Sorbonne Université Année 2023/2024

M2 Probabilités et finance, Probabilités et Modèles aléatoires

Premier semestre

TD 3: Théorème de Cochran et modèle linéaire

Exercice: Le principe de la régression linéaire est de modéliser une variable y à partir de variables explicatives $\mathbf{x} = (x_1, \dots, x_p)^T$, i.e de considérer

$$y = \beta_1 x_1 + \ldots + \beta_p x_p,$$

où $\beta = (x_1, \dots, x_p)$ est inconnu. En pratique, on dispose d'un échantillon $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, mais on obtient jamais réellement une droite (erreurs de mesures...). On va donc considérer le modèle linéaire

$$y = \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon$$

avec $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$. On parle alors de modèle linéaire gaussien. On suppose maintenant que les données suivent le modèle suivant :

$$Y_i = \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p} + \epsilon_i,$$

avec

- Y_i est une variable aléatoire et on observe les réalisations y_i .
- Les $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,p})^T$ sont déterministes.
- Le paramètre $\beta = (\beta_1, \dots, \beta_p)^T$ est inconnu et déterministe.
- Les ϵ_i sont i.i.d et $\epsilon_1 \sim \mathcal{N}(0, \sigma^2)$.
- 1. Ecrire le modèle de manière matricielle.
- 2. Donner la loi du vecteur Y. Quelle est la loi de Y_i ?
- 3. On considère à partir de maintenant que rang(X) = p, et on note D = Im(X). On s'intéresse à l'estimateur des moindres carrés défini par

$$\hat{\beta} = \arg\min_{h \in \mathbb{R}^p} \|Y - Xh\|^2$$

Montrer que la matrice X^TX est inversible et en déduire $\hat{\beta}$.

- 4. Montrer que $P_D = X (X^T X)^{-1} X^T$ est le projecteur orthogonale sur D parallèlement à D^{\perp} .
- 5. Que pouvez vous en déduire sur $X\hat{\beta}$?
- 6. Donner la loi de $X\hat{\beta}$ et en déduire celle de $\hat{\beta}$.

- 7. On suppose σ^2 connu. Soit $x_0 \in \mathbb{R}^p \setminus \{0\}$, donner un intervalle de confiance de niveau au moins 1α de $x_0^T \beta$.
- 8. On suppose maintenant que σ^2 est inconnu et on considère l'estimateur

$$\hat{\sigma}^2 = \frac{1}{n-p} \left\| Y - X \hat{\beta} \right\|^2$$

- (a) Expliquer ce choix d'estimateur.
- (b) Exprimer $\hat{\sigma}^2$ à l'aide de projections.
- (c) Enoncer le théorème de Cochran dans ce cas.
- (d) En déduire un intervalle de confiance pour $x_0^T \beta$.
- (e) En déduire un test de niveau α pour tout tester $\beta_j = \beta_{j,0}$.
- (f) Construire un intervalle de confiance pour σ .
- (g) Construire un test di niveau α pour tester $\beta = \beta_0$.
- 9. On considère maintenant une n + 1-ème donnée x_{n+1} et on souhaite prédire Y_{n+1} .
 - (a) Proposer un prédicteur \hat{Y}_{n+1} .
 - (b) Donner la loi de l'erreur de prédiction $\hat{\epsilon}_{n+1} = Y_{n+1} \hat{Y}_{n+1}$.
 - (c) En déduire un intervalle de prédiction.