Sujet S 11 - Exercice

1) Question de cours : Dans une série statistique associée à un échantillon, définition de la moyenne empirique et de la variance empirique.

On observe conjointement deux caractères quantitatifs X et Y sur un échantillon de taille n ($n \geqslant 2$) d'une population.
Ces observations sont donc constituées par un n-uplet $\big((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\big)$ d'éléments de $\mathbb{R}^2$. A chaque individu k de $[\![1, n]\!]$, on associe le couple d'observations (x_k, y_k) de $\mathbb{R}^2$. On suppose que, pour tout $(i, j) \in [\![1, n]\!]^2$, avec $i \neq j$, on a : $x_i \neq x_j$ et $y_i \neq y_j$.

2) Soit f la fonction définie sur $\mathbb{R}^2$ à valeurs réelles qui, à tout couple (a, b) de $\mathbb{R}^2$, associe :

$$f(a, b) = \sum_{k=1}^{n} \big(y_k - (ax_k + b)\big)^2.$$

a) Justifier que f est de classe C^2 sur $\mathbb{R}^2$.

b) Ecrire le système d'équations (S) permettant de déterminer les points critiques de f.

c) On pose : $\overline{x} = \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} x_k$. Etablir la formule : $\displaystyle\sum_{k=1}^{n} \left(x_k - \overline{x}\right)^2 = \left(\sum_{k=1}^{n} x_k^2\right) - n\overline{x}^2$.

En déduire que : $\left(\displaystyle\sum_{k=1}^{n} x_k^2\right) - n\overline{x}^2 > 0$.

d) Résoudre le système (S). En déduire que f admet un unique point critique, noté $(\hat{a}, \hat{b})$.

e) Montrer que f admet un minimum local en $(\hat{a}, \hat{b})$. Est-ce un minimum global pour f ?

3) On note $r_{x,y}$ le coefficient de corrélation linéaire de X et Y. Montrer que : $\left|r_{x,y}\right| \leqslant 1$.

Sujet S 11 - Exercice sans préparation

Soit p un entier supérieur ou égal à 2 et A une matrice de $\mathscr{M}_p(\mathbb{R})$ telle que :
$2A^4 - 2A^3 + I = 0$ (I désigne la matrice identité d'ordre p et 0 la matrice nulle carrée d'ordre p).

1) Déterminer l'ensemble des entiers $n \in \mathbb{Z}$ pour lesquels la matrice $A + \dfrac{n}{n^2+1} I$ est inversible.

2) Existe-t-il un entier naturel n tel que $((n^2+1)A^2 + nA)^n = B$ où B est la matrice carrée d'ordre p définie par $B = \begin{pmatrix} n & n & \ldots & n \\ \vdots & \vdots & \ldots & \vdots \\ n & n & \ldots & n \end{pmatrix}$?

R.

HEC 2010 S11 Correction de l'exercice

Q1 Soit $(x_i, n_i)_{i \in [\![1,p]\!]}$ une série statistique associée à un échantillon.

La moyenne empirique est $\bar{x} = \frac{1}{n} \sum_{i=1}^{p} n_i x_i$ où $n = \sum_{i=1}^{p} n_i$.

La variance empirique est $v_x = \frac{1}{n} \sum_{i=1}^{p} n_i (x_i - \bar{x})^2$ où $v_x = \frac{1}{n} \sum_{i=1}^{p} n_i x_i^2 - \bar{x}^2$

Q2 Notons que l'on redémontre un résultat de cours.

a) f est une fonction polynôme donc f est de classe C^2 sur $\mathbb{R}^2$.

b) Soit $(a,b) \in \mathbb{R}^2$. $\frac{\partial f}{\partial x}(a,b) = \sum_{k=1}^{n} 2(-x_k)(y_k - (a x_k + b))$ et $\frac{\partial f}{\partial y}(a,b) = \sum_{k=1}^{n} 2(-1)(y_k - (a x_k + b))$.

$$\nabla f(a,b) = 0_{\mathbb{R}^2} \Leftrightarrow \frac{\partial f}{\partial x}(a,b) = \frac{\partial f}{\partial y}(a,b) = 0 \Leftrightarrow \begin{cases} \sum_{k=1}^{n} 2(-x_k)(y_k - (a x_k + b)) = 0 \\ \sum_{k=1}^{n} 2(-1)(y_k - (a x_k + b)) = 0 \end{cases}$$

$$\nabla f(a,b) = 0_{\mathbb{R}^2} \Leftrightarrow \begin{cases} 0 = \sum_{k=1}^{n} x_k (y_k - a x_k - b) = \sum_{k=1}^{n} x_k y_k - a \sum_{k=1}^{n} x_k^2 - b \sum_{k=1}^{n} x_k \\ 0 = \sum_{k=1}^{n} (y_k - a x_k - b) = \sum_{k=1}^{n} y_k - a \sum_{k=1}^{n} x_k - nb \end{cases}$$

$$\nabla f(a,b) = 0_{\mathbb{R}^2} \Leftrightarrow \begin{cases} \left(\sum_{k=1}^{n} x_k^2\right) a + \left(\sum_{k=1}^{n} x_k\right) b = \sum_{k=1}^{n} x_k y_k \\ \left(\sum_{k=1}^{n} x_k\right) a + nb = \sum_{k=1}^{n} y_k \end{cases}$$

<u>Le système (S) permettant de déterminer les points critiques de f est</u>

$$\begin{cases} \left(\sum_{k=1}^{n} x_k^2\right) a + \left(\sum_{k=1}^{n} x_k\right) b = \sum_{k=1}^{n} x_k y_k \\ \left(\sum_{k=1}^{n} x_k\right) a + nb = \sum_{k=1}^{n} y_k \end{cases}$$

c) $\sum_{k=1}^{n}(x_k-\bar{x})^2 = \sum_{k=1}^{n}(x_k^2-2x_k\bar{x}+\bar{x}^2) = \sum_{k=1}^{n}x_k^2 - 2\bar{x}\underbrace{\sum_{k=1}^{n}x_k}_{n\bar{x}} + n\bar{x}^2 = \sum_{k=1}^{n}x_k^2 - 2\bar{x}(n\bar{x}) + n\bar{x}^2$

Alors $\sum_{k=1}^{n}(x_k-\bar{x})^2 = \sum_{k=1}^{n}x_k^2 - n\bar{x}^2$.

$\forall k \in [\![1,n]\!], (x_k-\bar{x})^2 \geqslant 0$ donc $\sum_{k=1}^{n}(x_k-\bar{x})^2 \geqslant 0$.

Supposons que $\sum_{k=1}^{n}(x_k-\bar{x})^2 = 0$ alors $\forall k \in [\![1,n]\!]$, $(x_k-\bar{x})^2=0$ (car pour tout $k \in [\![1,n]\!]$, $(x_k-\bar{x})^2 \geqslant 0$). Donc $\forall k \in [\![1,n]\!]$, $x_k - \bar{x} = 0$.

$\forall k \in [\![1,n]\!]$, $x_k = \bar{x}$. Alors $x_1 = x_2 = \cdots = x_n$, ce qui est contraire à l'hypothèse.

Donc $\sum_{k=1}^{n}(x_k-\bar{x})^2 \geqslant 0$ et $\sum_{k=1}^{n}(x_k-\bar{x})^2 \neq 0$. Ainsi $\sum_{k=1}^{n}(x_k-\bar{x})^2 > 0$

donc $\sum_{k=1}^{n}x_k^2 - n\bar{x}^2 > 0$.

> Notons que l'hypothèse $\forall (i,j) \in [\![1,n]\!]^2, i \neq j \Rightarrow x_i \neq x_j$ est trop forte. On peut se contenter de $x_1, x_2, \dots, x_n$ non tous égaux.

d) Posons, comme "d'habitude", $\bar{y} = \frac{1}{n}\sum_{k=1}^{n}y_k$, $\sigma_{x,y} = \frac{1}{n}\sum_{k=1}^{n}x_ky_k - \bar{x}\bar{y}$ et

$\sigma_x^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k-\bar{x})^2 = \frac{1}{n}\sum_{k=1}^{n}x_k^2 - \bar{x}^2$. Notons que $\sigma_x^2 > 0$ d'après ce qui précède.

Soit $(a,b) \in \mathbb{R}^2$.

$$\nabla f(a,b) = 0_{\mathbb{R}^2} \Leftrightarrow \begin{cases} \left(\sum_{k=1}^{n}x_k^2\right)a + \left(\sum_{k=1}^{n}x_k\right)b = \sum_{k=1}^{n}x_ky_k \\ \left(\sum_{k=1}^{n}x_k\right)a + nb = \sum_{k=1}^{n}y_k \end{cases}$$

En divisant chaque équation par n il vient

$$\nabla f(a,b) = 0_{\mathbb{R}^2} \Leftrightarrow \begin{cases} \frac{1}{n}\left(\sum_{k=1}^{n}x_k^2\right)a + \bar{x}b = \frac{1}{n}\sum_{k=1}^{n}x_ky_k \\ \bar{x}a + b = \bar{y} \end{cases} \Leftrightarrow \begin{cases} (\sigma_x^2+\bar{x}^2)a + \bar{x}b = \sigma_{xy} + \bar{x}\bar{y} \\ b = \bar{y} - a\bar{x} \end{cases}$$

$$\nabla f(a,b) = 0_{\mathbb{R}^2} \Leftrightarrow \begin{cases} b = \bar{y} - a\bar{x} \\ \sigma_{xy} + \bar{x}\bar{y} = (\sigma_x^2+\bar{x}^2)a + \bar{x}(\bar{y}-a\bar{x}) = \sigma_x^2 a + \bar{x}\bar{y} \end{cases} \underset{\sigma_x^2>0}{\Leftrightarrow} \begin{cases} b = \bar{y} - a\bar{x} \\ a = \frac{\sigma_{xy}}{\sigma_x^2} \end{cases}$$

f admet un point critique $(\hat{a}, \hat{b})$ et un seul.
$$\begin{cases} \hat{a} = \dfrac{\sigma_{x,y}}{\sigma_x^2} \\ \hat{b} = \bar{y} - \hat{a}\,\bar{x} \end{cases}$$

e) Soit $(a,b) \in \mathbb{R}^2$. $\dfrac{\partial f}{\partial x}(a,b) = -2\sum_{k=1}^{n} x_k(y_k - a x_k - b)$ et $\dfrac{\partial f}{\partial y}(a,b) = -2\sum_{k=1}^{n}(y_k - a x_k - b)$.

$\dfrac{\partial^2 f}{\partial x^2}(a,b) = 2\sum_{k=1}^{n} x_k^2$, $\dfrac{\partial^2 f}{\partial y^2}(a,b) = 2\sum_{k=1}^{n} 1 = 2n$ et $\dfrac{\partial^2 f}{\partial y \partial x}(a,b) = 2\sum_{k=1}^{n} x_k$.

"$rt - s^2$" $= \dfrac{\partial^2 f}{\partial x^2}(\hat{a},\hat{b})\dfrac{\partial^2 f}{\partial y^2}(a,b) - \left(\dfrac{\partial^2 f}{\partial y \partial x}(\hat{a},\hat{b})\right)^2 = 4\left(\sum_{k=1}^{n} x_k^2\right) \times n - 4\left(\sum_{k=1}^{n} x_k\right)^2$.

"$rt - s^2$" $= 4n\left[\sum_{k=1}^{n} x_k^2 - \dfrac{1}{n}(n\bar{x})^2\right] = 4n\left[\sum_{k=1}^{n} x_k^2 - n\bar{x}^2\right] \underset{c)}{>} 0$.

de plus $r = 2\sum_{k=1}^{n} x_k^2 \geq 0$.

le cours indique alors que f admet en $(\hat{a}, \hat{b})$ un minimum local. minimum global ?? →

Soit $(a,b) \in \mathbb{R}^2$. Posons $\alpha = a - \hat{a}$ et $\beta = b - \hat{b}$. Posons $\Delta(a,b) = f(a,b) - f(\hat{a},\hat{b})$.

$$\Delta(a,b) = \sum_{k=1}^{n}\left[(y_k - a x_k - b)^2 - (y_k - \hat{a}x_k - \hat{b})^2\right].$$

$$\Delta(a,b) = \sum_{k=1}^{n}\left[(y_k - a x_k - b - y_k + \hat{a}x_k + \hat{b})(y_k - a x_k - b + y_k - \hat{a}x_k - \hat{b})\right].$$

$$\Delta(a,b) = \sum_{k=1}^{n}\left[(-\alpha x_k - \beta)(2y_k - x_k(a + \hat{a}) - b - \hat{b})\right].$$

$$\Delta(a,b) = \sum_{k=1}^{n}\left[(-\alpha x_k - \beta)(2y_k - x_k(2\hat{a} + \alpha) - (2\hat{b} + \beta))\right].$$

$$\Delta(a,b) = \sum_{k=1}^{n}\left[(-\alpha x_k - \beta)(-\alpha x_k - \beta + 2(y_k - \hat{a}x_k - \hat{b}))\right].$$

$$\Delta(a,b) = \sum_{k=1}^{n}(\alpha x_k + \beta)^2 + 2\sum_{k=1}^{n}\left[(\alpha x_k + \beta)(\hat{a}x_k + \hat{b} - y_k)\right].$$

a $\sum_{k=1}^{n}\left[(\alpha x_k+\beta)(\hat{a}x_k+\hat{b}-\hat{y}_k)\right]=\alpha\left[\underbrace{(\sum_{k=1}^{n}x_k^2)\hat{a}+(\sum_{k=1}^{n}x_k)\hat{b}-\sum_{k=1}^{n}x_ky_k}_{=0}\right]+\beta\left[\underbrace{(\sum_{k=1}^{n}x_k)\hat{a}+n\hat{b}-\sum_{k=1}^{n}y_k}_{=0}\right]$

$\rightarrow$ $(\hat{a},\hat{b})$ est solution de (S).

Alors $\Delta(a,b)=\sum_{k=1}^{n}(\alpha x_k+\beta)^2\geqslant 0$. Donc $f(a,b)-f(\hat{a},\hat{b})\geqslant 0$; $f(a,b)\geqslant f(\hat{a},\hat{b})$.

$\forall (a,b)\in\mathbb{R}^2$, $f(a,b)\geqslant f(\hat{a},\hat{b})$. f admet un minimum global en $(\hat{a},\hat{b})$.

Exercice.. "Retrouver" ce résultat en utilisant la formule de Taylor-Lagrange à l'ordre 1.

(Q3) Rappel.. $r_{x,y}=\dfrac{\sigma_{x,y}}{\sqrt{\sigma_x^2\sigma_y^2}}$. $\quad r_{x,y}^2=\dfrac{\sigma_{x,y}^2}{\sigma_x^2\,\sigma_y^2}$.

Remarque.. $\sigma_y^2>0$ car $y_1,y_2,\dots,y_n$ ne sont pas tous égaux (voir Q2 c)).

$f(\hat{a},\hat{b})=\sum_{k=1}^{n}(y_k-\hat{a}x_k-\hat{b})^2\geqslant 0$. Rappelons que $\hat{b}=\bar{y}-a\bar{x}$.

$\sum_{k=1}^{n}(y_k-\hat{a}x_k-\hat{b})^2=\sum_{k=1}^{n}(y_k-\hat{a}x_k-\bar{y}+a\bar{x})^2=\sum_{k=1}^{n}\left((y_k-\bar{y})-\hat{a}(x_k-\bar{x})\right)^2$

$f(\hat{a},\hat{b})=\sum_{k=1}^{n}(y_k-\bar{y})^2-2\hat{a}\sum_{k=1}^{n}(y_k-\bar{y})(x_k-\bar{x})+\hat{a}^2\sum_{k=1}^{n}(x_k-\bar{x})^2$.

$f(\hat{a},\hat{b})=n\left[\sigma_y^2-2\hat{a}\underbrace{\frac{1}{n}\sum_{k=1}^{n}(y_k-\bar{y})(x_k-\bar{x})}_{\sigma_{x,y}}+\hat{a}^2\sigma_x^2\right]$. Rappelons que $\hat{a}=\dfrac{\sigma_{x,y}}{\sigma_x^2}$.

$f(\hat{a},\hat{b})=n\left[\sigma_y^2-2\dfrac{\sigma_{x,y}}{\sigma_x^2}\times\sigma_{x,y}+\dfrac{\sigma_{x,y}^2}{\sigma_x^4}\sigma_x^2\right]=n\left[\sigma_y^2-\dfrac{\sigma_{x,y}^2}{\sigma_x^2}\right]$.

$f(\hat{a},\hat{b})=n\sigma_y^2\left[1-\dfrac{\sigma_{x,y}^2}{\sigma_x^2\sigma_y^2}\right]=n\sigma_y^2(1-r_{x,y}^2)$. Or $f(\hat{a},\hat{b})\geqslant 0$ et $n\sigma_y^2>0$.

Donc $1-r_{x,y}^2\geqslant 0$. $r_{x,y}^2\leqslant 1$. Ainsi $\underline{|r_{x,y}|\leqslant 1}$.

R.

Question 11 **HEC 2010** $\boxed{\text{F 1}}$

Soit p un entier supérieur ou égal à 2 et A une matrice de $\mathcal{M}_p(\mathbb{R})$ telle que $2A^4 - 2A^3 + I = 0$ (I désigne la matrice identité d'ordre p et 0 la matrice nulle carrée d'ordre p).

Q1. Déterminer l'ensemble des entiers n de $\mathbb{Z}$ pour lesquels la matrice $A + \dfrac{n}{n^2+1} I$ est inversible.

Q2. Existe-t-il un entier naturel n tel que $\Big((n^2+1)\,A^2 + n\,A\Big)^n = B$ où B est la matrice de $\mathcal{M}_p(\mathbb{R})$ dont tous les coefficients sont égaux à n ?

Q1 $\operatorname{Sp} A \subset \{x \in \mathbb{R} \mid 2x^4 - 2x^3 + 1 = 0\}$.

Posons $\forall x \in \mathbb{R}$, $f(x) = 2x^4 - 2x^3 + 1$. f est dérivable sur $\mathbb{R}$ et $\forall x \in \mathbb{R}$, $f'(x) = 8x^3 - 6x^2 = 8x^2\left(x - \frac{3}{4}\right)$

f est croissante sur $\left[\frac{3}{4}, +\infty\right[$ et décroissante sur $\left]-\infty, \frac{3}{4}\right]$.

$\forall x \in \mathbb{R}$, $f(x) \geq f\left(\frac{3}{4}\right) = \frac{101}{128} > 0$. $\{x \in \mathbb{R} \mid 2x^4 - 2x^3 + 1 = 0\} = \emptyset$.

A n'a pas de valeur propre dans $\mathbb{R}$. $\forall \lambda \in \mathbb{R}$, $A - \lambda I$ est inversible.

Donc $\forall n \in \mathbb{Z}$, $A + \frac{n}{n^2+1} I$ est inversible. En particulier A est inversible ($n = 0$...)

Q2 Soit $n \in \mathbb{N}$. $A + \frac{n}{n^2+1} I$ et A sont inversibles.

Donc $A^2 + \frac{n}{n^2+1} A$ est inversible comme produit de deux matrices inversibles.

Comme $n^2 + 1 \neq 0$ $(n^2+1)A^2 + nA$ est inversible.

$\left((n^2+1)A^2 + nA\right)^n$ est alors inversible (clair pour $n = 0$ et $n = 1$; si $n \geq 2$: produit de n matrices inversibles).

Comme $p \geq 2$, $B = \begin{pmatrix} n & \cdots & n \\ \vdots & & \vdots \\ n & \cdots & n \end{pmatrix}$ n'est pas inversible(*). On ne peut donc pas avoir

$\left((n^2+1)A^2 + nA\right)^n = B$.

La réponse est NON.

(*) $\operatorname{rg} B = 1 < p$ ...