TD2 : Modèle linéaire gaussien

Exercice 1 (Quelques éléments d'algèbre linéaire)**.**

1. Soit M une matrice carrée $n \times n$ telle que ${}^tM = M = M^2$. Soit $W = \mathrm{Im}(M)$.
   (a) Soit $u \in W$, montrer que $Mu = u$.
   (b) Soit $v \in \mathbb{R}^n$, et $w \in W$, montrer que $v - Mv \perp w$.
   (c) Conclure que M est la matrice de projection sur $\mathrm{Im}(M)$.
2. Calculer la projection du vecteur $u = (1, 1, 0)$ sur le plan $\{x + y - z = 0\}$.
3. Déterminer la matrice de projection orthogonale Q de $\mathbb{R}^4$ sur le sous-espace vectoriel engendré par $(1, 2, 0, 0)$ et $(1, 0, 1, 1)$. Quelle est la projection de $(1, -2, 2, 2)$ sur cet espace ? Qu'en déduisez-vous ?

Exercice 2. On souhaite vérifier si le taux de sucre d'un soda respecte les recommandations d'un organisme de santé indépendant, à savoir une valeur inférieure à 0.100 g/ml. Dans des conditions expérimentales garantissant l'indépendance des observations, on mesure le taux de sucre de 10 bouteilles différentes. Sur ces observations, on observe un taux moyen de 0.103 g/ml de sucre et un écart-type de 0.010 g/ml. On suppose que les 10 observations suivent une même loi gaussienne.

1. Préciser le modèle statistique.
2. Construire un test en se plaçant du point de vue du directeur de l'usine de fabrication du soda. Quelle est sa conclusion au niveau 5% ?
3. Construire le test en se plaçant du point de vue de l'organisme de santé, quelle sera sa conclusion au niveau 5% ?

Exercice 3. Soit le modèle statistique

$$\left(\mathbb{R}^{n+p}, \{\mathcal{N}(m_1, \sigma^2)^{\otimes n} \otimes \mathcal{N}(m_2, \sigma^2)^{\otimes p}\}_{m_1, m_2 \in \mathbb{R}, \sigma^2 > 0}\right).$$

L'objectif de cet exercice est de construire un test pour $H_0 : m_1 = m_2$ contre $H_1 : m_1 \neq m_2$. Dans la suite, pour $m_1, m_2 \in \mathbb{R}$ et $\sigma^2 > 0$, on note $X = (X_1, \ldots X_2)$ de loi $\mathcal{N}(m_1, \sigma^2)$ et $Y = (Y_1, \ldots Y_p)$ de loi $\mathcal{N}(m_2, \sigma^2)$.

1. Déterminer les lois des statistiques

$$Q = \frac{\bar{X}_n - \bar{Y}_p}{\sqrt{1/n + 1/p}} \quad \text{et} \quad W^2 = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 + \sum_{j=1}^{p} (Y_j - \bar{Y}_p)^2.$$

2. En déduire la loi sous H_0 de la statistique $\sqrt{n+p-2}\frac{Q}{W}$.
3. Constuire un test de niveau $\alpha \in (0, 1)$ pour $_0$ contre H_1.
4. Les résultats d'un examen noté sur 20 sont répertoriés dans le tableau suivant.

Filles	2	12	4	4	2	15	10	17	11	12	2		
Garçons	7	3	6	8	9	10	9	11	17	16	14	12	2	6

On suppose que les notes sont indépendantes et issues de lois gaussiennes dont seule la moyenne dépend du genre de l'élève, la variance étant identique. Montrer qu'au niveau 5%, on ne peut déduire à l'existence d'une différence entre les moyennes des notes des filles et des garçons.

Exercice 4. Soit $(t_1, \dots t_n)$ des réels tels que $t_i \neq t_j$ pour deux indices $i \neq j$. On considère le modèle statistique

$$\left(\mathbb{R}^n, \{\mathcal{N}(\theta_1 + t_i\theta_2, \sigma^2)^{\otimes n} \otimes \mathcal{N}(m_2, \sigma^2)^{\otimes p}\}_{\theta_1,\theta_2 \in \mathbb{R}, \sigma^2>0}\right),$$

et on note $\mathbf{1}$ et $\mathbf{t}$ les vecteurs définis par ${}^t(1, \dots, 1)$ et ${}^t(t_1, \dots t_n)$ respectivement. On posera

$$\bar{\mathbf{t}} = \frac{1}{n}\sum t_i \quad \text{et} \quad s = \frac{1}{n}\|\mathbf{t}\|^2 - (\bar{\mathbf{t}})^2.$$

1. Montrer que $s \neq 0$.
2. Déterminer les estimateurs du maximum de vraisemblance de de $(\theta_1, \theta_2, \sigma^2)$.
3. Quelle est la loi de $(\hat{\theta}_1, \hat{\theta}_2)$? Montrer que ces deux vecteurs sont linéairement indépendants si et seulement si $\bar{\mathbf{t}} = 0$.
4. Construire un test au niveau α pour $H_0 : \theta_2 = 0$ contre $H_1 : \theta_2 \neq 0$.
5. Des prélèvements ont lieu dans la rivière aux jours $t_1 = 1, \dots, t_5 = 5$ et ont donné les valeurs suivantes pour la concentration d'un polluant

Jour	1	2	3	4	5
[C] (μ/l)	11	10	12	12	14

Peut-on rejeter au niveau 5% l'hypothèse que la concentration en polluant n'évolue pas au cours du temps ?

Exercice 5 (Régression polynomiale)**.** On s'intéresse au modèle statistique $Y = P(X) + \epsilon$, où P est un polynôme de degré d et ϵ un vecteur gaussien centré de matrice de covariance $\sigma^2 I_n$.

1. Poser ce modèle statistique sous la forme d'un modèle linéaire gaussien sous forme matricielle. Rappeler la formule donnant les estimateurs du maximum de vraisemblance de $a_d, \dots, a_0$ de P ainsi que σ^2.
2. Donner un estimateur non-biaisé de σ^2 noté $\hat{\sigma^2}_n$, ainsi que sa loi.
3. Déterminer un intervalle de confiance à 95% de σ^2. Commentez.
4. Que se passe-t-il dans le cas particulier $n = d$?
5. On note $H_0^{(k)} := P \in \mathbb{R}_{k-1}[X]$ et $H_1^{(k)} := P \in \mathbb{R}_k[X]$. Déterminer un test de niveau α de $H_0^{(k)}$ contre $H_1^{(k)}$.