

TD 5 : vecteurs gaussiens et modèle linéaire

Les questions marquées d'un astérisque (*) sont facultatives.

Exercice 1. (matrices définies positives)

- Rappeler la définition d'une matrice symétrique définie positive. D'une matrice symétrique semi-définie positive.
- Les matrices suivantes sont-elles symétriques définies positives ? Symétriques semi-définies positives ?

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & -2 \\ -2 & 2 \end{pmatrix}$$

Exercice 2. Corrigez tout ce qui est faux dans le raisonnement suivant :

« Soient X et Y deux v.a. gaussiennes d'espérance 1 telles que $\text{Var}(X) = 1$, $\text{Var}(Y) = 2$ et $\text{Cov}(X, Y) = -2$. Alors (X, Y) est un vecteur gaussien d'espérance $(1, 1)$ et de matrice de covariance

$$\begin{pmatrix} 2 & -2 \\ -2 & 1 \end{pmatrix}$$

donc le vecteur $(X + Y, X - Y)$ est un vecteur gaussien d'espérance $(2, 0)$ et de matrice de covariance

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & -2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 4 & -3 \end{pmatrix} \gg .$$

(*) De telles variables existent-elles ?

Exercice 3. (théorème de Cochran et applications)

- Énoncer le théorème de Cochran.
- Soit $n \geq 2$ un entier et $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur de v.a. i.i.d. de loi $\mathcal{N}(0, \sigma^2)$. Soit F le sous-espace vectoriel de $\mathbb{R}^n$ engendré par le vecteur $(1, 1, \dots, 1)$.

- $\mathbf{X}$ est-il un vecteur gaussien ?
- Calculer la projection orthogonale de $\mathbf{X}$ sur F .
- Calculer la projection orthogonale de $\mathbf{X}$ sur $F^\perp$.

Notons $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ et $\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

- Quelle est la loi de $(n-1)\hat{\sigma}_n^2$?
- Montrer que $\bar{X}_n$ et $\hat{\sigma}_n^2$ sont indépendants.

Soit $d \geq 1$ un entier, $Z \sim \mathcal{N}(0, 1)$ et $U \sim \chi^2(d)$ avec Z et U indépendantes. La loi de Student à d degrés de liberté $\mathcal{T}(d)$ est définie comme la loi de $\frac{Z}{\sqrt{U/d}}$.

- Montrer que $\frac{\bar{X}_n}{\hat{\sigma}_n/\sqrt{n}}$ suit la loi de Student à $n-1$ degrés de liberté.

- Suite à une panne, un laboratoire de biologie a remplacé un appareil de mesure. Il aimerait s'assurer que l'incertitude de mesure du nouvel appareil est la même que celle du précédent. L'écart-type de mesure du précédent appareil était de 0,7. Sur 100 mesures avec le nouvel appareil, on estime un écart-type de 1,12. Le laboratoire a-t-il des raisons de penser que les écart-types des deux appareils sont différents ?

ordre du quantile	0,025	0,05	0,95	0,975
$\chi^2(99)$	73,4	77,0	123,2	128,4
$\chi^2(100)$	74,2	77,9	124,3	129,6
$\chi^2(101)$	75,1	78,8	125,5	130,7

FIGURE 1 – Quelques quantiles de lois du χ^2

Exercice 4. (Modèle linéaire) (Rappel) Un *modèle linéaire* (multivarié) est un modèle statistique dans lequel un vecteur d'observations $\mathbf{Y} = (Y_i)_{1 \leq i \leq d}$ s'exprime en fonction de variables explicatives $\mathbf{X} = (X_{ij})_{1 \leq i \leq d, 1 \leq j \leq p}$ via la relation

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon,$$

où ε est un vecteur aléatoire de dimension d (le bruit) et β est un vecteur de taille p . La matrice $\mathbf{X}$ et le vecteur $\mathbf{Y}$ sont supposés observés. L'enjeu est généralement d'estimer β . Si $d = 1$, ce modèle s'écrit plus simplement

$$Y = \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon.$$

Dans les questions qui suivent, on suppose qu'on observe une variable aléatoire Y et des variables explicatives (X_1, X_2) . Lesquels des modèles suivants pouvez-vous réécrire sous forme de modèle linéaire? Précisez la transformation des X et Y effectuée.

1. $Y = w_1 X_1 + w_2 X_1^2 + w_3 X_1 X_2^2 + \varepsilon$,
2. $Y = \sqrt{w_1 + w_2 X_1 + w_3 X_2} + \varepsilon$,
3. $Y = X_1^{w_1} X_2^{w_2} 4^{w_3} \varepsilon$,
4. $\log(Y) = w_1 X_1 + w_2 \log(X_2) + \varepsilon$,
5. On observe une suite $Y^{(1)}, \dots, Y^{(t)}$ et $X^{(1)}, \dots, X^{(t)}$ dépendant de l'instant t de la manière suivante : pour tout $s \in \{0, \dots, t\}$, $Y^{(s)} = w_1 \cos(\pi s/6) + w_2 s + w_3 X^{(s)} + \varepsilon^{(s)}$.