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,

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

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

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

2 Student's t-statistics

We assume that for all $1 \leqslant i \leqslant n$, $Y_i = X_i^\top\theta_\star + \varepsilon_i$ for some unknown $\theta_\star \in \mathbb{R}^d$ where the $(\varepsilon_i)_{1\leqslant i\leqslant n}$ are i.i.d. random variables with distribution $\mathcal{N}(0,\sigma_*^2)$. Let $\varepsilon \in \mathbb{R}^n$ be the random vector such that for all $1 \leqslant i \leqslant n$, 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 $\widehat{\theta}_n = (X^\top X)^{-1}X^\top Y$ and $\widehat{\sigma}_n^2 = \|Y - X\widehat{\theta}_n\|^2/(n-d)$.

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

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

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 $1 \leqslant j \leqslant d$,

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

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

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

In addition,

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

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

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

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

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

Therefore,

$$I^{n-p}_{n,j}(\theta_\star) = \left[\widehat{\theta}_{n,j} - \widehat{\sigma}_n s^{n-d}_{1-\alpha/2}\sqrt{(X^\top X)^{-1}_{j,j}}\,;\,\widehat{\theta}_{n,j} + \widehat{\sigma}_n s^{n-d}_{1-\alpha/2}\sqrt{(X^\top X)^{-1}_{j,j}}\right]$$

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 $(\varepsilon_i)_{1\leqslant i\leqslant n}$ are i.i.d. centered Gaussian random variables with variance $\sigma^2_\star$ and independent of $(X_i)_{1\leqslant i\leqslant n}$ which are assumed to be random. Assume that $X^\top X$ has full rank and that $\theta_\star$ is estimated by

$$\widehat{\theta}_n = (X^\top X)^{-1}X^\top Y\,.$$

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

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

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

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

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

By the previous question,

$$\mathbb{E}[\mathsf{R}(\widehat{\theta}_n) - \mathsf{R}(\theta_\star)] = n^{-1}\mathbb{E}[(\theta_\star - \widehat{\theta}_n)^\top\mathbb{E}[X^\top X](\theta_\star - \widehat{\theta}_n)]\,.$$

Since $\widehat{\theta}_n$ is an unbiased estimate of $\theta_\star$,

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

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 $(\varepsilon_i)_{1\leqslant i\leqslant n}$ are i.i.d. centered Gaussian random variables with variance $\sigma^2_\star$. Assume that $X^\top X$ has full rank and that $\theta_\star$ and $\sigma_\star^2$ are estimated by

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

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

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

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 $\mathrm{rank}(L(X^\top X)^{-1}L^\top) = \mathrm{rank}(LL^\top) = q$. *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 $(L(X^\top X)^{-1}L^\top)^{-1/2}L(\widehat{\theta}_n - \theta_\star)/\sigma_\star \sim \mathcal{N}(0, I_q)$. *Therefore,*

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

On the other hand,we know that

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

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

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

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* $\mathcal{F}(q, n-p)$, *then*

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

Therefore,

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

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