CC2 2025/2026 – Durée 1h30

Les documents et appareils électroniques (calculatrice, téléphone, ordinateur, ...) sont interdits. ***Toutes les réponses doivent être justifiées****.*

Exercice 1 - Régression Polynomiale

Cet exercice est assez proche du cours. Vous avez le droit de l'invoquer, mais de manière précise et raisonnée (un "d'après le cours la réponse est celle de l'énoncé" ne suffira pas).

Pour $\theta = (\theta_0, \theta_1, \theta_2) \in \mathbb{R}^3$, on note Q_θ le polynôme $\theta_0 + \theta_1 x + \theta_2 x^2$.

Modèle sans répétition

Soient $\varepsilon_1, \varepsilon_2, \varepsilon_3$ des variables i.i.d. de loi $\mathcal{N}(0, \sigma^2)$ (σ^2 inconnu). On suppose que l'on observe $Y_1 = Q_\theta(-1) + \varepsilon_1$, $Y_2 = Q_\theta(0) + \varepsilon_2$, $Y_3 = Q_\theta(1) + \varepsilon_3$, pour un θ inconnu.

1. Écrire le modèle linéaire Gaussien associé (de la forme $Y = X\theta + \varepsilon$).
2. Ce modèle est-il identifiable ?
3. Donner l'expression de l'estimateur par moindre carrés de θ en fonction de Y et X, puis le calculer explicitement (plus de X dans la formule).
4. Donner la loi de $\hat{\theta}_0 + \hat{\theta}_2$.
5. En notant $\hat{Y} = X\hat{\theta}$ les valeurs ajustées, quelle est la loi de $\hat{Y} - Y$? Peut-on espérer en déduire un estimateur de σ^2 ?

Modèle avec répétition

On se donne $\tilde{Y}_{i,j}$, $i = 1, \ldots, n$, $j = 1, 2, 3$, indépendantes, et telles que $\tilde{Y}_{i,j} \sim Y_j$ (le Y_j du modèle précédent).

6. Écrire le modèle linéaire correspondant (forme $\tilde{Y} = X_n\theta + \tilde{\varepsilon}$). *Suggestion* : ranger dans l'ordre $\tilde{Y} = (\tilde{Y}_{1,1}, \ldots, \tilde{Y}_{n,1}, \tilde{Y}_{1,2}, \ldots, \tilde{Y}_{n,2}, \tilde{Y}_{1,3}, \ldots, \tilde{Y}_{n,3})^T$.
7. Trouver un estimateur de σ^2 (noté $\hat{\sigma}^2$, on pourra se contenter de donner son expression en fonction de X_n et Y).
8. On note W_0 le sous-espace vectoriel de $\mathbb{R}^3$ donné par $\{\theta_0 = \theta_2 = 0\}$, et on souhaite tester si $\theta \in W_0$. Donner les hypothèses du test de Fisher associé.
9. Donner la statistique de test (notée S_n, on pourra se contenter d'une expression 'géométrique'), sa loi lorsque $\theta \in H_0$, et le test de Fisher correspondant (noté T_n), pour un niveau $\alpha \in]0, 1[$ donné.
10. (*) Montrer que T_n est consistant (lorsque $n \to +\infty$).

Tournez S.V.P.

Exercice 2 - Circuit électrique

On mesure en parallèle la tension sur deux points très proches d'un même circuit électrique. Les deux sondes partagent des sources de bruit communes (température, alimentation, parasites électromagnétiques), ce qui introduit une corrélation entre leurs erreurs de mesure. Pour chaque expérience $i = 1, \dots, n$, on relève deux mesures bruitées :

$$X_{2i-1} = \mu - \frac{\delta}{2} + \varepsilon_{2i-1}, \qquad X_{2i} = \mu + \frac{\delta}{2} + \varepsilon_{2i},$$

où les $(\varepsilon_{2i}, \varepsilon_{2i-1})_{i=1,\dots,n}$ sont i.i.d. Gaussien $\mathcal{N}(0, \Sigma)$ avec $\Sigma_{1,1} = \Sigma_{2,2} = \sigma^2$ et $\Sigma_{1,2} = \rho\sigma^2$. On suppose que $|\rho| < 1$.

On cherche à estimer δ, σ et ρ. Pour cela, on introduit, pour $i = 1, \dots, n$,

$$P_i = \frac{X_{2i-1} + X_{2i}}{\sqrt{2}}, \qquad M_i = \frac{X_{2i} - X_{2i-1}}{\sqrt{2}},$$

ainsi que

$$S_M^2 = (1-\rho)^{-1} \sum_{i=1}^{n} (M_i - \overline{M_n})^2, \; S_P^2 = (1+\rho)^{-1} \sum_{i=1}^{n} (P_i - \overline{P_n})^2$$

1. On note $Y = (P_1, M_1, P_2, M_2, \dots, P_n, M_n)^T \in \mathbb{R}^{2n}$.
   (a) Montrer que Y est un vecteur Gaussien, calculer $\mathbb{E}[Y]$ et $\mathrm{Cov}[Y]$.
   (b) Écrire un modèle linéaire Gaussien de la forme $Y = X(\mu, \delta)^T + \Lambda\tilde{\varepsilon}$ avec deux matrices X et Λ à spécifier et où $\tilde{\varepsilon} \sim \mathcal{N}(0, I_{2n})$. Si $\rho \neq 0$, est-on dans le cadre du cours ?
2. (a) A partir de $(M_i)_{i=1,\dots,n}$, donner un estimateur $\widehat{\delta_n}$ de δ.
   (b) Construire un intervalle de confiance non-asymptotique de niveau $1-\alpha$ pour δ. On l'exprimera en fonction du quantile d'ordre $1-\alpha/2$ d'une loi classique.
3. On suppose dans cette question uniquement que ρ est connu. On considère, pour $a_n > 0$,
$$\widehat{\sigma^2} = a_n \left(S_M^2 + S_P^2 \right).$$
   (a) Quelle est la loi de $\widehat{\sigma^2}$? En déduire la valeur de a_n tel que $\widehat{\sigma^2}$ soit sans biais.
   (b) Déterminer un intervalle de confiance non-asymptotique de niveau $1-\alpha$ pour σ^2. On l'exprimera en fonction des quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ d'une loi classique.
4. (a) Quelle est la loi de S_M^2/S_P^2 ?
   (b) En déduire un intervalle de confiance de niveau $1-\alpha$ pour ρ. On l'exprimera en fonction de
$$r := \sum_{i=1}^{n} (M_i - \overline{M_n})^2 / \sum_{i=1}^{n} (P_i - \overline{P_n})^2,$$
   et des quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ d'une loi classique.

Exercice 3 - Encore une caractérisation Gaussienne

Soient $X_1, \dots, X_n$ i.i.d., centrées et de variance commune 1. Montrer l'équivalence suivante

$$X_1 \sim \mathcal{N}(0,1) \quad \Leftrightarrow \quad \exists \quad (a_1, \dots, a_n) \in \mathbb{R}^n \setminus (1, 0, \dots, 0) \quad \sum_{i=1}^{n} a_i X_i \sim X_1.$$

Indication pour le sens indirect : On pourra montrer ou admettre le résultat intermédiaire suivant : si g est une fonction vérifiant

— $g(0) = 0$;

— g est continue en 0 ;

— pour tout $t \in \mathbb{R}$, $g(t) = \sum_{i=1}^{n} a_i^2 g(a_i t)$, avec $\sum_{i=1}^{n} a_i^2 = 1$, et, pour tout $1 \leq i \leq n$, $|a_i| < 1$,

alors g est constante.

Solution 3 -

1. Comme $Q_\theta(-1) = \theta_0 - \theta_1 + \theta_2$, $Q_\theta(0) = \theta_0$ et $Q_\theta(1) = \theta_0 + \theta_1 + \theta_2$, le modèle s'écrit
$$Y = X\theta + \varepsilon,$$
avec $Y = (Y_1, Y_2, Y_3)^T$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)^T$, et
$$X = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$
2. Si $(\theta', \sigma') \neq (\theta, \sigma)$, comme X est de rang 3, on a
$$\mathcal{N}(X\theta', \sigma'^2 I_3) \not\sim \mathcal{N}(X\theta, \sigma^2 I_3),$$
il suffit par exemple de regarder les vecteurs moyenne et la matrice de covariance. Le modèle est alors identifiable.
3. On a (formule de cours ou minimisation du risque) $\hat{\theta} = (X^T X)^{-1} X^T Y$. Le reste est de l'algèbre linéaire :
$$X^T X = \begin{pmatrix} 3 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix},$$
$$(X^T X)^{-1} = \frac{1}{\mathrm{Det}(X^T X)} \begin{pmatrix} 4 & 0 & -4 \\ 0 & 2 & 0 \\ -4 & 0 & 6 \end{pmatrix},$$
avec $\mathrm{Det}(X^T X) = 4$. On en déduit
$$\begin{aligned} \hat{\theta}_0 &= Y_2, \\ \hat{\theta}_1 &= \frac{1}{2}(Y_3 - Y_1), \\ \hat{\theta}_2 &= \frac{1}{2}(Y_3 + Y_1) - Y_2. \end{aligned}$$
4. Toujours d'après le cours,
$$\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2 (X^T X)^{-1}),$$
avec $(X^T X)^{-1}$ calculé plus haut. On a alors
$$\hat{\theta}_0 + \hat{\theta}_2 \sim \mathcal{N}(\theta_0 + \theta_2, \sigma^2 a),$$

avec

$$\begin{aligned} a &= \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} (X^T X)^{-1} \begin{pmatrix} 1 & 0 & 1 \end{pmatrix} \\ &= 1/2. \end{aligned}$$

5. On peut recalculer $X\hat{\theta}$ manuellement, ou se rendre compte que comme $V(X)$ (espace vectoriel engendré par les colonnes de X) vaut $\mathbb{R}^3$, et que $\hat{Y} = \pi_{V(X)}(Y)$, on a nécessairement $\hat{Y} = Y$, et donc $\hat{Y} - Y \sim \delta_0$. On ne peut donc pas espérer en tirer un estimateur de σ^2.
6. Avec $\tilde{Y} = (\tilde{Y}_{1,1}, \dots, \tilde{Y}_{n,1}, \tilde{Y}_{1,2}, \dots, \tilde{Y}_{n,2}, \tilde{Y}_{1,3}, \dots, \tilde{Y}_{n,3})^T$, on a

$$\tilde{Y} = X_n\theta + \tilde{\varepsilon},$$

où $\tilde{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I_{3n})$ et

$$X_n = \begin{pmatrix} 1 & -1 & 1 \\ & \vdots_n & \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ & \vdots_n & \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ & \vdots_n & \\ 1 & 1 & 1 \end{pmatrix} \in M_{3n\times 3}(\mathbb{R}).$$

7. En notant $\hat{Y} = \pi_{V(X_n)}(\tilde{Y}) = X_n(X_n^T X_n)^{-1} X_n^T \tilde{Y}$, on pose

$$\hat{\sigma}^2 = \frac{\|\tilde{Y} - \hat{Y}\|^2}{3n-3} \sim \sigma^2 \chi^2(3n-3)/(3n-3),$$

en utilisant le cours et $\dim(V(X_n)) = 3$.
8. Le test de Fisher permet de tester $H_0 : \theta \in W_0$ contre $H_1 : \theta \notin W_0$.
9. En notant $V_0 = \{X\theta \mid \theta \in W_0\}$, la statistique de test est

$$S(\tilde{Y}) = \frac{\|\pi_{V(X_n)\cap V_0^\perp}(\tilde{Y})\|^2/2}{\hat{\sigma^2}},$$

comme $\dim(V(X_n) \cap V_0^\perp) = 2$. Sous H_0, $S(\tilde{Y}) \sim \mathcal{F}(2, 3n-3)$. Si q_n est le quantile d'ordre $1-\alpha$ d'une loi $S(\tilde{Y}) \sim \mathcal{F}(2, 3n-3)$, le test de Fisher s'écrit

$$T_n(\tilde{Y}) = \mathbb{1}_{S_n(\tilde{Y}) \geq q_n}.$$

10. Soit $\theta \notin H_0$. Notons $\bar{\theta}$ sa projection sur W_0 et $\delta = \|\bar{\theta} - \theta\| > 0$. On a

$$\|X\theta - X\bar{\theta}\|^2 \geq \lambda_{\min}\|\bar{\theta} - \theta\| = \lambda_{\min}\delta^2,$$

où $\lambda_{\min}$ est la plus petite valeur propre de $X_n^T X_n$. Or $X_n^T X_n = nX^T X$, et $X^T X$ est définie positive. On en déduit alors que $\lambda_{\min} \geq n\lambda$, pour un $\lambda > 0$, et donc $\|X\theta - X\bar{\theta}\| \geq \sqrt{n\lambda}\delta$.

Comme $\|\pi_{V(X_n)\cap V_0^\perp}(\tilde{Y})\|^2 \sim \chi^2(2, \|X(\theta - \bar{\theta})\|)$, on en déduit que

$$S_n(\tilde{Y}) \sim \frac{((\|X(\theta - \bar{\theta})\| + Z_1)^2 + Z_2^2)/2}{\sum_{j=1}^{3n-3} Z_j^2/(3n-3)},$$

où les Z_j sont i.i.d. de loi $\mathcal{N}(0,1)$. Comme $\|X\theta - X\bar{\theta}\| \geq \sqrt{n\lambda}\delta$, on déduit que $((\|X(\theta - \bar{\theta})\| + Z_1)^2 + Z_2^2)/2 \xrightarrow[n\to+\infty]{\mathbb{P}} +\infty$. Par ailleurs, comme $\mathbb{E}(Z_1^2) = 1 < +\infty$, la loi des grands nombres donne $\sum_{j=1}^{3n-3} Z_j^2/(3n-3) \xrightarrow[n\to+\infty]{\mathbb{P}} 1$. On en déduit que

$$S_n(\tilde{Y}) \xrightarrow[n\to+\infty]{\mathbb{P}} +\infty.$$

Par ailleurs, en utilisant le même raisonnement et le Lemme de Slutsky, on peut montrer que $\mathcal{F}(2, 3n-3) \rightsquigarrow \chi^2(2)$. En notant q le quantile d'ordre $1-\alpha$ d'une loi $\chi^2(2)$, comme $\chi^2(2)$ est une loi diffuse et à densité strictement positive en q, on a $q_n \xrightarrow[n\to+\infty]{} q$. On en déduit alors que

$$T_n(\tilde{Y}) = \mathbb{1}_{S_n(\tilde{Y})\geq q_n} \xrightarrow[n\to+\infty]{\mathbb{P}} 1,$$

et donc que T_n est consistant.

Solution 3 -

1. (a) $\varepsilon = (\varepsilon_1, \dots, \varepsilon_{2n})$ est un vecteur gaussien. Y est un transformation linéaire de ε donc Gaussien. Par linéarité de l'espérance :

$$\mathbb{E}[Y] = (\sqrt{2}\mu, \frac{1}{\sqrt{2}}\delta, \dots).$$

Pour le calcul de la matrice de covariance :

$$2\operatorname{Cov}[M_i, P_i] = \sigma^2\rho - \sigma^2 + \sigma^2 - \sigma^2\rho = 0,$$

$$\operatorname{Var}[M_i] = \sigma^2(1-\rho), \quad \operatorname{Var}[P_i] = \sigma^2(1+\rho).$$

Par indépendance de $(\varepsilon_{2i-1}, \varepsilon_{2i})_i$, les autre coefficient de $\operatorname{Cov}[Y]$ sont nuls. D'où

$$\operatorname{Cov}[Y] = \sigma^2\operatorname{Diag}(1+\rho, 1-\rho, 1+\rho, \dots).$$

(b) On a $\Lambda = \sigma\,\mathrm{Diag}(\sqrt{1+\rho}, \sqrt{1-\rho}, \sqrt{1+\rho}, \dots)$ et $X = (X_1, X_2)$ où $X_1 = (\sqrt{2}, 0, \sqrt{2}, \dots)^T$ et $X_2 = (0, 1/\sqrt{2}, 0, 1/\sqrt{2}, \dots)^T$. On n'est pas dans le cadre du cours car les variance des erreurs ne sont pas égales. Le modèle n'est pas homoscédastique.

2. (a) $\widehat{\delta_n} = \sqrt{2}\,\overline{M_n}$.

(b) On estime $\mathrm{Var}[M_1] = (1-\rho)\sigma^2$ par l'estimateur de la variance empirique usuel :

$$\widehat{\mathrm{Var}[M_1]} := \frac{1}{n-1}\sum_{i=1}^{n}(M_i - \overline{M_n})^2.$$

Avec Cochran, on sait que $\widehat{\mathrm{Var}[M_1]} = \mathrm{Var}[M_1]\chi^2(n-1)/(n-1)$ en loi et que $\widehat{\mathrm{Var}[M_1]}$ est indépendant de $\overline{M_n}$. Donc, en loi

$$\sqrt{n}\frac{\overline{M_n} - \delta/\sqrt{2}}{\sqrt{\widehat{\mathrm{Var}[M_1]}}} = \frac{N(0,1)}{\sqrt{\chi^2(n-1)/(n-1)}} = T_{n-1}.$$

Avec T_{n-1} une loi de student à $n-1$ degré de libertés et indépendance entre le numérateur et le dénominateur. Après calculs, on a l'IC :

$$\left[\sqrt{2}\overline{M_n} \pm \sqrt{\frac{2\widehat{\mathrm{Var}[M_1]}}{n}}\,q_{1-\alpha/2}(T_{n-1})\right].$$

3. (a) D'après 1), a), $\mathrm{Cov}[Y]$ est diagonale par bloc. Donc $(M_i)_i$ et $(P_i)_i$ sont indépendantes. De plus, avec Cochran, $S_M^2 \sim (1-\rho)^{-1}(1-\rho)\sigma^2\chi^2(n-1) = \sigma^2\chi^2(n-1)$ et $S_P^2 \sim (1+\rho)^{-1}(1+\rho)\sigma^2\chi^2(n-1) = \sigma^2\chi^2(n-1)$. En sommant deux loi du chi2 indépendantes, on a donc $\widehat{\sigma^2} \sim a_n\sigma^2\chi^2(2n-2)$. Donc $a_n = (2n-2)^{-1}$.

(b) D'après a),

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

Après calculs, on a l'IC

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

4. (a) Avec les mêmes arguments qu'en 3)a) $S_M^2/S_P^2 \sim F(n-1, n-1)$ (loi de Fisher).

(b) Avec proba $1-\alpha$ on a

$$S_M^2/S_P^2 = r(1+\rho)/(1-\rho) \in [q_{\alpha/2}(F(n-1,n-1)), q_{1-\alpha/2}(F(n-1,n-1))].$$

La fonction $f : \rho \in (-1,1) \mapsto (1+\rho)/(1-\rho)$ est strictement croissante et $f^{-1}(y) = (y-1)/(1+y)$. D'où l'IC

$$\left[\frac{q_{\alpha/2}/r - 1}{q_{\alpha/2}/r + 1}, \frac{q_{1-\alpha/2}/r - 1}{q_{1-\alpha/2}/r + 1}\right] = \left[\frac{q_{\alpha/2} - r}{q_{\alpha/2} + r}, \frac{q_{1-\alpha/2} - r}{q_{1-\alpha/2} + r}\right].$$

Solution 3 -

Pour le sens direct, prendre par exemple $a_i = 1/n$. Pour le sens réciproque, on commence par montrer le résultat intermédiaire. Soit $\delta = \max_{i=1,\ldots,n} |a_i| < 1$, et $\mathrm{B}(t)$ la boule fermée de rayon t. Pour tout $t > 0$, notons $g_+(t) = \sup_{\mathrm{B}(t)} |g|$. Comme $\sum_{i=1}^n a_i^2 = 1$ et $|a_i| < \delta$, $g(t) = \sum_{i=1}^n a_i^2 g(a_i t)$ implique que $g_+(t) \leq g_+(\delta t)$. Une récurrence immédiate donne alors $g_+(t) \leq g_+(\delta^n t) \underset{n\to+\infty}{\longrightarrow} |g|(0) = 0$.

On peut maintenant prouver le sens indirect. Soit $(a_1, \ldots, a_n)$ comme dans l'énoncé. L'égalité des variances fournit $\sum_{i=1}^n a_i^2 = 1$. Soit ϕ la fonction caractéristique de X_1. L'égalité en loi fournit, pour tout t,

$$\phi(t) = \prod_{i=1}^n \phi(a_i t).$$

Comme $\mathbb{E}X_1^2 < +\infty$, ϕ est C^2 sur $\mathbb{R}$. Pour pouvoir passer au log (ce qu'on aimerait bien), il faut montrer que $|\phi(t)| > 0$ pour tout t. Pour ce faire, on note $\phi_-(t) = \min_{\mathrm{B}(t)} |\phi(u)|$. Avec les mêmes notations que précédemment on a $\phi_-(t) \geq \phi_-(\delta t)^n \geq \phi_-(\delta^k t)^{nk}$. Or $\phi(0) = 1 > 0$, donc, pour k assez grand $\phi_-(\delta^k t) > 0$, ce dont on déduit $\phi \neq 0$.

On peut donc poser $f = \log(\phi)$ qui est alors aussi C^2, et en dérivant deux fois

$$f''(t) = \sum_{i=1}^n (f(a_i t))'' = \sum_{i=1}^n a_i^2 f''(a_i t).$$

En appliquant le résultat intermédiaire à $f'' - f''(0)$, on a que $f'' \equiv c$, avec $c = f''(0) = -\mathbb{E}(X_1^2) = -1$. On a alors $\phi(t) = e^{-t^2/2}$ ce qui permet de conclure.