Informatique Appliquée au Calcul Scientifique 2

Séance 3

Inégalité de Cauchy-Schwarz

Droite des moindres carrés

Table des matières

I. Inégalité de Cauchy-Schwarz *2*

II. Droites des moindres carrés *3*

III. TP2 : Régression linéaire *6*

Cours de B Moreau

I. Inégalité de Cauchy-Schwarz

Théorème :
Soient $a_1, \dots, a_n \in \mathbb{R}$ et $b_1, \dots, b_n \in \mathbb{R}$, on a l'inégalité suivante :

$$\left(\sum_{i=1}^{n} a_i . b_i\right)^2 \leq \sum_{i=1}^{n} {a_i}^2 . \sum_{i=1}^{n} {b_i}^2 \quad (1)$$

On la retrouve aussi sous cette forme :

$$\sum_{i=1}^{n} |a_i b_i| \leq \sqrt{\sum_{i=1}^{n} {a_i}^2} . \sqrt{\sum_{i=1}^{n} {b_i}^2} \quad (2)$$

Démonstration :
Considérons le polynôme suivant :

$$P(X) = \sum_{i=1}^{n} (a_i X + b_i)^2$$

Il s'agit d'une somme de carré, donc le résultat, un polynôme de degré 2, est obligatoirement positif ou nul.
Il y a donc soit pas de racine, soit une racine double.
Donc, le discriminant est négatif ou nul.
De plus,

$$(a_i X + b_i)^2 = {a_i}^2 X^2 + 2a_i . b_i + {b_i}^2$$

On peut donc écrire :

$$P(X) = \sum_{i=1}^{n} (a_i X + b_i)^2 = \left(\sum_{i=1}^{n} {a_i}^2\right) X^2 + 2\left(\sum_{i=1}^{n} a_i . b_i\right) X + \sum_{i=1}^{n} {b_i}^2$$

$$\Delta = \left(2\sum_{i=1}^{n} a_i . b_i\right)^2 - 4.\left(\sum_{i=1}^{n} {a_i}^2\right).\left(\sum_{i=1}^{n} {b_i}^2\right) = 4\left(\sum_{i=1}^{n} a_i . b_i\right)^2 - 4\left(\sum_{i=1}^{n} {a_i}^2\right).\left(\sum_{i=1}^{n} {b_i}^2\right)$$

$$= 4\left(\left(\sum_{i=1}^{n} a_i . b_i\right)^2 - \left(\sum_{i=1}^{n} {a_i}^2\right).\left(\sum_{i=1}^{n} {b_i}^2\right)\right) \leq 0$$

$$\Leftrightarrow \left(\sum_{i=1}^{n} a_i . b_i\right)^2 - \left(\sum_{i=1}^{n} {a_i}^2\right).\left(\sum_{i=1}^{n} {b_i}^2\right) \leq 0$$

$$\Leftrightarrow \left(\sum_{i=1}^{n} a_i . b_i\right)^2 \leq \sum_{i=1}^{n} {a_i}^2 . \sum_{i=1}^{n} {b_i}^2$$

Cas d'égalité :
Soient 2 vecteurs $\vec{a}$ et $\vec{b}$ de coordonnées réelles respectives $(a_1, \dots, a_n)$ et $(b_1, \dots, b_n)$:

$$\left(\sum_{i=1}^{n} a_i . b_i\right)^2 = \sum_{i=1}^{n} {a_i}^2 . \sum_{i=1}^{n} {b_i}^2 \Leftrightarrow \vec{a} \text{ et } \vec{b} \text{ sont colinéaires}$$

La démonstration de l'égalité se fait de la même manière.
Il y a égalité si et seulement si $\Delta = 0$, donc si et seulement si il existe $\lambda \in \mathbb{R}$ tel que $P(\lambda) = 0 \Leftrightarrow b_i = -\lambda a_i$ pour tout $i \in [\![1, n]\!]$.
Interprétation géométrique :
En dimension 2 ou 3, dans le cas de la forme (2), dans le membre de gauche on reconnait le produit scalaire de deux vecteurs et dans le membre de droite, le produit de leur norme.

Cours de B Moreau

Utilisation avec les moyennes :
Si $(x_j)_{j=1,..,N}$ désigne une famille de nombres réels, on pose :

$$< x >= \frac{1}{N}\sum_{j=1}^{N} x_j$$

On a donc :

$$< a.b >= \frac{1}{N}\sum_{i=1}^{N} a_i.b_i$$

$$< a^2 >= \frac{1}{N}\sum_{j=1}^{N} {a_i}^2$$

$$< b^2 >= \frac{1}{N}\sum_{j=1}^{N} {b_i}^2$$

L'inégalité de Cauchy-Schwarz peut dont s'écrire :

$$< a.b >^2 \leq < a^2 >< b^2 >$$

Avec $b_i = 1$ pour tout j, $< b^2 > \geq\ 1$ et $< a.b > = < a >$.
On obtient ainsi :

$$< a >^2 \leq < a^2 >$$

On peut la retranscrire en français : « le carré de la moyenne est inférieur ou égal à la moyenne des carré ».

Cours de B Moreau

II. Droites des moindres carrés

Les moindres carrés sont une méthode statistique utilisée pour trouver la meilleure approximation possible d'un ensemble de données par une fonction. Cette technique est très courante en analyse de données, en économétrie, en physique, en ingénierie, et dans de nombreux autres domaines où il est nécessaire de modéliser une relation entre des variables.

La méthode des moindres carrés vise à minimiser la somme des carrés des écarts entre les valeurs observées et les valeurs prédites par un modèle. En d'autres termes, elle cherche à ajuster un modèle de manière à ce qu'il soit le plus proche possible des données.

A chaque couple (x_i, y_i), pour $i = 1,2, \dots , n$, on fait correspondre un point M_i. On obtient un nuage de point.

On cherche à ajuster une droite d'équation $y = ax + b$ au nuage de points. Le critère d'ajustement est la distance totale entre les points du nuage $M_i(x_i, y_i)$ et les points $P_i(x_i,\ ax_i + b)$ correspondant sur la droite d'ajustement.
On cherche alors le couple $(\hat{a}, \hat{b})$ qui minimise :

$$f(a, b) = \sum_{i=1}^{n}(y_i - (\ ax_i + b))^2 = \sum_{i=1}^{n}(y_i -\ ax_i - b)^2$$

Conditions nécessaires du minimum :

$$\frac{\partial f}{\partial a}(\hat{a}, \hat{b}) = 0\ et\ \frac{\partial f}{\partial b}(\hat{a}, \hat{b}) = 0$$

$$\frac{\partial f}{\partial a}\left(\hat{a},\hat{b}\right) = 2\sum_{i=1}^{n}((y_i - \ \hat{a}x_i - \hat{b})\ (-x_i) = 0$$

$$\Leftrightarrow \sum_{i=1}^{n}\left((y_i - \ \hat{a}x_i - \hat{b}\right)\ x_i = 0$$

$$\Leftrightarrow \sum_{i=1}^{n}\ x_i y_i - \hat{a}\sum_{i=1}^{n}\ {x_i}^2 - \hat{b}\sum_{i=1}^{n}\ x_i = 0$$

$$\frac{\partial f}{\partial b}\left(\hat{a},\hat{b}\right) = 2\sum_{i=1}^{n}((y_i - \ \hat{a}x_i - \hat{b})(-1)$$

$$\Leftrightarrow \sum_{i=1}^{n}(y_i - \ \hat{a}x_i - \hat{b}) = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} y_i - \hat{a}\sum_{i=1}^{n}\ x_i - n\hat{b} = 0$$

En divisant par n nos deux conditions, on obtient :

$$\frac{1}{n}\sum_{i=1}^{n}\ x_i y_i - \hat{a}\frac{1}{n}\sum_{i=1}^{n}\ {x_i}^2 - \hat{b}\frac{1}{n}\sum_{i=1}^{n}\ x_i = 0\ et\ \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{a}\frac{1}{n}\sum_{i=1}^{n}\ x_i - \hat{b} = 0$$

Cours de B Moreau

Rappel de quelques formules :
Moyennes :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}\ x_i\ \ et\ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Variances :

$${s_x}^2 = \frac{1}{n}\sum_{i=1}^{n}(\ x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}\ {x_i}^2 - (\bar{x})^2\ et\ {s_y}^2 = \frac{1}{n}\sum_{i=1}^{n}\ {y_i}^2 - (\bar{y})^2$$

Covariance :

$$cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}\ (x_i - \bar{x})\ (y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{n}\ x_i y_i - \bar{x}\bar{y}$$

Nos deux conditions deviennent :

$$cov(x,y) + \bar{x}\bar{y} - \hat{a}({s_x}^2 + (\bar{x})^2 - \hat{b}\bar{x} = 0 \; et \; \bar{y} - \hat{a}\bar{x} - \hat{b} = 0$$

De la seconde condition, on en déduit que :

$$\hat{b} = \bar{y} - \hat{a}\bar{x}$$

Qui nous donne par substitution dans la première condition :

$$cov(x,y) + \bar{x}\bar{y} - \hat{a}({s_x}^2 + (\bar{x})^2 - (\bar{y} - \hat{a}\bar{x})\bar{x} = 0$$

$$\Leftrightarrow cov(x,y) - \hat{a}{s_x}^2 = 0 \Leftrightarrow \hat{a} = \frac{cov(x,y)}{{s_x}^2}$$

On a donc le couple :

$$\hat{a} = \frac{cov(x,y)}{{s_x}^2} \; et \; \hat{b} = \bar{y} - \hat{a}\bar{x}$$

On admet que ce couple minimise $f(a,b)$.

La droite des moindres carrés (ou de régression linéaire) de y en x a donc pour équation :

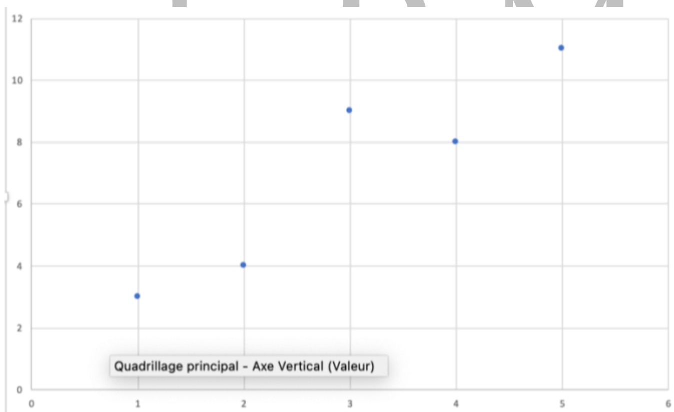
$$D_{y/x} : y = \hat{a}x + \hat{b}, \text{avec } \hat{a} = \frac{cov(x,y)}{{s_x}^2} \; et \; \hat{b} = \bar{y} - \hat{a}\bar{x}$$

Exemple :

On considère la série double statistique suivante :

x_i	2	3	5	1	4
y_i	4	9	11	3	8

On obtient le nuage de points suivant :



La droite de régression de y en x a pour équation : $y = \hat{a}x + \hat{b}$, avec $\hat{a} = \frac{cov(x,y)}{{s_x}^2}$ et $\hat{b} = \bar{y} - \hat{a}\bar{x}$.

x_i	y_i	$x_i y_i$	${x_i}^2$
2	4	8	4
3	9	27	9
5	11	55	25
1	3	3	1
4	8	32	16
15	35	125	55

On a donc :

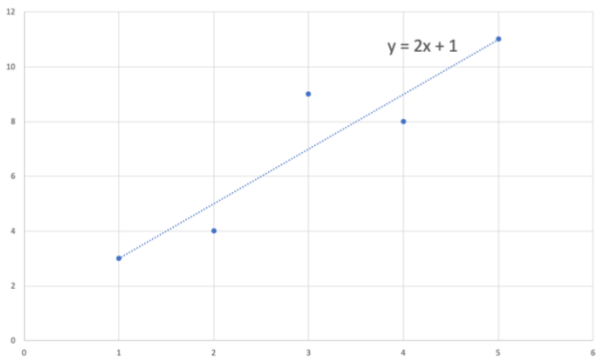
$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{15}{5} = 3$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{35}{5} = 7$

$cov(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = \frac{1}{5} \times 125 - 3 \times 7 = 4,$

${s_x}^2 = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - (\bar{x})^2 = \frac{1}{5} \times 55 - 3^2 = 2.$

On trouve :

$$\text{avec } \hat{a} = \frac{cov(x,y)}{{s_x}^2} = \frac{4}{2} = 2 \; et$$
$$\hat{b} = \bar{y} - \hat{a}\bar{x} = 7 - 2 \times 3 = 1$$



La droite de régression a donc pour équation :
$y = 2x + 1.$

Coefficient de corrélation linéaire entre x et y :

Le coefficient de corrélation linéaire est donné par :

$$r_{x,y} = \frac{cov(x,y)}{s_x s_y}$$

$|r_{x,y}| = 1$ si et seulement si les points M_i sont alignés sur $D_{y/x}$.
Plus $|r_{x,y}|$ est proche de 1, meilleur est l'ajustement de la droite des moindres carrés au nuage de points.

Coefficient de détermination linéaire :

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = {r^2}_{x,y} \text{ où } \hat{y}_i \; est \; la \; valeur \; prédite \; par \; le \; modèle$$

Dans notre exemple, nous avons :

$$r_{x,y} = \frac{cov(x,y)}{s_x s_y} = \frac{4}{\sqrt{2 \times 9.2}} \approx 0{,}9325$$

Cours de B Moreau

Cela montre une corrélation linéaire positive forte.

$$R^2 = {r^2}_{x,y} \approx 0{,}8696$$

86,96% des variations de y peuvent être expliquées par la variable x.

III. TP2 : Régression linéaire

Écrire un programme sous Python qui permettra de calculer les paramètres de la droite de régression linéaire et l'affichage de R^2 ainsi que le tracé de votre droite et des points données.
Le tester avec l'exemple du cours pour le valider.
Point Python :
Avec numpy :

```
import numpy as np
# Données d'exemple
x = np.array([2, 3, 5, 1, 4])
y = np.array([4, 9, 11, 3, 8])
```

x+y donnera la somme matricielle de x et y. ici : `[6, 12, 16, 4, 12]`
x**2 donnera le carré des coordonnées du vecteur x. Ici, `[ 4  9 25  1 16]`
`np.sum(x)` ajoutera tous les éléments de notre matrice ligne. Ici : 15.
`np.mean(x)` fera la moyenne des éléments de la matrice.

Cours de B Moreau