Sorbonne Université | Année 2023/2024

M2 Statistique | Premier semestre

Feuille de TD 4

Exercice 1 : Soient X et Y deux variables aléatoires indépendantes telles que $X \sim \mathcal{N}(0,1)$ et Y suit une loi de Rademacher de paramètre $p \in (0,1)$, i.e $\mathbb{P}[Y=1] = p$ et $\mathbb{P}[Y=-1] = 1-p$.

1. Montrer que $XY \sim \mathcal{N}(0,1)$.
2. Montrer que $\operatorname{Cov}(X, XY) = 2p-1$
3. Les variables X et XY sont-elles indépendantes ?
4. Montrer que le vecteur (X, XY) n'est pas gaussien.

Exercice 2 : On considère le modèle

$$Y_i = m + \sigma\epsilon_i, \qquad 1 \leq i \leq n$$

où les v.a. ϵ_i sont i.i.d. de loi commune $\mathcal{N}(0,1)$, pour des paramètres $m \in \mathbb{R}$ et $\sigma > 0$. On note $\overline{Y}_n = n^{-1} \sum_{i=1}^n Y_i$.

1. On suppose que σ est connu.
   (a) Déterminer un intervalle de confiance symétrique pour m de niveau $1-\alpha$.
   (b) Pour $\sigma = 3$, combien d'observations doit-on avoir pour que la longueur de l'intervalle de confiance de niveau 95% soit inférieure à 2 ? Donner la forme de cet intervalle au niveau 95% pour $\sigma = 3$, $n = 25$ et $\bar{y}_{25} = \bar{Y}_{25}(\omega) = 20$. (Indication : $\Phi^{-1}(0.975) \approx 2$.)
   (c) Soit $\alpha \in [0,1]$. Proposer un test de niveau α pour l'hypothèse $H_0 : m = m_0$ contre $H_1 : m \neq m_0$. Pour $\sigma = 3$, $n = 25$, $\bar{y}_{25} = \bar{Y}_{25}(\omega) = 20$ et $m_0 = 18.9$, quelle est la p-valeur de ce test ? Peut-on accepter l'hypothèse H_0 aux niveaux 1%, 5% et 10% ? (Indication : $\Phi\left(\frac{5.5}{3}\right) \simeq \Phi(1.83) \simeq 0.97$.)
2. On ne suppose plus que σ est connu. On pose $\hat{\sigma}_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2}$.
   (a) Écrire le modèle de régression associé aux hypothèses énoncées plus haut et donner l'estimateur des moindres carrés.
   (b) Énoncer le théorème de Cochran dans ce cas.
   (c) Montrer que $\hat{\sigma}_n^2$ est un estimateur sans biais et consistant de σ^2.
   (d) Donner la loi exacte de $\sqrt{n}\frac{\bar{Y}_n - m}{\hat{\sigma}_n}$.

(e) Tester l'hypothèse $H_0 : \sigma^2 = 3$ contre $H_1 : \sigma^2 \neq 3$ au niveau α.

(f) Déterminer un intervalle de confiance de niveau $1-\alpha$ pour m. En déduire un test pour l'hypothèse $H_0 : m = m_0$ contre $H_1 : m \neq m_0$ au niveau α.

(g) Tester maintenant $H_0 : m \geq m_0$ contre $H_1 : m < m_0$ au niveau α. Calculer la p-valeur lorsque $m_0 = 12.5$, $n = 25$, $\bar{y}_{25} = \bar{Y}_{25}(\omega) = 12$ et $\hat{\sigma}_n^2(\omega) = 1.69$? Peut-on accepter l'hypothèse H_0 au niveau 5% ? (Indication : $F_{\mathcal{T}(24)}(-1.92) \simeq 0.03$.)

Exercice 3 : On considère le modèle suivant

$$Y_i = a + bt_i + \varepsilon_i, \quad 1 \leq i \leq n,$$

où les variables aléatoires ε_i sont i.i.d. $\mathcal{N}(0,\sigma^2)$, les réels $(t_i)_{1\leq i\leq n}$ sont connus et a, b, σ^2 sont trois paramètres réels inconnus. On suppose que $\sum_{i=1}^n t_i = 0$ et on note

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i, \quad v_t = \frac{1}{n}\sum_{i=1}^n t_i^2 > 0, \quad v_Y = \frac{1}{n}\sum_{i=1}^n Y_i^2 - \bar{Y}^2 \quad \text{et} \quad \rho = \frac{1}{n}\sum_{i=1}^n Y_i t_i.$$

1. Préciser les conditions d'identifiabilité du modèle.
2. Calculer les estimateurs des moindres carrés $\hat{a}$, $\hat{b}$ et $\hat{\sigma}^2$ de a, b et σ^2 en fonction de $\bar{Y}$, v_t, v_Y et ρ. Quelle est leur loi jointe ?
3. Soit $\alpha \in [0,1]$. Donner un intervalle de confiance de niveau $1-\alpha$ pour chacun des paramètres a et b. En déduire un rectangle de confiance de niveau 95% pour le paramètre (a,b).
4. Montrer que
$$\frac{1}{2\hat{\sigma}^2}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix} - \begin{pmatrix}a\\ b\end{pmatrix}\right)^T \mathbf{X}^T\mathbf{X}\left(\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix} - \begin{pmatrix}a\\ b\end{pmatrix}\right) \sim \mathcal{F}_{2,n-2}.$$
5. En déduire une ellipse de confiance de niveau 95% pour le paramètre (a,b).
6. Donner un intervalle de confiance pour $5a - 8b$, de niveau 95%, lorsque $n = 18$.
7. Tester l'hypothèse H_0 : "$a = b$" contre H_1 : "$a \neq b$" au niveau 1% lorsque $n = 22$.