TD 6 : Vecteurs gaussiens

Exercice 1. Soit $r \geq 1$ et $n_1, \ldots, n_r$ des entiers strictement positifs. Soient $Y_1, \ldots, Y_r$ des vecteurs gaussiens indépendants à valeurs respectivement dans $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_r}$. Montrer que le vecteur aléatoire $Y = {}^t(Y_1, \ldots, Y_r)$, de dimension $n_1 + \ldots + n_r$, est gaussien et expliciter sa matrice de covariance en fonction de celles des Y_i, $1 \leq i \leq r$.

Exercice 2. Soit X un vecteur gaussien centré de matrice de covariance

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

Trouver un endomorphisme U de $\mathbb{R}^3$ tel que les composantes de $Y = UX$ soient indépendantes. Préciser la loi de Y.

Exercice 3. Soient $U_1, \ldots, U_n$ des v.a. réelles i.i.d. de loi $\mathcal{N}(0,1)$ et $\theta \in \mathbb{R}$. On pose $X_1 = U_1$ et pour tout $2 \leq k \leq n$, $X_k = \theta X_{k-1} + U_k$.

a) Montrer que le vecteur aléatoire $X = {}^t(X_1, \ldots, X_n)$ peut s'exprimer sous la forme $X = AU$, pour une matrice A à déterminer. Montrer que X est un vecteur gaussien.

b) Déduire de l'expression précédente l'espérance $\vec{m}$ et la matrice de covariance Σ de X.

c) Montrer que X admet une densité, que l'on calculera en fonction de Σ.

d) Quelle est la loi de X_n ? Discuter en fonction de θ le comportement de X_n lorsque $n \to \infty$.

Exercice 4. On souhaite étudier le rendement d'une certaine variété de pommes. On dispose de p pommiers et on note X_j le rendement observé du $j$ème arbre, $j \in \{1, \cdots, p\}$. On suppose que chaque X_j est la somme d'une moyenne théorique m ne dépendant que de la variété, et d'une erreur ϵ_j :

$$X_j = m + \epsilon_j, \qquad \epsilon_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \qquad j \in \{1, \ldots, p\}.$$

On ne connaît ni m ni σ^2, on considère donc le modèle paramétrique $\{\mathbb{P}_{m,\sigma},\ m > 0,\ \sigma > 0\}$.

a) Proposer un estimateur pour m et donner ses propriétés (biais, consistance, vitesse).

b) On cherche à présent à estimer σ^2. Pour ce faire, on pose

$$S_p = \frac{1}{p} \sum_{j=1}^{p} X_j^2 - \left(\bar{X}_p\right)^2 = \frac{1}{p} \sum_{j=1}^{p} \left(X_j - \bar{X}_p\right)^2.$$

Pour tout $j \in \{1, \dots, p\}$, on définit

$$Z_j = \begin{pmatrix} X_j^2 \\ X_j \end{pmatrix}.$$

Donner l'espérance et la matrice de covariance de Z_1 sous $\mathbb{P}_{m,\sigma}$. En déduire le comportement asymptotique du vecteur aléatoire $\sqrt{p}\big(\overline{Z}_p - \mathbb{E}_{m,\sigma}[Z_1]\big)$, lorsque $p \to \infty$.

c) En utilisant la méthode delta avec une fonction adéquate, montrer que

$$\sqrt{p}\big(S_p - \sigma^2\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, V(\sigma)\big) \qquad \text{lorsque } p \to \infty,$$

pour une fonction $V(\sigma)$ indépendante de m à déterminer.

Exercice 5. A l'aide de la transformée de Fourier et du résultat en dimension 1, démontrer le théorème central limite vectoriel.

Exercice 6. Soit Y un vecteur de $\mathbb{R}^n$. On dit que Y suit un modèle linéaire Gaussien lorsque

$$Y = \mu + \varepsilon \quad \text{où } \varepsilon \text{ est un vecteur Gaussien de loi } \mathcal{N}_n(0, \sigma^2\, I_n)$$

pour deux paramètres $\mu \in \mathbb{R}^n$ et $\sigma^2 > 0$ inconnus. Nous faisons de plus l'hypothèse que μ appartient à un sous espace vectoriel V de $\mathbb{R}^n$.

Remarque : Il s'agit de la donnée du 1-échantillon Y de loi $\mathbb{P}_{(\mu,\sigma^2)} = \mathcal{N}_n(\mu, \sigma^2\, I_n)$ avec $(\mu, \sigma^2) \in V \times]0, +\infty[$. Il s'agit aussi de la donnée de $Y = {}^t(Y_1, \dots, Y_n)$, où Y_i suit la loi $\mathcal{N}(\mu_i, \sigma^2)$ pour tout $i \in \{1, \dots, n\}$, mais attention ce n'est pas a priori un n-échantillon : ici les Y_i sont indépendantes mais pas de même loi.

1) Quelle est la densité du vecteur Gaussien Y de loi $\mathbb{P}_{(\mu,\sigma^2)} = \mathcal{N}_n(\mu, \sigma^2\, I_n)$ avec $(\mu, \sigma^2) \in V \times]0, +\infty[$?

2)a) Notons Π_V la projection orthogonale sur V. Montrer que

$$\forall x \in \mathbb{R}^n \qquad \inf_{m \in V} \|x - m\| = \|x - \Pi_V x\|$$

b) En déduire que l'estimateur du maximum de vraisemblance de (μ, σ^2) est $(\widehat{\mu}, \widehat{s}_n^2)$ avec

$$\widehat{\mu} = \Pi_V Y \qquad \text{et} \qquad \widehat{s}_n^2 = \frac{1}{n}\|Y - \Pi_V Y\|^2$$

3) Notons $p = \dim V$. Montrer que $\widehat{\mu}$ et $\widehat{s}_n^2$ sont indépendants et que

$$\widehat{\mu} \sim \mathcal{N}_n(\mu, \sigma^2 \Pi_V) \qquad \text{et} \qquad \frac{n\,\widehat{s}_n^2}{\sigma^2} \sim \chi^2(n-p)$$

4) Posons

$$\widehat{\sigma}_n^2 = \frac{1}{n-p}\|Y - \Pi_V Y\|^2 = \frac{n\,\widehat{s}_n^2}{n-p}$$

Soit $x \in \mathbb{R}^n$ tel que $\langle \Pi_V x, x \rangle = {}^t\!x \Pi_V x \neq 0$. Donner la loi de la variable aléatoire

$$\frac{{}^t\!x\widehat{\mu} - {}^t\!x\mu}{\sqrt{\widehat{\sigma}_n^2 \; {}^t\!x\Pi_V x}}$$

En déduire un test de H_0 : " ${}^t\!x\mu = 0$" contre H_1 : " ${}^t\!x\mu \neq 0$" au niveau α.

A quoi ça sert tout ça ? Supposons qu'une certaine quantité y est une fonction affine des quantités $x_1, \ldots, x_p$:

$$y = \beta_1 x_1 + \cdots + \beta_p x_p \qquad \textit{avec } \beta_1, \ldots, \beta_p \textit{ inconnus}$$

Observons alors n réalisations indépendantes de $(y, x_1, \ldots, x_p)$:

$$(Y_1, X_{1,1}, \ldots, X_{1,p}), \qquad \ldots \qquad (Y_n, X_{n,1}, \ldots, X_{n,p})$$

Nous pouvons alors supposer que $Y = \mu + \varepsilon$ avec $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$ et $\mu_j = \beta_1 X_{j,1} + \cdots + \beta_p X_{j,p}$ pour tout $j \in \{1, \ldots, n\}$. Le variable aléatoire ε représentre le bruit, c'est-à-dire l'impact de toutes les variables aléatoires d'environnement non explicitement prises en compte dans la relation linaire. Nous pouvons donc estimer μ, déterminer un intervalles de confiance, faire des tests... Et, sous des hypothèses d'inversibilité, nous en déduisons l'estimation de $(\beta_1, \ldots, \beta_p)$, des intervalles de confiance, des tests...