Correction du TD 6

Exercice 1

1. On suppose que σ est connu.
   (a) Comme $\overline{Y}_n \sim \mathcal{N}(m, \sigma^2/n)$, on a $\sqrt{n}(\overline{Y}_n - m)/\sigma \sim \mathcal{N}(0,1)$. On a donc pour tout $u \geq 0$,
   $$\mathbb{P}\left(\sqrt{n}\left|\frac{\overline{Y}_n - m}{\sigma}\right| \leq u\right) = 2\Phi(u) - 1.$$
   Ainsi, on obtient l'intervalle de confiance de niveau $1-\alpha$:
   $$I(m) = \left[\overline{Y}_n - \sigma\frac{\Phi^{-1}(1-\alpha/2)}{\sqrt{n}}\ ;\ \overline{Y}_n + \sigma\frac{\Phi^{-1}(1-\alpha/2)}{\sqrt{n}}\right].$$
   Notez bien que σ est connu, donc il s'agit bien d'une quantité calculable.
   (b) La longueur de l'intervalle de confiance de niveau $1-\alpha$ est donnée par la fonction
   $$L(n) = 2\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2).$$
   Pour $\sigma = 3$ et $1-\alpha = 0.95$, on a pour tout n,
   $$L(n) \leq 2 \Leftrightarrow \frac{3}{\sqrt{n}}\Phi^{-1}(0.975) \leq 1 \Leftrightarrow (3\Phi^{-1}(0.975))^2 \leq n.$$
   De $0 < \Phi^{-1}(0.975) \leq 2$ on déduit que $(3\Phi^{-1}(0.975))^2 \leq 36$. Ainsi, il faut au minimum $n_0 = 36$ observations pour que l'intervalle de confiance $I(m)$ soit de longueur plus petite que 2.
   Pour $1-\alpha = 0.95$, $\sigma = 3$, $n = 25$ et $\overline{y}_{25} = 20$, la réalisation de l'intervalle de confiance $I(m)$ est :
   $$\begin{aligned} & \left[\overline{y}_n - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2)\ ;\ \overline{y}_n + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2)\right] \\ = & \left[20 - \frac{3}{\sqrt{25}}\Phi^{-1}(0.975)\ ;\ 20 + \frac{3}{\sqrt{25}}\Phi^{-1}(0.975)\right] \\ \simeq & [18.8\ ;\ 21.2]. \end{aligned}$$
   Attention à bien distinguer la valeur observée $\overline{y}_n = \overline{Y}_n(\omega)$ de la v.a. $\overline{Y}_n$, et la réalisation de l'intervalle de confiance obtenue à partir de $\overline{y}_n$ de l'intervalle de confiance aléatoire obtenu avec $\overline{Y}_n$.
   (c) On rejette H_0 si
   $$m_0 \notin \left[\overline{Y}_n - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2), \overline{Y}_n + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2)\right],$$
   c'est-à-dire
   $$\sqrt{n}\left|\frac{\overline{Y}_n - m_0}{\sigma}\right| > \Phi^{-1}(1-\alpha/2),$$

ou encore

$$\alpha > 2\left(1 - \Phi\left(\sqrt{n}\left|\frac{\overline{Y}_n - m_0}{\sigma}\right|\right)\right).$$

La p-valeur du test est donc donnée par

$$\alpha_0 = 2\left(1 - \Phi\left(\sqrt{n}\left|\frac{\overline{y}_n - m_0}{\sigma}\right|\right)\right).$$

Pour $\sigma = 3$, $n = 25$, $\overline{y}_{25} = 20$ et $m_0 = 18.9$, on obtient donc la p-valeur

$$\alpha_0 = 2\left(1 - \Phi\left(\sqrt{25}\left|\frac{20 - 18.9}{3}\right|\right)\right) \simeq 2(1 - \Phi(1.83)) \simeq 2(1 - 0.97) \simeq 0.06.$$

On rejette donc H_0 au niveau 10%, et on accepte H_0 aux niveaux 5% et 1%.

2. (a) Le modèle de régression associé est $Y = X\beta + \sigma\varepsilon$, avec $X = [1, \cdots, 1]^t \in \mathbb{R}^n$, $\beta = m \in \mathbb{R}$ et $\varepsilon = [\varepsilon_1, \cdots, \varepsilon_n]^t \in \mathbb{R}^n$. L'estimateur des moindres carrés est ainsi $\widehat{\beta} = (X^tX)^{-1}X^tY = \overline{Y}_n$.

   (b) Étant donné le modèle de régression, $Y \sim \mathcal{N}(m(1 \cdots 1)^t, \sigma^2 I_n)$ et la matrice de projection associée est :

$$P = X(X^tX)^{-1}X^t = \frac{1}{n}\begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}.$$

   De plus, $PY = (\overline{Y}_n \cdots \overline{Y}_n)^t$, $P_\perp Y = (I_n - P)Y = Y - (\overline{Y}_n \cdots \overline{Y}_n)^t$, $P(m(1 \cdots 1)^t) = m(1 \cdots 1)^t$ et $P_\perp(m(1 \cdots 1)^t) = 0$. D'après le théorème de Cochran,
   - $PY \sim \mathcal{N}(m(1 \cdots 1)^t), \sigma^2 P)$, $P_\perp Y \sim \mathcal{N}(0, \sigma^2 P_\perp)$;
   - PY et $P_\perp Y$ sont indépendants ;
   - $\frac{\|P(Y - m(1 \cdots 1)^t)\|^2}{\sigma^2} = \frac{n(\overline{Y}_n - m)^2}{\sigma^2} \sim \chi^2(1)$, $\frac{\|Y - (\overline{Y}_n \cdots \overline{Y}_n)^t\|^2}{\sigma^2} = \frac{\sum_{i=1}^n (Y_i - \overline{Y}_n)^2}{\sigma^2} \sim \chi^2(n-1)$.

   (c) On vient de voir que

$$(n-1)\widehat{\sigma}_n^2/\sigma^2 = \sum_{i=1}^n (Y_i - \overline{Y}_n)^2/\sigma^2 \sim \chi^2(n-1).$$

   Ainsi, $\mathbb{E}[(n-1)\widehat{\sigma}_n^2/\sigma^2] = n - 1$, c'est-à-dire $\mathbb{E}[\widehat{\sigma}_n^2] = \sigma^2$ donc $\widehat{\sigma}_n^2$ est sans biais. De plus, la loi des grands nombres et le théorème de continuité impliquent que

$$s_n^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2 = \frac{1}{n}\sum_{i=1}^n Y_i^2 - \overline{Y}_n^2 \xrightarrow[n\to\infty]{p.s.} \mathbb{E}[Y_1^2] - m^2 = \sigma^2,$$

   i.e. s_n^2 est un estimateur convergent de σ^2. Ainsi, $\widehat{\sigma}_n^2 = \frac{n}{n-1}s_n^2 \xrightarrow{P} \sigma^2$.

   (d) D'après le cours et ce qui vient d'être dit, un intervalle de confiance de niveau $(1 - \alpha)$ pour σ^2 est donné par

$$\left[\frac{(n-1)\widehat{\sigma}_n^2}{c_{n-1}(1 - \alpha/2)}, \frac{(n-1)\widehat{\sigma}_n^2}{c_{n-1}(\alpha/2)}\right]$$

   où $c_{n-1}(\alpha/2)$ et $c_{n-1}(1 - \alpha/2)$ sont les quantiles d'ordres $\alpha/2$ et $1 - \alpha/2$ d'une loi χ^2_{n-1} (ici $p = 1$ si p est le nombre de variables explicatives). Ainsi, le test consistant à rejeter H_0 si

$$3 \notin \left[\frac{(n-1)\widehat{\sigma}_n^2}{c_{n-1}(1 - \alpha/2)}, \frac{(n-1)\widehat{\sigma}_n^2}{c_{n-1}(\alpha/2)}\right]$$

   i.e., si

$$\widehat{\sigma}_n^2 > 3\frac{c_{n-1}(1 - \alpha/2)}{n-1} \quad \text{ou} \quad \widehat{\sigma}_n^2 < 3\frac{c_{n-1}(\alpha/2)}{n-1}$$

   est de niveau α.

(e) D'après le cours, on sait que :

$$\frac{\widehat{\beta}-\beta}{\widehat{\sigma}_n\sqrt{1/n}} = \sqrt{n}\frac{\overline{Y}_n - m}{\widehat{\sigma}_n} \sim \mathcal{T}(n-1).$$

(f) D'après le cours et ce qui vient d'être dit, un intervalle de confiance de niveau $1-\alpha$ pour m est donné par

$$J(m) = \left[\overline{Y}_n - \widehat{\sigma}_n\frac{t_{n-1}(1-\alpha/2)}{\sqrt{n}}, \overline{Y}_n + \widehat{\sigma}_n\frac{t_{n-1}(1-\alpha/2)}{\sqrt{n}}\right].$$

où $t_{n-1}(1-\alpha/2)$ est le quantile d'ordre $1-\alpha/2$ d'une loi de Student $\mathcal{T}(n-1)$.
On peut construire un test de niveau α en rejetant H_0 si $m_0 \notin J(m)$, c'est-à-dire si

$$\sqrt{n}\left|\frac{\overline{Y}_n - m_0}{\widehat{\sigma}_n}\right| > t_{n-1}(1-\alpha/2).$$

(g) D'après la question précédente, un intervalle de confiance unilatère de niveau $1-\alpha$ pour m est :

$$\left]-\infty, \overline{Y}_n - \frac{\widehat{\sigma}_n}{\sqrt{n}}t_{n-1}(\alpha)\right],$$

dont l'intersection avec $\Theta_0 = [m_0, +\infty]$ est vide si et seulement si

$$\overline{Y}_n < m_0 + \frac{\widehat{\sigma}_n}{\sqrt{n}}t_{n-1}(\alpha).$$

Le test consistant à rejeter H_0 lorsque $\overline{Y}_n < m_0 + \frac{\widehat{\sigma}_n}{\sqrt{n}}t_{n-1}(\alpha)$ est donc de niveau α.
Calculons maintenant la p-valeur : on rejette H_0 si et seulement si

$$\begin{aligned}\overline{Y}_n < c_\alpha(Y) &\Leftrightarrow \sqrt{n}\frac{\overline{Y}_n - m_0}{\widehat{\sigma}_n} < t_{n-1}(\alpha)\\ &\Leftrightarrow F_{\mathcal{T}(n-1)}\left(\sqrt{n}\frac{\overline{Y}_n - m_0}{\widehat{\sigma}_n}\right) < \alpha.\end{aligned}$$

Ainsi, la p-valeur du test est :

$$\alpha_0 = F_{\mathcal{T}(n-1)}\left(\sqrt{n}\frac{\overline{Y}_n(\omega) - m_0}{\widehat{\sigma}_n(\omega)}\right).$$

Pour $m_0 = 12,5$, $n = 25$, $\overline{Y}_{25}(\omega) = 12$ et $\widehat{\sigma}_n^2(\omega) = 1,69$, la p-valeur est :

$$\alpha_0 = F_{\mathcal{T}(24)}\left(\sqrt{25}\left(\frac{12-12.5}{\sqrt{1.69}}\right)\right) \simeq F_{\mathcal{T}(24)}(-1,92) \simeq 0.03.$$

On rejette donc H_0 au niveau 5%.
Remarque : On aurait pu anticiper la forme du test : $T = \mathbb{1}_{\overline{Y}_n < C_{\alpha,n}}$, où $C_{\alpha,n}$ est ici une quantité aléatoire, comme nous allons le voir. Par calibration :

$$\begin{aligned}\sup_{m\geq m_0}\mathbb{P}(\overline{Y}_n < C_{\alpha,n}) &= \sup_{m\geq m_0}\mathbb{P}\left(\sqrt{n}\frac{\overline{Y}_n - m}{\widehat{\sigma}_n} < \sqrt{n}\frac{C_{\alpha,n} - m}{\widehat{\sigma}_n}\right)\\ &\leq \mathbb{P}\left(\sqrt{n}\frac{\overline{Y}_n - m}{\widehat{\sigma}_n} < \sqrt{n}\frac{C_{\alpha,n} - m_0}{\widehat{\sigma}_n}\right)\\ &= F_{\mathcal{T}(n-1)}\left(\sqrt{n}\frac{C_{\alpha,n} - m_0}{\widehat{\sigma}_n}\right),\end{aligned}$$

si l'on suppose que $\frac{C_{\alpha,n}-m_0}{\widehat{\sigma}_n}$ est déterministe. Le dernier terme est égal à α lorsque

$$C_{\alpha,n} = m_0 + \frac{\widehat{\sigma}_n}{\sqrt{n}} t_{n-1}(\alpha),$$

et $\frac{C_{\alpha,n}-m_0}{\widehat{\sigma}_n}$ est bien déterministe. Vérifions maintenant le niveau du test :

$$\begin{aligned}
\sup_{m \geq m_0} \mathbb{P}(T=1) &= \sup_{m \geq m_0} \mathbb{P}\left(\sqrt{n}\frac{\overline{Y}_n - m}{\widehat{\sigma}_n} < (m_0 - m)\frac{\sqrt{n}}{\widehat{\sigma}_n} + t_{n-1}(\alpha)\right) \\
&\leq \mathbb{P}\left(\sqrt{n}\frac{\overline{Y}_n - m}{\widehat{\sigma}_n} < t_{n-1}(\alpha)\right) \\
&= F_{\mathcal{T}(n-1)}\left(t_{n-1}(\alpha)\right) \\
&= \alpha,
\end{aligned}$$

car pour tout $m \geq m_0$, $(m_0 - m)\frac{\sqrt{n}}{\widehat{\sigma}_n} \leq 0$. Le test T est donc bien de taille α.

Exercice 2

1. On note $\beta = (a,b)^t$ et

$$X = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}.$$

   Le modèle est identifiable si et seulement si X est de rang $p=2$, c'est-à-dire si et seulement si les vecteurs $(t_i)_{1\leq i \leq n}$ et $(1)_{1 \leq i \leq n}$ ne sont pas colinéaires. Ceci est équivalent à l'existence de $i \neq j$ tel que $t_i \neq t_j$, ou, considérant l'hypothèse $\sum_{i=1}^n t_i = 0$, à l'existence de $i \in [\![1,n]\!]$ tel que $t_i \neq 0$.
   On peut remarquer aussi que les deux vecteurs colonnes sont orthogonaux et non-nuls, ou que X^tX est inversible (déterminant non-nul).

2. L'estimateur des moindres carrés de β est donné par :

$$\begin{aligned}
\widehat{\beta} &= (X^tX)^{-1}X^tY = \begin{pmatrix} n & \sum_i t_i \\ \sum_i t_i & \sum_i t_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_i Y_i \\ \sum_i t_i Y_i \end{pmatrix} \\
&= \begin{pmatrix} n & 0 \\ 0 & n v_t \end{pmatrix}^{-1} \begin{pmatrix} n\overline{Y} \\ n\rho \end{pmatrix} = \begin{pmatrix} \overline{Y} \\ \frac{\rho}{v_t} \end{pmatrix},
\end{aligned}$$

   donc

$$\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} = \begin{pmatrix} \overline{Y} \\ \frac{\rho}{v_t} \end{pmatrix}.$$

   Par ailleurs, l'estimateur de σ^2 est défini par :

$$\begin{aligned}
\widehat{\sigma}^2 &= \frac{1}{n-2}\left\|Y - X\widehat{\beta}\right\|^2 = \frac{1}{n-2}\sum_i \left(Y_i - \overline{Y} - \frac{\rho}{v_t}t_i\right)^2 \\
&= \frac{1}{n-2}\Big(\sum_i (Y_i - \overline{Y})^2 + \frac{\rho^2}{v_t^2}\sum_i t_i^2 - 2\frac{\rho}{v_t}\sum_i t_i Y_i + 0\Big) \\
&= \frac{1}{n-2}\Big(n v_Y + \frac{\rho^2}{v_t^2} n v_t - 2\frac{\rho}{v_t}\, n\rho\Big) = \frac{n}{n-2}\Big(v_Y - \frac{\rho^2}{v_t}\Big).
\end{aligned}$$

Pour rappel, $\widehat{\beta}$ consiste en les coordonnées dans la base des colonnes de X, de la projection de Y sur l'espace engendré par les colonnes de X, noté $\mathcal{M}(X)$, et

$$\widehat{\sigma}^2 = \frac{1}{n-2}\|Y - X\widehat{\beta}\|^2 = \frac{1}{n-2}\|P_{\mathcal{M}(X)^\perp}Y\|^2.$$

D'après le cours, les estimateurs $\widehat{\beta}$ et $\widehat{\sigma}^2$ sont indépendants et ont pour loi :

$$\begin{aligned}\begin{pmatrix}\widehat{a}\\ \widehat{b}\end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix}a\\ b\end{pmatrix}, \sigma^2(X^tX)^{-1}\right) = \mathcal{N}\left(\begin{pmatrix}a\\ b\end{pmatrix}, \sigma^2\begin{pmatrix}\frac{1}{n} & 0\\ 0 & \frac{1}{nv_t}\end{pmatrix}\right);\\ \frac{(n-2)\widehat{\sigma}^2}{\sigma^2} &\sim \chi^2_{(n-2)}.\end{aligned}$$

3. D'après le cours, on a les intervalles de confiance de niveau $1-\alpha$ suivants :
— pour a
$$\left[\widehat{a} - \frac{t_{n-2}(1-\alpha/2)\widehat{\sigma}}{\sqrt{n}}, \widehat{a} + \frac{t_{n-2}(1-\alpha/2)\widehat{\sigma}}{\sqrt{n}}\right]$$
— pour b
$$\left[\widehat{b} - \frac{t_{n-2}(1-\alpha/2)\widehat{\sigma}}{\sqrt{nv_t}}, \widehat{b} + \frac{t_{n-2}(1-\alpha/2)\widehat{\sigma}}{\sqrt{nv_t}}\right]$$
Comme $\mathbb{P}(A\cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$, on en déduit que

$$\begin{aligned}&\mathbb{P}\left(|\widehat{a}-a| > \frac{\widehat{\sigma}}{\sqrt{n}}t_{n-2}(1-\alpha/4) \text{ ou } |\widehat{b}-b| > \frac{\widehat{\sigma}}{\sqrt{nv_t}}t_{n-2}(1-\alpha/4)\right)\\ &\leq \mathbb{P}\left(|\widehat{a}-a| > \frac{\widehat{\sigma}}{\sqrt{n}}t_{n-2}(1-\alpha/4)\right) + \mathbb{P}\left(|\widehat{b}-b| > \frac{\widehat{\sigma}}{\sqrt{nv_t}}t_{n-2}(1-\alpha/4)\right)\\ &\leq \alpha/2 + \alpha/2 = \alpha.\end{aligned}$$

Donc
$$\mathbb{P}\left(|\widehat{a}-a| \leq \frac{\widehat{\sigma}}{\sqrt{n}}t_{n-2}(1-\alpha/4) \text{ et } |\widehat{b}-b| \leq \frac{\widehat{\sigma}}{\sqrt{nv_t}}t_{n-2}(1-\alpha/4)\right) \geq 1-\alpha.$$
Ceci implique que la région rectangulaire donnée par

$$\left\{(a,b)\in\mathbb{R}^2 \mid |\widehat{a}-a| \leq \frac{\widehat{\sigma}}{\sqrt{n}}t_{n-2}(0.9875) \text{ et } |\widehat{b}-b| \leq \frac{\widehat{\sigma}}{\sqrt{nv_t}}t_{n-2}(0.9875)\right\}$$

est une région de confiance de niveau 0.95 pour (a,b).
Remarque : il faut prendre les quantiles à l'ordre $1-\alpha/4$ et non plus $1-\alpha/2$ et les événements $\{|\widehat{a}-a| > \widehat{\sigma}t_{n-2}(1-\alpha/4)/\sqrt{n}\}$ et $\{|\widehat{b}-b| > \widehat{\sigma}t_{n-2}(1-\alpha/4)/\sqrt{nv_t}\}$ ne sont pas indépendants.

4. D'après le cours, un ellipsoïde de niveau $1-\alpha$ pour (a,b) est donné par

$$\begin{aligned}\mathcal{E} &= \left\{(x,y)\in\mathbb{R}^2\colon \frac{1}{2\widehat{\sigma}^2}\left(\widehat{a}-x, \widehat{b}-y\right)\begin{pmatrix}n & 0\\ 0 & nv_t\end{pmatrix}\begin{pmatrix}\widehat{a}-x\\ \widehat{b}-y\end{pmatrix} \leq f^2_{n-2}(1-\alpha)\right\}\\ &= \left\{(x,y)\in\mathbb{R}^2\colon (\widehat{a}-x)^2 + v_t(\widehat{b}-y)^2 \leq \frac{2}{n}\widehat{\sigma}^2 f^2_{n-2}(1-\alpha)\right\},\end{aligned}$$

où $f^2_{n-2}(1-\alpha)$ est le quantile d'ordre $1-\alpha$ d'une loi de Fisher $\mathcal{F}^2_{n-2}$.
Remarque : la forme de cette région diffère de la région rectangulaire de la question 3.

5. On pose $B=(5,-8)$, de sorte que $B\beta = 5a-8b$. La loi de $B(\widehat{\beta}-\beta)$ est la loi normale d'espérance nulle et de variance :

$$B\sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{nv_t} \end{pmatrix} B^t = \sigma^2 \left(\frac{25}{n} + \frac{64}{nv_t} \right).$$

On peut alors construire un intervalle de confiance à partir de la loi de Student

$$\frac{(5\widehat{a}-8\widehat{b})-(5a-8b)}{\sqrt{(\frac{25}{n}+\frac{64}{nv_t})\widehat{\sigma}^2}} \sim \mathcal{T}_{(18-2)}.$$

En effet, l'indépendance entre numérateur et dénominateur est une conséquence du Théorème de Cochran et

$$\underbrace{\frac{(5\widehat{a}-8\widehat{b})-(5a-8b)}{\sqrt{(\frac{25}{n}+\frac{64}{nv_t})\sigma^2}}}_{\mathcal{N}(0,1)} \times \underbrace{\sqrt{\frac{\sigma^2}{\widehat{\sigma}^2(n-2)}}}_{1/\sqrt{\chi^2(n-2)}} \times \underbrace{\frac{1}{\frac{1}{\sqrt{n-2}}}}_{1/(1/\sqrt{\text{ddl}})} \sim \mathcal{T}_{(18-2)}.$$

On en déduit que

$$\mathbb{P}\left(\left| \frac{(5\widehat{a}-8\widehat{b})-(5a-8b)}{\sqrt{\left(\frac{25}{n}+\frac{64}{nv_t}\right)\widehat{\sigma}^2}} \right| \leq t_{16}(0.975) \right) = 0.95,$$

ce qui conduit à l'intervalle de confiance pour $5a-8b$ de niveau 95% :

$$\left[(5\widehat{a}-8\widehat{b}) - t_{16}(0.975)\sqrt{\left(\frac{25}{n}+\frac{64}{nv_t}\right)\widehat{\sigma}^2},\ (5\widehat{a}-8\widehat{b}) + t_{16}(0.975)\sqrt{\left(\frac{25}{n}+\frac{64}{nv_t}\right)\widehat{\sigma}^2} \right].$$

6. Notons tout d'abord que H_0 s'écrit encore $a-b=0$, de sorte qu'on effectue les mêmes calculs qu'à la question précédente avec $B=(1,-1)$. Cette fois, $B\widehat{\beta}$ suit la loi normale de moyenne $a-b$ et de variance :

$$B\sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{nv_t} \end{pmatrix} B^t = \sigma^2 \left(\frac{1}{n} + \frac{1}{nv_t} \right),$$

ce qui donne pour $n=22$:

$$\sqrt{22}\frac{(\widehat{a}-\widehat{b})-(a-b)}{\widehat{\sigma}\left(1+\frac{1}{v_t}\right)^{1/2}} \sim \mathcal{T}_{(22-2)}.$$

Sous H_0, $a=b$, donc

$$\mathbb{P}\left(\sqrt{22}\frac{|\widehat{a}-\widehat{b}|}{\widehat{\sigma}\left(1+\frac{1}{v_t}\right)^{1/2}} > t_{20}(0.995) \right) = 0.01.$$

donc le test rejetant H_0 lorsque

$$|\widehat{a}-\widehat{b}| > \frac{\widehat{\sigma}}{\sqrt{22}}\left(1+\frac{1}{v_t}\right)^{1/2} t_{20}(0.995)$$

est de niveau 1%.

7. Dans la question précédente, on a construit le test

$$T_1 = \mathbb{1}_{|\widehat{a}-\widehat{b}| > \frac{\widehat{\sigma}}{\sqrt{n}}\left(1+\frac{1}{v_t}\right)^{1/2} t_{n-2}(1-\alpha/2)} = \mathbb{1}_{(\widehat{a}-\widehat{b})^2 > \frac{\widehat{\sigma}^2}{n}\left(1+\frac{1}{v_t}\right) t_{n-2}(1-\alpha/2)^2},$$

que l'on cherche à comparer à $T_2 = \mathbb{1}_{\Theta_0 \cap \mathcal{E} = \emptyset}$, où l'ellipsoïde de confiance de niveau $1-\alpha$ pour (a,b) est donné par

$$\mathcal{E} \quad = \quad \left\{ (x,y) \in \mathbb{R}^2 \colon (\widehat{a}-x)^2 + v_t(\widehat{b}-y)^2 \leq \frac{2}{n}\widehat{\sigma}^2 f^2_{n-2}(1-\alpha) \right\},$$

avec $f^2_{n-2}(1-\alpha)$ le quantile d'ordre $1-\alpha$ de la loi de Fisher $\mathcal{F}^2_{n-2}$.
Vérifions d'abord que T_2 est de niveau α (la démonstration est similaire à celle du cours pour les intervalles) :

$$\begin{aligned}
\sup_{(a,b)\in\Theta_0} \mathbb{P}(T_2 = 1) &= \sup_{(a,b)\in\Theta_0} \mathbb{P}(\Theta_0 \cap \mathcal{E} = \emptyset) \\
&\leq \sup_{(a,b)\in\Theta_0} \mathbb{P}((a,b) \notin \mathcal{E}) \\
&= \alpha,
\end{aligned}$$

où l'on a utilisé que si $(a,b) \in \Theta_0$ et $\Theta_0 \cap \mathcal{E} = \emptyset$, alors $(a,b) \notin \mathcal{E}$.
Comparons maintenant T_2 à T_1 en rapprochant leurs régions de rejet :

$$\begin{aligned}
T_2 = 1 &\iff \forall (x,y) \in \Theta_0 : (x,y) \notin \mathcal{E} \\
&\iff \forall x \in \mathbb{R} : (\widehat{a}-x)^2 + v_t(\widehat{b}-x)^2 > \frac{2}{n}\widehat{\sigma}^2 f^2_{n-2}(1-\alpha) \\
&\iff \forall x \in \mathbb{R} : Ax^2 + Bx + C > 0,
\end{aligned}$$

où $A = 1 + v_t$, $B = -2(\widehat{a} + v_t\widehat{b})$ et $C = \widehat{a}^2 + v_t\widehat{b}^2 - \frac{2}{n}\widehat{\sigma}^2 f^2_{n-2}(1-\alpha)$. Cela revient à s'assurer que le trinôme de coefficients (A,B,C) n'a pas de racines, c'est-à-dire que son discriminant est strictement négatif :

$$\begin{aligned}
B^2 - 4AC < 0 &\iff 4(\widehat{a} + v_t\widehat{b})^2 - 4(1+v_t)\left(\widehat{a}^2 + v_t\widehat{b}^2 - \frac{2}{n}\widehat{\sigma}^2 f^2_{n-2}(1-\alpha)\right) < 0 \\
&\iff \widehat{a}^2 + v_t^2\widehat{b}^2 + 2v_t\widehat{a}\widehat{b} - \widehat{a}^2 - v_t\widehat{a}^2 - v_t\widehat{b}^2 - v_t^2\widehat{b}^2 + (1+v_t)\frac{2}{n}\widehat{\sigma}^2 f^2_{n-2}(1-\alpha) < 0 \\
&\iff 2v_t\widehat{a}\widehat{b} - v_t\widehat{a}^2 - v_t\widehat{b}^2 < -(1+v_t)\frac{2}{n}\widehat{\sigma}^2 f^2_{n-2}(1-\alpha) \\
&\iff -v_t(\widehat{a}-\widehat{b})^2 < -(1+v_t)\frac{2}{n}\widehat{\sigma}^2 f^2_{n-2}(1-\alpha) \\
&\iff (\widehat{a}-\widehat{b})^2 > \frac{\widehat{\sigma}^2}{n}\left(1+\frac{1}{v_t}\right) 2 f^2_{n-2}(1-\alpha).
\end{aligned}$$

Ainsi, T_1 et T_2 sont différents, puisque les quantiles le sont. En effet, on peut montrer que $t_{n-2}(1-\alpha/2)^2 = f^1_{n-2}(1-\alpha) < 2f^2_{n-2}(1-\alpha)$. Par exemple, pour $\alpha = 1\%$ et $n = 22$, $t_{n-2}(1-\alpha/2)^2 \approx 8.1$ et $2f^2_{n-2}(1-\alpha) \approx 11.7$. On remarque qu'alors T_2 rejette moins souvent que T_1. Ceci n'est pas surprenant car T_1 est de taille α, ce qui n'est pas le cas de T_2.

8. La valeur prédite est

$$\widehat{Y}^{(p)} = [1,2]\widehat{\beta} = \widehat{a} + 2\widehat{b} = \overline{Y} + 2\frac{\rho}{v_t}.$$

Prenons $\alpha = 0.02$. Si on note t_α le quantile d'ordre $(1-\frac{\alpha}{2})$ de la loi de Student à $(n-2)$ degrés de liberté, alors

$$\left[\widehat{Y}^{(p)} - t_\alpha \widehat{\sigma}\sqrt{1+\frac{1}{n}\left(1+\frac{4}{v_t}\right)}\ ;\ \widehat{Y}^{(p)} + t_\alpha \widehat{\sigma}\sqrt{1+\frac{1}{n}\left(1+\frac{4}{v_t}\right)}\right]$$

est un intervalle de prédiction de niveau $1-\alpha$ (qui vaut ici 98%) pour Y.

Exercice 3

1. Le modèle s'écrit $Y = X\beta + \varepsilon$, avec

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} 1 & W_1 & Z_1 \\ \vdots & \vdots & \vdots \\ 1 & W_n & Z_n \end{pmatrix}, \beta = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \text{ et } \varepsilon \sim \mathcal{N}(0, \sigma^2 I_n).$$

Les estimateurs des moindres carrés de a,b et c s'écrivent :

$$\widehat{\beta} = \begin{pmatrix} \widehat{a} \\ \widehat{b} \\ \widehat{c} \end{pmatrix} = (X^t X)^{-1} X^t Y.$$

On calcule

$$\begin{aligned} X^t X &= \begin{pmatrix} 1 & \dots & 1 \\ W_1 & \dots & W_n \\ Z_1 & \dots & Z_n \end{pmatrix} \begin{pmatrix} 1 & W_1 & Z_1 \\ \vdots & \vdots & \vdots \\ 1 & W_n & Z_n \end{pmatrix} = \begin{pmatrix} \|1\|^2 & \langle W, 1\rangle & \langle Z, 1\rangle \\ \langle W, 1\rangle & \|W\|^2 & \langle W, Z\rangle \\ \langle Z, 1\rangle & \langle W, Z\rangle & \|Z\|^2 \end{pmatrix} \\ &= \begin{pmatrix} n & n\overline{W} & n\overline{Z} \\ n\overline{W} & r^2 & r^2\sin\theta \\ n\overline{Z} & r^2\sin\theta & r^2 \end{pmatrix}. \end{aligned}$$

D'après l'énoncé, $\overline{W} = \overline{Z} = 0$ et $\cos\theta > 0$, donc :

$$\begin{aligned} X^t X &= \begin{pmatrix} n & 0 & 0 \\ 0 & r^2 & r^2\sin\theta \\ 0 & r^2\sin\theta & r^2 \end{pmatrix} \\ (X^t X)^{-1} &= \begin{pmatrix} 1/n & 0 & 0 \\ 0 & \frac{1}{r^2\cos^2\theta} & \frac{-\sin\theta}{r^2\cos^2\theta} \\ 0 & \frac{-\sin\theta}{r^2\cos^2\theta} & \frac{1}{r^2\cos^2\theta} \end{pmatrix}, \end{aligned}$$

puisque l'inverse de la matrice $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ est définie lorsque $\Delta = ad - bc \neq 0$ par $\frac{1}{\Delta}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. On obtient finalement :

$$\begin{aligned} \widehat{\beta} = \begin{pmatrix} \widehat{a} \\ \widehat{b} \\ \widehat{c} \end{pmatrix} &= (X^t X)^{-1} X^t Y = (X^t X)^{-1} \begin{pmatrix} \langle 1, Y\rangle \\ \langle W, Y\rangle \\ \langle Z, Y\rangle \end{pmatrix} \\ &= \begin{pmatrix} \overline{Y} \\ (r^2\cos^2\theta)^{-1}\langle W - (\sin\theta)Z, Y\rangle \\ (r^2\cos^2\theta)^{-1}\langle -(\sin\theta)W + Z, Y\rangle \end{pmatrix}. \end{aligned}$$

D'après le cours, on sait que :

$$\widehat{\beta}-\beta \sim \mathcal{N}(0,\sigma^2(X^tX)^{-1})$$

est indépendant de $\widehat{\sigma}^2 = \|Y - X\widehat{\beta}\|^2/(n-3)$, avec

$$\frac{(n-3)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-3)}.$$

2. Rappelons que $\cos\theta > 0$ par hypothèse. D'après le cours, un intervalle de confiance de niveau $1-\alpha$ pour c est donc donné par :

$$\left[\widehat{c} - \frac{\widehat{\sigma}}{r\cos\theta} t_{n-3}(1-\alpha/2), \widehat{c} + \frac{\widehat{\sigma}}{r\cos\theta} t_{n-3}(1-\alpha/2)\right].$$

Cet intervalle a pour longueur

$$\mathcal{L}(\theta) = \frac{2\widehat{\sigma}}{r\cos\theta} t_{n-3}(1-\alpha/2).$$

Donc

$$\mathcal{L}^2(\theta) = \frac{4\widehat{\sigma}^2}{r^2\cos^2\theta} t^2_{n-3}(1-\alpha/2)$$

et

$$\mathbb{E}\big(\mathcal{L}^2(\theta)\big) = \frac{4\mathbb{E}(\widehat{\sigma}^2)}{r^2\cos^2\theta} t^2_{n-3}(1-\alpha/2) = \frac{4\sigma^2}{r^2\cos^2\theta} t^2_{n-3}(1-\alpha/2),$$

car $\widehat{\sigma}^2$ est un estimateur sans biais de σ^2.
Ainsi, $\theta \mapsto \mathbb{E}\big(\mathcal{L}^2(\theta)\big)$ a les variations inverses de la fonction cosinus sur $]-\pi/2,\pi/2[$; l'espérance de la longueur est minimum pour $\theta = 0$, alors que l'espérance de la longueur tend vers $+\infty$ lorsque θ approche $-\pi/2$ ou $\pi/2$. Autrement dit, l'intervalle de confiance est le meilleur dans le cas $W \perp Z$ et est le pire dans le cas limite où W et Z sont colinéaires (ceci est normal car le modèle n'est plus identifiable en b, c dans ce cas).

3. D'après le cours, un intervalle de confiance de niveau $1-\alpha/3$ pour a est donné par

$$\mathbb{P}\left(|\widehat{a}-a| \leq \frac{\widehat{\sigma}}{\sqrt{n}} t_{n-3}(1-\alpha/6)\right) = 1-\alpha/3,$$

et pour b par

$$\mathbb{P}\left(|\widehat{b}-b| \leq \frac{\widehat{\sigma}}{r\cos\theta} t_{n-3}(1-\alpha/6)\right) = 1-\alpha/3,$$

ce qui implique (en utilisant la question 2)

$$\mathbb{P}\left(|\widehat{a}-a| \geq \frac{\widehat{\sigma}}{\sqrt{n}} t_{n-3}(1-\alpha/6)\right) \leq \alpha/3,$$
$$\mathbb{P}\left(|\widehat{b}-b| \geq \frac{\widehat{\sigma}}{r\cos\theta} t_{n-3}(1-\alpha/6)\right) \leq \alpha/3,$$
$$\mathbb{P}\left(|\widehat{c}-c| \geq \frac{\widehat{\sigma}}{r\cos\theta} t_{n-3}(1-\alpha/6)\right) \leq \alpha/3\,.$$

Ainsi, par la borne de l'union, l'événement

$$\left\{|\widehat{a}-a| \geq \frac{\widehat{\sigma}t_{n-3}(1-\alpha/6)}{\sqrt{n}}\right\} \cup \left\{|\widehat{b}-b| \geq \frac{\widehat{\sigma}t_{n-3}(1-\alpha/6)}{r\cos\theta}\right\} \cup \left\{|\widehat{c}-c| \geq \frac{\widehat{\sigma}t_{n-3}(1-\alpha/6)}{r\cos\theta}\right\}$$

est de probabilité inférieure à α. On en déduit un parallélépipède rectangle de confiance de niveau $97\% = 1-\alpha$ (on choisit $\alpha = 0.03$) pour (a,b,c) et pour $n = 27$ observations :

$$\left\{(a,b,c) \in \mathbb{R}^3 \colon |\widehat{a} - a| \leq \frac{\widehat{\sigma}}{3\sqrt{3}} t_{24}(0.995) \text{ et } \max(|\widehat{b} - b|, |\widehat{c} - c|) \leq \frac{\widehat{\sigma}}{r\cos\theta} t_{24}(0.995)\right\}.$$

4. D'après le cours, un ellipsoïde de confiance de niveau $1-\alpha = 0.97$ pour $\beta = (a,b,c)$ lorsque $n = 27$ est donné par

$$\begin{aligned}
\mathcal{E}(X) &= \left\{\beta \in \mathbb{R}^3 \colon \frac{1}{3\widehat{\sigma}^2}(\widehat{\beta} - \beta)^t (X^t X)(\widehat{\beta} - \beta) \leq f_{24}^3(0.97)\right\} \\
&= \left\{(a,b,c) \in \mathbb{R}^3 \colon \frac{1}{3\widehat{\sigma}^2}\left(n(\widehat{a} - a)^2 + r^2(\widehat{b} - b)^2 + r^2(\widehat{c} - c)^2 + 2r^2 \sin\theta(\widehat{b} - b)(\widehat{c} - c)\right) \leq f_{24}^3(0.97)\right\}
\end{aligned}$$

où $f_{24}^3(0.97)$ est le quantile d'ordre 0.97 d'une loi de Fisher $\mathcal{F}_{27}^3$.

Exercice 4

1. On note $G^j = (g_1^j \dots g_n^j)^t$, pour $j = 1, \dots, p$. Par hypothèse, les G^j sont deux à deux orthogonaux. On a

$$G^t G = \begin{pmatrix} (G^1)^t \\ \vdots \\ (G^p)t \end{pmatrix} (G^1 \dots G^p) = \begin{pmatrix} (G^1)^t G^1 & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & (G^p)^t G^p \end{pmatrix} = \begin{pmatrix} \delta_1^2 & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & \delta_p^2 \end{pmatrix}.$$

L'estimateur des moindres carrés de γ est donné par

$$\widehat{\gamma} = (G^t G)^{-1} G^t Z$$

avec

$$(G^t G)^{-1} = \begin{pmatrix} 1/\delta_1^2 & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & 1/\delta_p^2 \end{pmatrix}.$$

Ainsi

$$\widehat{\gamma}_j = \langle G^j, Z\rangle / \delta_j^2.$$

L'estimateur des moindres carrés de σ^2 est donné par :

$$\widehat{\sigma}^2 = \frac{1}{n-p} \|Z - G\widehat{\gamma}\|^2.$$

D'après le cours, $\widehat{\sigma}^2$ est indépendant de $\widehat{\gamma}$, et on a

$$(n-p)\frac{\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-p)}$$

et

$$\widehat{\gamma} - \gamma \sim \mathcal{N}(0, \sigma^2 (G^t G)^{-1}).$$

2. On en déduit un intervalle de confiance au niveau $1-\alpha$:

$$\mathbb{P}\left(\frac{(n-p)\widehat{\sigma}^2}{F^{-1}_{\chi^2_{(n-p)}}(1-\alpha/2)} \leq \sigma^2 \leq \frac{(n-p)\widehat{\sigma}^2}{F^{-1}_{\chi^2_{(n-p)}}(\alpha/2)}\right) = 1-\alpha.$$

Pour $n=25$ et $p=5$, on obtient l'intervalle de confiance pour σ^2 au niveau 95% :

$$\left[\frac{20\widehat{\sigma}^2}{F^{-1}_{\chi^2_{20}}(0.975)}, \frac{20\widehat{\sigma}^2}{F^{-1}_{\chi^2_{20}}(0.025)}\right],$$

où $F^{-1}_{\chi^2_{20}}(0.025) \simeq 9.6$ et $F^{-1}_{\chi^2_{20}}(0.975) \simeq 34.2$.

3. On sait que

$$\widehat{\gamma} - \gamma \sim \mathcal{N}(0, \sigma^2(G^tG)^{-1}) = \mathcal{N}\left(0, \sigma^2\begin{pmatrix} 1/\delta_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1/\delta_p^2 \end{pmatrix}\right).$$

Comme le vecteur $\widehat{\gamma}$ est gaussien et que les composantes $\widehat{\gamma}_k$ de $\widehat{\gamma}$ sont non corrélées, elles sont indépendantes. De plus, leurs lois marginales sont données par

$$\widehat{\gamma}_k - \gamma_k \sim \mathcal{N}\left(0, \frac{\sigma^2}{\delta_k^2}\right).$$

Par l'indépendance, la loi de $\sum_{j=1}^p \widehat{\gamma}_j$ est tout simplement :

$$\sum_{j=1}^p \widehat{\gamma}_j \sim \mathcal{N}\left(\sum_{j=1}^p \gamma_j, \sigma^2\sum_{j=1}^p \frac{1}{\delta_j^2}\right) = \mathcal{N}\left(\sum_{j=1}^p \gamma_j, \frac{\sigma^2}{\delta^2}\right).$$

4. On déduit de la question précédente

$$\frac{\sum_{j=1}^p \widehat{\gamma}_j - \sum_{j=1}^p \gamma_j}{\sigma/\delta} \sim \mathcal{N}(0,1)$$

mais on ne connaît pas σ^2 ! Il faut donc l'estimer et utiliser la loi de Student. En utilisant la question 1, les v.a. $\widehat{\gamma}-\gamma$ et $\frac{(n-p)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-p)}$ sont indépendantes. On a alors :

$$\frac{\frac{\sum_{j=1}^p \widehat{\gamma}_j - \sum_{j=1}^p \gamma_j}{\sigma/\delta}}{\sqrt{\frac{\widehat{\sigma}^2}{\sigma^2}}} \sim \mathcal{T}(n-p),$$

c'est-à-dire

$$\frac{\delta}{\widehat{\sigma}}\left(\sum_{j=1}^p \widehat{\gamma}_j - \sum_{j=1}^p \gamma_j\right) \sim \mathcal{T}(n-p).$$

Les calculs habituels mènent à l'intervalle de confiance suivant au niveau $1-\alpha$ pour $\sum_{j=1}^p \gamma_j$:

$$\left[\sum_{j=1}^p \widehat{\gamma}_j - \frac{\widehat{\sigma}}{\delta}F^{-1}_{\mathcal{T}(n-p)}(1-\alpha/2), \sum_{j=1}^p \widehat{\gamma}_j + \frac{\widehat{\sigma}}{\delta}F^{-1}_{\mathcal{T}(n-p)}(1-\alpha/2)\right].$$

Pour $n = 32$ et $p = 7$, on construit le test à partir de l'intervalle de confiance : on rejette H_0 : $\sum_{j=1}^{p} \gamma_j = 0$ au niveau 1% si

$$0 \notin \left[\sum_{j=1}^{7} \widehat{\gamma}_j - \frac{\widehat{\sigma}}{\delta} F_{\mathcal{T}(25)}^{-1}(0.995),\ \sum_{j=1}^{7} \widehat{\gamma}_j + \frac{\widehat{\sigma}}{\delta} F_{\mathcal{T}(25)}^{-1}(0.995) \right],$$

c'est-à-dire

$$\left| \sum_{j=1}^{7} \widehat{\gamma}_j \right| > \frac{\widehat{\sigma}}{\delta} F_{\mathcal{T}(25)}^{-1}(0.995).$$

Exercice 5

1. Réécriture sous forme de modèle linéaire.
   (a) La difficulté vient de ce que f est une fonction, donc l'espace des paramètres est énorme (dimension infinie). C'est un exemple typique de modèle non paramétrique.
   (b) Pour le modèle proposé, nous pouvons écrire pour tout $x \in [0, 1]$,

   $$f(x) = \sum_{j=1}^{p} \beta_j \phi_j(x)$$

   où $p = 2K + 1$ et pour tout $1 \leq k \leq K$ et $x \in [0, 1]$,

   $$\begin{cases} \beta_1 = a_0 \\ \beta_{2k} = a_k \\ \beta_{2k+1} = b_k \end{cases} ; \quad \begin{cases} \phi_1(x) = 1 \\ \phi_{2k}(x) = \cos(2\pi k x) \\ \phi_{2k+1}(x) = \sin(2\pi k x) \end{cases}$$

   Ainsi, en posant pour tout $1 \leq i \leq n$ et $1 \leq j \leq p$

   $$X_{i,j} = \phi_j(i/n),$$

   nous obtenons le modèle linéaire gaussien

   $$Y_i = \sum_{j=1}^{p} \beta_j \phi_j(i/n) + \varepsilon_i = (X\beta)_i + \varepsilon_i.$$

   (c) Si $p \leq n$, on peut vérifier que les vecteurs colonnes X_j, $1 \leq j \leq p$, de X satisfont la propriété suivante :

   $$X_j^t X_{j'} = 0 \text{ pour } j \neq j',\ \|X_1\|^2 = n \text{ et } \|X_j\|^2 = n/2 \text{ pour } j > 1. \tag{1}$$

   Pour montrer cela, nous utilisons de manière répétée que pour tout $\ell \in \{-(n-1), \ldots, (n-1)\}$,

   $$\sum_{i=1}^{n} e^{\mathbf{i}2\pi \ell i/n} = 0 \text{ si } \ell \neq 0 \text{ et } \sum_{i=1}^{n} e^{\mathbf{i}2\pi \ell i/n} = n \text{ si } \ell = 0.$$

   Allons-y, prouvons courageusement (1). Le fait que $\|X_1\|^2 = n$ ne pose pas problème. Pour tout $0 \leq k \leq K$, $1 \leq k' \leq K$, on a $1 \leq k + k' \leq 2K < p$, ce qui donne sympathiquement

   $$\begin{aligned} X_{2k+1}^t X_{2k'} &= \sum_{i=1}^{n} \sin(2\pi k i/n) \cos(2\pi k' i/n) \\ &= \sum_{i=1}^{n} \Im \left(e^{\mathbf{i}2\pi(k+k')i/n} - e^{\mathbf{i}2\pi(k'-k)i/n} \right) /2 \\ &= 0 \end{aligned}$$

car $e^{\mathbf{i}2\pi(k+k')i/n} \neq 1$ et $\sum_{i=1}^n e^{\mathbf{i}2\pi(k'-k)i/n} \in \mathbb{R}$. De plus, pour tout $0 \leq k, k' \leq K$ avec $k+k' > 0$,

$$\begin{aligned} X_{2k+1}^t X_{2k'+1} &= \sum_{i=1}^n \sin(2\pi k i/n) \sin(2\pi k' i/n) \\ &= \sum_{i=1}^n \Re\left(e^{\mathbf{i}2\pi(k-k')i/n} - e^{\mathbf{i}2\pi(k'+k)i/n}\right)/2 \\ &= \sum_{i=1}^n \Re\left(e^{\mathbf{i}2\pi(k-k')i/n}\right)/2 \end{aligned}$$

car $e^{\mathbf{i}2\pi(k'+k)i/n} \neq 1$ puisque $0 < k+k' < p \leq n$. Lorsque $k \neq k'$, on a $e^{\mathbf{i}2\pi(k-k')i/n} \neq 1$ donc $X_{2k+1}^t X_{2k'+1} = 0$. Si par contre $k = k' > 0$, alors $X_{2k+1}^t X_{2k'+1} = n/2$. De la même façon, on montre que pour tout $1 \leq k, k' \leq K$, $X_{2k}^t X_{2k'} = 0$ pour $k \neq k'$ et $= n/2$ si $k' = k$. Tout ceci prouve finalement que (1) est vraie !

Ainsi lorsque $n \geq p$, X est de plein rang et le modèle est identifiable. Nous pouvons calculer l'estimateur des moindres carrés à l'aide de la méthode usuelle :

$$\widehat{\beta} = (X^t X)^{-1} X^t Y,$$

Ici, par l'orthogonalité (1), $X^t X$ est la matrice de diagonale $(n, n/2, \ldots, n/2)$, ce qui donne pour tout $1 \leq k \leq K$

$$\begin{cases} \widehat{\beta}_1 = \overline{Y_n} \\ \widehat{\beta}_{2k} = 2n^{-1} \sum_{i=1}^n Y_i \cos(2\pi k i/n) \\ \widehat{\beta}_{2k+1} = 2n^{-1} \sum_{i=1}^n Y_i \sin(2\pi k i/n) \end{cases}$$

Ceci conduit aux estimateurs

$$\widehat{\mu}_i = \overline{Y_n} + \sum_{k=1}^K \left(\widehat{\beta}_{2k} \cos(2\pi k i/n) + \widehat{\beta}_{2k+1} \sin(2\pi k i/n)\right),$$

et

$$\widehat{f}(x) = \overline{Y_n} + \sum_{k=1}^K \left(\widehat{\beta}_{2k} \cos(2\pi k x) + \widehat{\beta}_{2k+1} \sin(2\pi k x)\right)$$

2. Overfitting et choix de modèle.
   (a) D'après le cours, on sait que $\|Y - X\widehat{\beta}\|^2/\sigma^2 \sim \chi^2_{n-p}$ donc

   $$r_n = n^{-1}\mathbb{E}(\|Y - X\widehat{\beta}\|^2) = \left(1 - \frac{p}{n}\right)\sigma^2.$$

   Ainsi, lorsque p est fixe et n tend vers l'infini, c'est-à-dire lorsque l'on dispose de plus en plus de données, le modèle permet de retrouver les données initiales Y_i à une erreur σ près (en écart-type). Cette erreur asymptotique est logique et due au fait que les données observées sont elles-mêmes bruitées.

   (b) Si $p = n$, alors $r_n = 0$ c'est-à-dire que $X\widehat{\beta} = Y$. Ainsi la fonction $\widehat{f}$ passe par les points de coordonnées $(i/n, Y_i)$, $1 \leq i \leq n$. Pour autant, comme le montre la question suivante, ceci n'est pas souhaitable car l'estimateur $\widehat{f}$ "colle" trop aux données : c'est le phénomène de sur-ajustement (ou overfitting).

   (c) La qualité de l'approximation dépend fortement du modèle choisi :
      - si $p = 3$, le modèle ne contient pas la fonction f ! Il y a donc un biais et l'approximation est trop "plate".

- si $p = 81$, le modèle contient bien la fonction f, mais on voit que le modèle est trop gros, ce qui se traduit par un sur-ajustement aux données. En fait, on estime plein de coefficients qui sont en réalité égaux à zéro donc inutilement.
- si $p = 11$, il s'agit du bon modèle, donc l'approximation est assez bonne (autant que σ le permet).

La moralité est que le choix du modèle est très important. Il faut prendre un modèle assez grand pour avoir peu de biais, mais pas trop grand pour avoir une variance petite. C'est le fameux compromis biais-variance.

Exercice 6

1. La projection orthogonale de X sur $F = \mathrm{vect}((1, \cdots, 1)^t)$ est
$$= F(F^tF)^{-1}F^tX = \overline{X}_n(1\cdots 1)^t.$$
Ainsi, le théorème de Cochran assure que $\overline{X}_n(1\cdots 1)^t$ est indépendant de $X - \overline{X}_n(1\cdots 1)^t$. Ceci implique que $\overline{X}_n$ est indépendant de $s_n^2 = n^{-1}||X - \overline{X}_n(1\cdots 1)^t||^2$. De plus, nous avons $\sqrt{n}\overline{X}_n \sim \mathcal{N}(0, 1)$ et le théorème de Cochran nous dit que
$$||X - \overline{X}_n(1\cdots 1)^t||^2 = ns_n^2 \sim \chi^2(n-1).$$
2. (a) Posons $Y_i = X_i - m$, de sorte que $\mathbb{E}Y_i = 0$ pour tout i. Ainsi, par définition
$$ns_n^2 = \sum_{i=1}^n (X_i - \overline{X}_n)^2 = \sum_{i=1}^n (Y_i - \overline{Y}_n)^2 = \sum_{i=1}^n Y_i^2 - 2\overline{Y}_n \sum_{i=1}^n Y_i + n(\overline{Y}_n)^2 = \sum_{i=1}^n Y_i^2 - n(\overline{Y}_n)^2.$$
Ainsi, on a
$$\begin{aligned}\mathbb{E}(ns_n^2) &= \sum_{i=1}^n \mathbb{E}(Y_i^2) - n\mathbb{E}\left(\left(n^{-1}\sum_{i=1}^n Y_i\right)^2\right)\\ &= n\sigma^2 - n\,\mathrm{Var}\left(n^{-1}\sum_{i=1}^n Y_i\right)\\ &= n\sigma^2 - \sigma^2\\ &= (n-1)\sigma^2.\end{aligned}$$
<u>Remarque</u> : on obtient le même résultat en travaillant directement sur X_i plutôt que sur Y_i et en remarquant que $\mathbb{E}(X_1^2) = \mathrm{Var}(X_1) + \mathbb{E}(X_1)^2 = \mathrm{Var}(X_1) + \mathbb{E}(\overline{X}_n)^2$.
Par suite, comme $\overline{X}_n$ et ns_n^2 sont indépendants et que les X_i sont i.i.d., on a pour tout $t \in \mathbb{R}$,
$$\mathbb{E}\left(s_n^2 \exp\left(itn\overline{X}_n\right)\right) = \mathbb{E}\big(s_n^2\big)\mathbb{E}\left(\exp\left(it\sum_{i=1}^n X_i\right)\right) = \mathbb{E}\big(s_n^2\big)\left(\mathbb{E}\big(\exp\left(itX_1\right)\big)\right)^n = \phi^n(t)\mathbb{E}\big(s_n^2\big).$$
(b) Remarquons tout d'abord que $\phi'(t) = i\,\mathbb{E}\big(X_1\exp\left(itX_1\right)\big)$ et $\phi''(t) = -\mathbb{E}\big(X_1^2\exp\left(itX_1\right)\big)$. En particulier, $\phi'(0) = im$. De plus, comme
$$\begin{aligned}ns_n^2 &= \sum_{j=1}^n X_j^2 - n(\overline{X}_n)^2\\ &= \sum_{j=1}^n X_j^2 - n^{-1}\sum_{k,\ell=1}^n X_kX_\ell\\ &= (1 - 1/n)\sum_{j=1}^n X_j^2 - n^{-1}\sum_{k\neq\ell} X_kX_\ell,\end{aligned}$$

nous obtenons

$$\begin{aligned}
\mathbb{E}\left(ns_n^2 e^{itn\overline{X}_n}\right) =&(1-1/n)\sum_{j=1}^n \mathbb{E}\left(X_j^2\prod_{h=1}^n e^{itX_h}\right) - n^{-1}\sum_{k\neq\ell}\mathbb{E}\left(X_kX_\ell\prod_{h=1}^n e^{itX_h}\right)\\
=&(1-1/n)\sum_{j=1}^n \mathbb{E}\left(X_j^2 e^{itX_j}\prod_{h\neq j} e^{itX_h}\right)\\
&-n^{-1}\sum_{k\neq\ell}\mathbb{E}\left(X_k e^{itX_k} X_\ell e^{itX_\ell}\prod_{h\notin\{k,\ell\}} e^{itX_h}\right)\\
=&(1-1/n)\sum_{j=1}^n \mathbb{E}(X_j^2 e^{itX_j})\mathbb{E}\left(\prod_{h\neq j} e^{itX_h}\right)\\
&-n^{-1}\sum_{k\neq\ell}\mathbb{E}\left(X_k e^{itX_k}\right)\mathbb{E}\left(X_\ell e^{itX_\ell}\right)\mathbb{E}\left(\prod_{h\notin\{k,\ell\}} e^{itX_h}\right),
\end{aligned}$$

en utilisant l'indépendance des X_i. Finalement, comme les X_i sont aussi identiquement distribués,

$$\mathbb{E}\left(ns_n^2 e^{itn\overline{X}_n}\right) = -(n-1)\phi''(t)\phi^{n-1}(t) + (n-1)(\phi'(t))^2\phi^{n-2}(t).$$

En utilisant la question (a), on obtient donc

$$\phi^n(t)(n-1)\sigma^2 = -(n-1)\phi''(t)\phi^{n-1}(t) + (n-1)(\phi'(t))^2\phi^{n-2}(t),$$

ce qui donne la relation voulue

$$\frac{\phi''}{\phi} - \left(\frac{\phi'}{\phi}\right)^2 = -\sigma^2, \quad \phi(0)=1, \phi'(0)=im.$$

(c) On résout l'équation différentielle. Comme

$$(\ln\phi)''(t) = \frac{\phi''}{\phi} - \left(\frac{\phi'}{\phi}\right)^2,$$

elle s'écrit

$$(\ln\phi)''(t) = -\sigma^2.$$

D'où $\ln\phi(t) = -\sigma^2 t^2/2 + at + b$, pour $a, b\in\mathbb{R}$, et les conditions initiales $\phi(0)=1$ et $\phi'(0)=im$ donnent

$$\phi(t) = \exp\left(\frac{-\sigma^2 t^2}{2} + imt\right) \text{ pour tout } t\in\mathbb{R}.$$

On reconnaît la fonction caractéristique de la loi $\mathcal{N}(m,\sigma^2)$ et on en déduit que les v.a. X_i suivent la loi $\mathcal{N}(m,\sigma^2)$, pour tout $i=1,\dots,n$.