Université François-Rabelais de Tours **M1 Mathématiques**
2018-2019 **Probabilités-2**

Controle continu 2

Exercice 1 Soient $n \geq 1$ et les réels $x_1, \ldots x_n$. Soient $\sigma^2 > 0$ et Z un vecteur gaussien de dimension d, de loi $\mathcal{N}(0, \sigma^2 I_d)$. Soient enfin $a, b \in \mathbb{R}$ et Y le vecteur gaussien de coordonnées $Y_i = ax_i + b + \sigma Z_i$.

1) Trouver les variables aléatoires $\hat{A}, \hat{B} \in \mathbb{R}$ minimisant pour tout $\omega \in \Omega$ la quantité

$$\sum_{i=1}^{n} (Y_i - \hat{A}x_i - \hat{B})^2.$$

2) On pose $\hat{Y}_i = \hat{A}x_i + \hat{B}$ pour $1 \leq i \leq n$. Démontrer que $\sum_{i=1}^{n}(Y - \bar{Y})^2 = \sum_{i=1}^{n}(Y - \hat{Y})^2 + \sum_{i=1}^{n}(\hat{Y} - \bar{Y})^2$.
3) Vérifier que $\sum_{i=1}^{n}(Y - \hat{Y})^2 = \sum_{i=1}^{n}(Y - \bar{Y})^2 - \hat{A}^2 \sum_{i=1}^{n}(\hat{x} - \bar{x})^2$.
4) En déduire la formule $SCE = n\sigma_Y^2(1 - \rho_{x,Y}^2)$, où SCE désigne la somme des carrés des écarts au modèle, σ_Y^2 la variance observée de $(Y_i)_i$ et $\rho_{x,Y}$ le coefficient de corrélation.

Exercice 2 Soient $n \geq 1$ et les réels $x_1, \ldots x_n$. Pour $n+1 \leq i \leq 2n$ on pose $x_i = x_{i-n}$. On considère les vecteurs $\vec{e}_1, \vec{e}_2 \in \mathbb{R}^{2n}$ définis par $\vec{e}_j = (e_{i,j})$ où

$$e_{i,1} = \begin{cases} x_i & \text{si} \quad 1 \leq i \leq n \\ -x_{i-n} & \text{si} \quad n+1 \leq i \leq n. \end{cases}$$

$$e_{i,2} = \begin{cases} x_i & \text{si} \quad 1 \leq i \leq n_1 \\ x_{i-n} & \text{si} \quad n+1 \leq i \leq 2n. \end{cases}$$

1) Étude théorique
   a) Soit B la matrice de vecteurs colonnes $(\vec{e}_1, \vec{e}_2)$. Calculer la matrice B^*B et la matrice $(B^*B)^{-1}B^*$.
   b) Soient $\vec{\lambda} = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$, et $\vec{Y}$ un vecteur gaussien de loi $\mathcal{N}(B\vec{\lambda}, \sigma^2 I_n)$. On note $(X_i)_{1 \leq i \leq n}$ ses coordonnées. Soit $\hat{\Lambda} = \begin{pmatrix} \hat{A} \\ \hat{B} \end{pmatrix}$ défini par $\hat{\Lambda} = (B^*B)^{-1}B^*X$. Donner les expressions de $\hat{A}$ et $\hat{B}$ en fonction des variables aléatoires $C_1 = \frac{1}{n}\sum_{i=1}^{n} x_i Y_i$ et $C_2 = \frac{1}{n}\sum_{i=n+1}^{2n} x_i Y_i$.
   c) Donner les lois de $\hat{A}$ et $\hat{B}$.
   d) Les variables aléatoires $\hat{A}$ et $\hat{B}$ sont-elles indépendantes ?
   e) Donner l'expression de $\hat{Y} = B\hat{\Lambda}$ en fonction des variables aléatoires C_1 et C_2.
   f) Quelle est la loi de $\hat{Y}$?
   g) Quelle est la loi de $\vec{Y} - \hat{Y}$? Préciser le rang de sa matrice de covariance.
   h) Les vecteurs aléatoires $\vec{Y} - \hat{Y}$ et $\hat{Y}$ sont-ils indépendants ?
   i) On note $(\hat{Y}_i)_{1 \leq i \leq n}$ les coordonnées du vecteur aléatoire $\hat{Y}$. Les vecteurs aléatoires $((Y_i - \hat{Y}_i)^2)_{1 \leq i \leq 2n}$ et $(\hat{Y}_i^2)_{1 \leq i \leq 2n}$ sont-ils des vecteurs gaussiens ? sont-ils indépendants ?
   j) Quelle est la loi de $SCE = \|\vec{Y} - \hat{Y}\|^2$?

k) Les variables aléatoires $\|\vec{Y} - \hat{Y}\|^2$ et $\hat{A}$ sont-elles indépendantes ?

On note I l'intervalle de confiance de niveau $\alpha \in]0,1[$ sur le paramètre a.

l) Quelle est la statistique T, de loi de Student, permettant le calcul de I.

m) Donner l'expression des bornes de I.

On note J_i l'intervalle de confiance de niveau $\alpha \in]0,1[$ sur $E(Y_i)$, $i = 1, \ldots, 2n$.

n) Quelle est la statistique T_i, de loi de Student, permettant le calcul de J_i.

o) Donner l'expression des bornes de J_i.

2) Application numérique :

Afin de comparer deux engrais $\mathcal{E}_1$ et $\mathcal{E}_2$ on relève la taille moyenne y d'une plante à 1 mois, pour différents dosages x d'engrais. Les observations sont données dans le tableau suivant

Plante i	1	2	3	4	5	6
Engrais	$\mathcal{E}_1$	$\mathcal{E}_1$	$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_2$	$\mathcal{E}_2$
Dose x_i	1	2	3	1	2	3
Taille y_i	$0,9$	$1,5$	$2,9$	1	$0,6$	$1,6$

On donne

$$\sum_{i=1}^{6} x_i^2 = 28, \ \sum_{i=1}^{6} y_i^2 = 15,39, \ \sum_{i=1}^{3} x_i y_i = 12,6, \ \sum_{i=4}^{6} x_i y_i = 7$$

a) Calculer la valeur $\hat{a}$ prise par l'estimateur $\hat{A}$ de a.

b) Calculer la valeur $\hat{b}$ prise par l'estimateur $\hat{B}$ de b.

c) En déduire les estimations $\hat{y}_i$ de $E(Y_i)$, $i = 1, \ldots, 6$.

d) Calculer le résidus SCE du modèle.

e) Donner son nombre de degrés de liberté.

On note $I = [c;\ d]$ l'intervalle de confiance au niveau $1 - \alpha = 0,95$ pour le coefficient a.

f) Déterminer le fractile de la loi de Student intervenant dans la calcul de I.

g) Déterminer c.

h) Déterminer d.

i) D'après cette étude, peut-on penser qu'un engrais favorise plus la croissance que l'autre (expliciter le critère pris utilisé).

Fractiles $t_n(p)$ de la loi de STUDENT à n degrés de liberté.

	0,6	0,7	0,8	0,9	0,95	0,975	0,99	0,995	0,999	0,9995
n= 2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,327	31,599
n= 3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,215	12,924
n= 4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
n= 5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	4,032	5,893	6,869
n= 6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
n= 7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,408
n= 8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
n= 9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
n= 10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587