

TD 4

Exercice 1

Soit Y qui suit un modèle linéaire non-gaussien, et soit $T \in \mathbb{R}^n$ un vecteur déterministe. Montrer que

$$\mathbb{E}(\|T - Y\|^2) = n\sigma^2 + \|T - X\theta\|^2.$$

Exercice 2 Théorème de Gauss-Markov

On se propose de montrer dans cet exercice que, pour le modèle linéaire non gaussien, l'estimateur des moindres carrés $\hat{\theta}$ reste optimal mais maintenant seulement parmi les estimateurs linéaires sans biais.

L'optimalité veut dire que si $\tilde{\theta}$ est un autre estimateur linéaire sans biais :

$$\text{Var}(\tilde{\theta}) - \text{Var}(\hat{\theta}) \text{ est une matrice semi-définie positive,}$$

ou encore, ce qui est équivalent, que pour toute combinaison linéaire $C'\theta$ des paramètres où C est un vecteur de même taille que θ , i.e., $C \in \mathbb{R}^{k+1}$

$$\text{Var}(C'\tilde{\theta}) \geq \text{Var}(C'\hat{\theta}).$$

- Posons $\tilde{\theta} = MY$ où M est une matrice de taille $(k+1, n)$. Montrer que $MX = I$.
- Écrire $\hat{\theta} = TP_{[X]}Y$, et montrer que $MP_{[X]} = TP_{[X]}$.
- Montrer que $\tilde{\theta} = \hat{\theta} + MP_{[X]^\perp}Y$, les deux termes de la somme étant non-corrélés. Conclure.

Exercice 3

Soit θ_1 et θ_2 deux paramètres réels inconnus et soit :

Y_1 un estimateur sans biais de $\theta_1 + \theta_2$ et de variance σ^2 ;

Y_2 un estimateur sans biais de $2\theta_1 - \theta_2$ et de variance $4\sigma^2$;

Y_3 un estimateur sans biais de $6\theta_1 + 3\theta_2$ et de variance $9\sigma^2$,

les estimateurs Y_1 , Y_2 et Y_3 étant indépendants. Quels estimateurs de θ_1 et θ_2 proposeriez vous? (on pourra utiliser l'exercice précédent).

Exercice 4 Modèle passant par l'origine

Considérons le modèle gaussien

$$Y_i = \beta X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

où $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, i.i.d. et X_i sont déterministes. Rappelons que la vraisemblance est la densité des observations Y_1, \dots, Y_n vue comme une fonction des paramètres β et σ^2 . La densité d'une variable aléatoire gaussienne $\mathcal{N}(\mu, \sigma^2)$ est

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

1. Montrer que $-2 \times \log$ -vraisemblance vaut

$$L(\beta, z) = n \log(2\pi) + n \log z + \frac{1}{z} \sum_{i=1}^n (Y_i - \beta X_i)^2,$$

en notant $z = \sigma^2$ pour dériver plus facilement.

2. Montrer que les estimateurs du maximum de vraisemblance de β et σ^2 sont

$$\hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \beta X_i)^2.$$

3. Les estimateurs $\hat{\beta}$ et $\hat{\sigma}^2$ sont ils sans biais? Calculez les variances de $\hat{\beta}$ et $\hat{\sigma}^2$. (Rappel : $\mathbb{E}(\chi^2(n)) = n$, $\text{Var}(\chi^2(n)) = 2n$.)

4. On dispose de deux jeux de données suivants.

X		8	11	5	6	4	X		47	43	58	45	36
Y		17	21	10	11	8	Y		95	87	118	89	74

Lequel choisissez vous pour estimer β ? Au moment du recueil de données, que devons nous faire pour diminuer les variances de $\hat{\beta}$ et $\hat{\sigma}^2$?