_{\star}) = \left[\hat{\theta}_{n,j} - \hat{\sigma}_n s_{1-\alpha/2}^{n-d} \sqrt{(X^T X)_{j,j}^{-1}}; \hat{\theta}_{n,j} + \hat{\sigma}_n s_{1-\alpha/2}^{n-d} \sqrt{(X^T X)_{j,j}^{-1}} \right]$$

is a confidence interval for $\theta_{\star,j}$ with confidence level $1 - \alpha$.

3 Random design

Consider the regression model given by

$$Y = X\theta_{\star} + \varepsilon,$$

where $X \in \mathbb{R}^{n \times d}$, the $(\varepsilon_i)_{1 \leq i \leq n}$ are i.i.d. centered Gaussian random variables with variance σ_{\star}^2 and independent of $(X_i)_{1 \leq i \leq n}$ which are assumed to be random. Assume that $X^T X$ has full rank and that θ_{\star} is estimated by

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

1. Compute the excess risk $R(\theta) - R(\theta_{\star})$, where $R(\theta) = n^{-1} \mathbb{E}[\|Y - X\theta\|_2^2]$.

By definition, using that $\mathbb{E}[\varepsilon] = 0$,

$$\begin{aligned} R(\theta) &= n^{-1} \mathbb{E}[\|Y - X\theta\|_2^2] = n^{-1} \mathbb{E}[\|X\theta_{\star} + \varepsilon - X\theta\|_2^2], \\ &= n^{-1} \mathbb{E}[\|X\theta_{\star} - X\theta\|_2^2] + n^{-1} \mathbb{E}[\|\varepsilon\|_2^2], \\ &= (\theta_{\star} - \theta)^T n^{-1} \mathbb{E}[X^T X] (\theta_{\star} - \theta) + \sigma_{\star}^2. \end{aligned}$$

Therefore, $R(\theta) - R(\theta_{\star}) = (\theta_{\star} - \theta)^T n^{-1} \mathbb{E}[X^T X] (\theta_{\star} - \theta)$.

2. Compute then the excess risk $\mathbb{E}[R(\hat{\theta}_n) - R(\theta_{\star})]$.

By the previous question,

$$\mathbb{E}[R(\hat{\theta}_n) - R(\theta_{\star})] = n^{-1} \mathbb{E}[(\theta_{\star} - \hat{\theta}_n)^T \mathbb{E}[X^T X] (\theta_{\star} - \hat{\theta}_n)].$$

Since $\hat{\theta}_n$ is an unbiased estimate of θ_{\star} ,

$$\begin{aligned} \mathbb{E}[R(\hat{\theta}_n) - R(\theta_{\star})] &= n^{-1} \mathbb{E}[(\theta_{\star} - \mathbb{E}[\hat{\theta}_n])^T \mathbb{E}[X^T X] (\theta_{\star} - \mathbb{E}[\hat{\theta}_n])], \\ &= n^{-1} \text{Trace} \left(\mathbb{E}[X^T X] \mathbb{V}[\hat{\theta}_n] \right), \\ &= \frac{\sigma_{\star}^2}{n} \text{Trace} \left(\mathbb{E}[X^T X] \mathbb{E}[(X^T X)^{-1}] \right). \end{aligned}$$

4 Fisher statistics (bonus)

Consider the regression model given by

$$Y = X\theta_{\star} + \varepsilon,$$

where $X \in \mathbb{R}^{n \times d}$ and the $(\varepsilon_i)_{1 \leq i \leq n}$ are i.i.d. centered Gaussian random variables with variance σ_{\star}^2 . Assume that $X^T X$ has full rank and that θ_{\star} and σ_{\star}^2 are estimated by

$$\hat{\theta}_n = (X^T X)^{-1} X^T Y \quad \text{and} \quad \hat{\sigma}_n^2 = \frac{\|Y - X\hat{\theta}_n\|^2}{n-d}.$$

1. Let L be a $\mathbb{R}^{q \times d}$ matrix with rank $q \leq d$. Show that

$$\frac{(\hat{\theta}_n - \theta_\star)^\top L^\top (L(X^\top X)^{-1} L^\top)^{-1} L(\hat{\theta}_n - \theta_\star)}{q \hat{\sigma}_n^2} \sim \mathcal{F}(q, n-d),$$

where $\mathcal{F}(q, n-d)$ is the Fisher distribution with q and $n-d$ degrees of freedom, i.e. the law of $(X/q)/(Y/(n-d))$ where $X \sim \chi^2(q)$ is independent of $Y \sim \chi^2(n-d)$.

Note that $\text{rank}(L(X^\top X)^{-1} L^\top) = \text{rank}(LL^\top) = q$. The matrix $L(X^\top X)^{-1} L^\top$ is therefore positive definite. There exists a diagonal matrix $D \in \mathbb{R}^{q \times q}$ with positive diagonal terms and an orthogonal matrix $Q \in \mathbb{R}^{q \times q}$ such that $L(X^\top X)^{-1} L^\top = QDQ^{-1}$. The matrix $(L(X^\top X)^{-1} L^\top)^{-1/2}$ may be defined as $(L(X^\top X)^{-1} L^\top)^{-1/2} = QD^{-1/2}Q^{-1}$.

It is then enough to note that $(L(X^\top X)^{-1} L^\top)^{-1/2} L(\hat{\theta}_n - \theta_\star)/\sigma_\star \sim \mathcal{N}(0, I_q)$. Therefore,

$$\begin{aligned} \sigma_\star^{-2} \|(L(X^\top X)^{-1} L^\top)^{-1/2} L(\hat{\theta}_n - \theta_\star)\|^2 \\ = (\hat{\theta}_n - \theta_\star)^\top L^\top (L(X^\top X)^{-1} L^\top)^{-1} L(\hat{\theta}_n - \theta_\star) / \sigma_\star^2 \sim \chi^2(q). \end{aligned}$$

On the other hand, we know that

$$(n-d)\sigma_\star^{-2} \hat{\sigma}_n^2 \sim \chi^2(n-d).$$

The proof is concluded by noting that $\hat{\theta}_n$ and $\hat{\sigma}_n^2$ are independent.

2. Using the previous question, build a confidence region with confidence level $1 - \alpha \in (0, 1)$ for θ_\star .

By the previous question, for $\alpha \in (0, 1)$, if $f_{1-\alpha}^{q, n-d}$ denotes the quantile of order $1 - \alpha$ of the law $\mathcal{F}(q, n-d)$, then

$$\mathbb{P} \left(\theta_\star \in \left\{ \theta \in \mathbb{R}^d; (\hat{\theta}_n - \theta)^\top L^\top (L(X^\top X)^{-1} L^\top)^{-1} L(\hat{\theta}_n - \theta) \leq q \hat{\sigma}_n^2 f_{1-\alpha}^{q, n-d} \right\} \right) = 1 - \alpha.$$

Therefore,

$$I_n^{q, n-d}(\theta_\star) = \left\{ \theta \in \mathbb{R}^d; (\hat{\theta}_n - \theta)^\top L^\top (L(X^\top X)^{-1} L^\top)^{-1} L(\hat{\theta}_n - \theta) \leq q \hat{\sigma}_n^2 f_{1-\alpha}^{q, n-d} \right\}$$

is a confidence region for θ_\star with confidence level $1 - \alpha$.