Nom :

Mathématiques approfondies
Cours ECG 2

Partie VIII

Chapitres

Lycée Saint Louis 2024/2025

CHAPITRE 17

Compléments sur les fonctions de plusieurs variables

If there is a 50-50 chance that something can go wrong, then nine times out of 10 it will.

PAUL HARVEY
Animateur radio américain (1918-2009)

1 Rappels sur les fonctions d'une variable réelle

Avec la continuité

Rappelons la définition de l'image d'une fonction $f : \mathrm{I} \to \mathbb{R}$, noté $f(\mathrm{I})$, comme l'ensemble des réels qui admettent un antécédent par f. $\quad f(\mathrm{I}) = \{y \in \mathbb{R} \mid \exists\, x \in \mathrm{I},\ f(x) = y\} = \{f(x) \mid x \in \mathrm{I}\}$.

Théorème 1 (image d'un segment)

Pour toute application continue $f : \mathrm{I} \to \mathbb{R}$ où I est un segment de $\mathbb{R}$, l'image $f(\mathrm{I})$ est aussi un segment. Alors, f admet un minimum et un maximum et $f(\mathrm{I}) = \left[\min_{\mathrm{I}} f, \max_{\mathrm{I}} f\right]$.

Remarque. On dit alors que f est bornée et atteint ses bornes. Il existe α, β tels que $\forall\, x \in \mathrm{I}, \quad f(\alpha) \leqslant f(x) \leqslant f(\beta)$.

Avec la dérivée

Théorème 2 (condition nécessaire pour un extrema)

Soit $f : \mathrm{I} \to \mathbb{R}$.

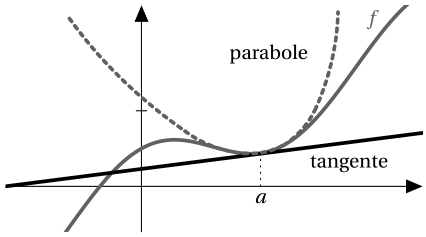
Si
- f admet un extremum local en a,
- f est dérivable en a, et
- a appartient à un *intervalle ouvert* inclus dans I, **alors** $f'(a) = 0$.

Remarque. La réciproque est fausse : tout point critique ne donne pas un extremum. La fonction $x \in \mathbb{R} \mapsto x^3$ en 0 fournit un contre-exemple.

Avec les dérivées successives

Théorème 3 (formule de Taylor-Young)

Soient $f \in \mathscr{C}^n(\mathrm{I})$, $a \in \mathrm{I}$ et $n \in \mathbb{N}$. Au voisinage du réel a, $\quad f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + o\big((x-a)^n\big)$.



Exemples. • $n=1$

$$f(x)=\underbrace{f(a)+f'(a)(x-a)}^{\text{équation de la tangente}}+o(x-a).$$

• $n=2$

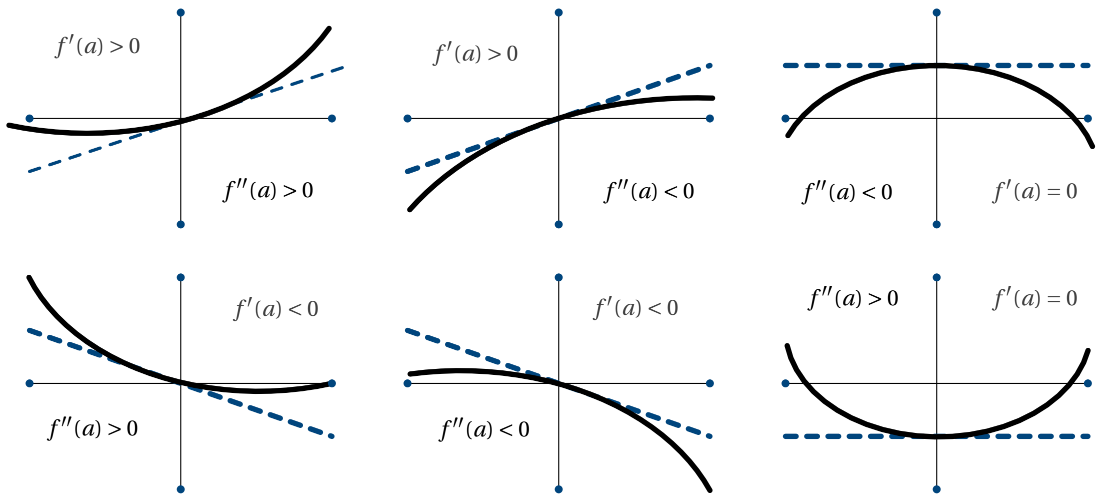
$$f(x)=\overbrace{f(a)+f'(a)(x-a)+\frac{f''(a)}{2}(x-a)^2}^{\text{équation d'une parabole si } f''(a)\neq 0}+o_a\big((x-a)^2\big).$$

Allure du graphe d'une fonction au voisinage d'un point

En reprenant le cas $n=2$, au voisinage de a, la courbe représentative de f est « proche » de la courbe d'équation

$$y=f(a)+f'(a)(x-a)+\frac{f''(a)}{2}(x-a)^2.$$

Lorsque $f''(a)\neq 0$, la courbe est une parabole.
Distinguons suivant le signe de $f'(a)$ et $f''(a)$. En gras, la parabole, en pointillés, la tangente.



Exercice 1

✦✦ **Égalité de Taylor-Lagrange**
Soient f une fonction de classe $\mathscr{C}^2$ sur un intervalle I et $a\in$ I. Justifier que pour tout $x\in$ I, il existe c compris entre a et x tel que

$$f(x)=f(a)+f'(a)(x-a)+\frac{1}{2}(x-a)^2f''(c).$$

On pourra introduire la fonction φ définie sur I *par* $\varphi(t)=f(x)-f(t)-f'(t)(x-t)-(x-t)^2\lambda/2$ *pour un réel λ choisi afin d'avoir* $\varphi(a)=0$.

Avec la convexité

Théorème 4 (condition suffisante pour un minimum global)

Soient $f:\mathrm{I}\to\mathbb{R}$ et a appartenant à un intervalle ouvert inclus dans I.

Si
- f est convexe.
- f est dérivable en a.
- a est un point critique ($f'(a)=0$).

Alors f admet un minimum global en a.

Remarque. On a un énoncé similaire dans le cas d'une fonction concave où on obtient un maximum global.

2 Fonctions définies sur une partie de $\mathbb{R}^n$

2.1 Fermés, ouverts et bornés

Dans la suite, on considère le produit scalaire canonique sur $\mathbb{R}^n$ et on note $\|\cdot\|$ la norme associée.

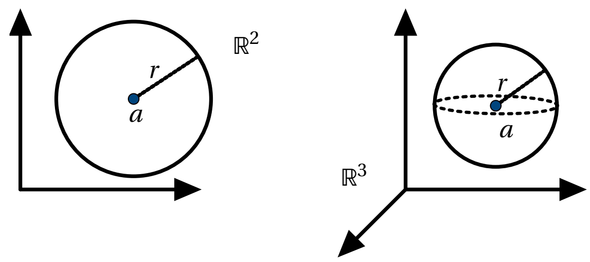
Définition 5 (boule ouverte)

Soient $a \in \mathbb{R}^n$ et $r \in \mathbb{R}_*^+$, on définit la **boule ouverte** de centre a et de rayon r par :

$$\mathscr{B}(a,r) = \left\{x \in \mathbb{R}^n \;\middle|\; \|x-a\| < r\right\}.$$

C'est l'ensemble des points de $\mathbb{R}^n$ situés à une distance strictement inférieure à r du point a.

À gauche, une boule ouverte dans $\mathbb{R}^2$ et à droite dans $\mathbb{R}^3$. Il faut exclure dans les deux cas la « frontière ».



Définition 6 (partie ouverte)

Une partie $\mathscr{O}$ de $\mathbb{R}^n$ est dite **ouverte** si, en chaque point de $\mathscr{O}$, il existe une boule ouverte centrée en ce point et contenue dans $\mathscr{O}$. Autrement dit,

$$\forall\, a \in \mathscr{O}, \quad \exists\, r \in \mathbb{R}_*^+, \qquad \mathscr{B}(a,r) \subset \mathscr{O}.$$

Exemples.
→ $\mathbb{R}^n$ est une partie ouverte puisqu'une boule ouverte en un point quelconque de $\mathbb{R}^n$ est toujours contenue dans $\mathbb{R}^n$.
→ Toute boule ouverte est un ouvert.
→ $\mathbb{R}^n$ privé d'un ensemble fini de points est ouvert.
→ Soient $\mathrm{I}_1, \mathrm{I}_2, \ldots, \mathrm{I}_n$, n intervalles ouverts de $\mathbb{R}$. La partie $\mathrm{I}_1 \times \mathrm{I}_2 \times \ldots \times \mathrm{I}_n$ est une partie ouverte de $\mathbb{R}^n$ (on parle de pavé ouvert).

Remarques.
→ Une union quelconque d'ouverts est un ouvert.
→ Une intersection *finie* d'ouverts reste un ouvert. Par contre, cela devient faux pour une intersection infinie. Un contre-exemple est donné par $\bigcap_{n \in \mathbb{N}^*} \mathscr{B}(a, 1/n) = \{a\}$.

Définition 7 (partie fermée)

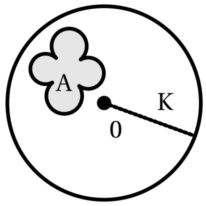
Une partie F de $\mathbb{R}^n$ est dite **fermée** de $\mathbb{R}^n$ si son complémentaire $\overline{\mathrm{F}}$ est un ouvert de $\mathbb{R}^n$.

Proposition 8 (conditions suffisantes pour une partie ouverte/fermée)

Soient $\varphi : \mathbb{R}^n \to \mathbb{R}$, une fonction continue sur $\mathbb{R}^n$ et $r \in \mathbb{R}$.
→ La partie $\left\{x \in \mathbb{R}^n \mid \varphi(x) < r\right\}$ est un ouvert de $\mathbb{R}^n$.
→ La partie $\left\{x \in \mathbb{R}^n \mid \varphi(x) \leqslant r\right\}$ est un fermé de $\mathbb{R}^n$.
→ La partie $\left\{x \in \mathbb{R}^n \mid \varphi(x) = r\right\}$ est un fermé de $\mathbb{R}^n$.

Définition 9 (partie bornée)

Soit A une partie de $\mathbb{R}^n$.
La partie A est dite **bornée** s'il existe $K \in \mathbb{R}^+$ tel que pour tout $x \in A$, $\|x\| \leqslant K$.



Remarque. Autrement dit, une partie A de $\mathbb{R}^n$ est bornée si elle est incluse dans une boule de centre l'origine :

$$\exists K \in \mathbb{R}_*^+, \qquad A \subset \mathscr{B}(0, K).$$

Exercice 2

Exemples

✦ 1. Parmi les parties suivantes, préciser les parties ouvertes, fermées, bornées.

$$A_1 = \left\{(x, y) \in \mathbb{R}^2 \mid x + 2y = 5\right\}, \quad A_2 = \left\{(x, y) \in \mathbb{R}^2 \mid x > 0, y > 0 \text{ et } x + y < 1\right\}$$
$$A_3 = \left\{x \in \mathbb{R}^n \mid \langle u, x\rangle \geqslant 0\right\}, \quad A_4 = \left\{x \in \mathbb{R}^n \mid \langle u, x\rangle = 0\right\} \quad \text{où} \quad u \in \mathbb{R}^n \setminus \{0\}.$$

✦✦ 2. Justifier que tout sous-espace vectoriel de $\mathbb{R}^n$ est un fermé.
On pourra exprimer le sous-espace comme intersection de parties du type A_4.

2.2 Rappels et compléments sur le cas $\mathscr{C}^1$

Les définitions vues au premier semestre s'étendent à une partie ouverte de $\mathbb{R}^n$. Ainsi, une fonction $f : \mathscr{O} \to \mathbb{R}$ est de classe $\mathscr{C}^1$ sur $\mathscr{O}$ si les dérivées partielles existent sur $\mathscr{O}$ et sont continues en chaque point de $\mathscr{O}$. On rappelle que pour tout $a \in \mathscr{O}$, la fonction f admet en a un unique développement limité à l'ordre 1. C'est-à-dire, il existe $\varepsilon : \mathscr{O} \to \mathbb{R}$ telle que, pour $h \in \mathbb{R}^n$ avec $a + h \in \mathscr{O}$,

$$f(a+h) = f(a) + \left\langle \nabla(f)(a), h\right\rangle + \|h\|\varepsilon(h) \quad \text{et} \quad \varepsilon(h) \underset{h\to 0}{\longrightarrow} 0.$$

3 Fonctions de classe $\mathscr{C}^2$

3.1 Définitions et exemples

Définition 10 (dérivées partielles d'ordre 2)

Soient un ouvert $\mathscr{O}$ de $\mathbb{R}^n$, une fonction $f : \mathscr{O} \to \mathbb{R}$, $a \in \mathscr{O}$ et $i, j \in [\![1; n]\!]$. On suppose que la dérivée partielle première $\partial_j f$ est définie sur $\mathscr{O}$. Si la dérivée partielle $\partial_i\left(\partial_j f\right)$ est définie en a, on dit que f admet une **dérivée partielle d'ordre 2 d'indice** (i, j) **en** a et on note $\partial^2_{i,j} f(a)$ pour $\partial_i\left(\partial_j f\right)(a)$.

Remarque. Si la dérivée partielle $a \in \mathscr{O} \mapsto \partial_i\left(\partial_j f\right)$ est définie sur $\mathscr{O}$, on dit que f admet, sur $\mathscr{O}$, une dérivée partielle d'ordre 2 d'ordre (i, j) et on note

$$\partial^2_{i,j} f = \partial_i\left(\partial_j f\right).$$

On obtient alors une nouvelle fonction de plusieurs variables $\partial^2_{i,j} f : \mathscr{O} \mapsto \mathbb{R}$.

Exercice 3

✧ Calculer $\partial^2_{2,1} f$ et $\partial^2_{1,1} f + \partial^2_{2,2} f + \partial^2_{3,3} f$ dans les deux cas suivants :

I. $f : (x_1, x_2, x_3) \in \mathbb{R}^3 \mapsto {x_1}^2 + {x_2}^2 - 2{x_3}^2 - x_2 + x_1 + x_1 x_2$ **II.** $f : x \in \mathbb{R}^3 \setminus \{0_{\mathbb{R}^3}\} \mapsto 1/\|x\|$.

Définition 11 (fonction de classe $\mathscr{C}^2$)

Soient un ouvert $\mathscr{O}$ de $\mathbb{R}^n$ et une fonction $f : \mathscr{O} \to \mathbb{R}$. On dit que f est une fonction de **classe** $\mathscr{C}^2$ sur $\mathscr{O}$:

Si
- → La fonction f admet des dérivées partielles d'ordre 2 d'indice (i, j) pour tout $(i, j) \in [\![1; n]\!]^2$.
- → Les dérivées partielles d'ordre 2 sont des fonctions continues de $\mathscr{O}$ dans $\mathbb{R}$.

Exemple. Les fonctions polynomiales sont de classe $\mathscr{C}^2$ sur $\mathbb{R}^n$ car les dérivées partielles d'ordre 2 existent et restent polynomiales (donc continues sur $\mathbb{R}^n$).

Remarque. Comme pour le cas $\mathscr{C}^1$, les combinaisons linéaires, les produits et quotients de fonctions de classe $\mathscr{C}^2$ sur un ouvert $\mathscr{O}$ ou encore la composition par une fonction d'une variable de classe $\mathscr{C}^2$ restent de classe $\mathscr{C}^2$ sur $\mathscr{O}$.

Définition 12 (la matrice hessienne)

Soit $f : \mathscr{O} \subset \mathbb{R}^n \to \mathbb{R}$ admettant des dérivées partielles d'ordre 2 pour tout couple $(i, j) \in [\![1; n]\!]^2$.
La **matrice hessienne** de f au point $a \in \mathscr{O}$, notée $\nabla^2 f(a) \in \mathscr{M}_n(\mathbb{R})$, est définie par

$$\nabla^2 f(a) = \left[\partial^2_{i,j} f(a)\right]_{(i,j)\in[\![1;n]\!]^2} = \begin{bmatrix} \partial^2_{1,1} f(a) & \partial^2_{1,2} f(a) & \cdots & \partial^2_{1,n} f(a) \\ \partial^2_{2,1} f(a) & \partial^2_{2,2} f(a) & \cdots & \partial^2_{2,n} f(a) \\ \vdots & \vdots & \vdots & \vdots \\ \partial^2_{n,1} f(a) & \partial^2_{n,2} f(a) & \cdots & \partial^2_{n,n} f(a) \end{bmatrix}.$$

3.2 Le théorème de Schwarz

Théorème 13 (de Schwarz)

Soit f définie sur un ouvert $\mathscr{O}$ de $\mathbb{R}^n$.

Si f est de classe $\mathscr{C}^2$ sur $\mathscr{O}$,

alors pour tout couple $(i, j) \in [\![1; n]\!]^2$ et pour tout point $a \in \mathscr{O}$, on a $\partial^2_{j,i} f(a) = \partial^2_{i,j} f(a)$.

Remarque. D'après le théorème de Schwarz, la matrice hessienne d'une fonction de classe $\mathscr{C}^2$ est symétrique. En particulier, la matrice est diagonalisable dans une base orthonormée (théorème spectral).

Exercice 4

✦ Soit f la fonction définie sur $\mathbb{R}^3$ par :

$$\forall (x, y, z) \in \mathbb{R}^3, \qquad f(x, y, z) = x e^{x(y^2+z^2+1)}.$$

1. Montrer que f est de classe $\mathscr{C}^2$ sur $\mathbb{R}^3$.
2. Déterminer le seul point critique a de f.
3. Former la hessienne de f au point a et vérifier qu'elle est diagonale.

3.3 Forme quadratique et développement limité d'ordre 2

Préliminaires

• **Formes quadratiques**
Rappelons que la forme quadratique associée à une matrice symétrique $A \in \mathscr{M}_n(\mathbb{R})$ est l'application définie sur $\mathbb{R}^n$ par

$$q(h) = {}^t\!HAH$$

où H est la matrice des coordonnées de h dans la base canonique de $\mathbb{R}^n$.

• **Signe d'une forme quadratique**
Nous avons vu aussi que le signe de la forme quadratique est précisé par le signe des valeurs propres de la matrice associée. Plus précisément

$$\textbf{i)}\,\forall\, u \in \mathrm{E}, \quad q(u) \geqslant 0 \iff \mathrm{Sp}(\mathrm{A}) \subset \mathbb{R}^+\,; \qquad \textbf{ii)}\,\forall\, u \in \mathrm{E}, \quad q(u) \leqslant 0 \iff \mathrm{Sp}(\mathrm{A}) \subset \mathbb{R}^-.$$

On a vu un énoncé plus précis encore connu sous le nom d'encadrement de Rayleigh

$$\forall\, h \in \mathbb{R}^n, \qquad \|h\|^2 \min \mathrm{Sp}(\mathrm{A}) \leqslant q(h) \leqslant \|h\|^2 \max \mathrm{Sp}(\mathrm{A}).$$

• **Formes quadratiques et matrices hessiennes**
Soient $\mathscr{O}$ un ouvert de $\mathbb{R}^n$, $f : \mathscr{O} \to \mathbb{R}$ de classe $\mathscr{C}^2$.
On peut considérer la forme quadratique associée à la matrice hessienne de f en a, notée $q_a : \mathbb{R}^n \to \mathbb{R}$ et définie par

$$\forall\, h = (h_1, \ldots, h_n) \in \mathbb{R}^n, \qquad q_a(h) = {}^t\mathrm{H}\big(\nabla^2 f(a)\big)\,\mathrm{H} \quad \text{où} \quad \mathrm{H} = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}.$$

Cela s'écrit aussi : $$q_a(h) = \sum_{(i,j)\in[\![1,n]\!]^2} \partial^2_{i,j} f(a)\, h_i h_j = \sum_{i=1}^n \partial^2_{ii} f(a)\, {h_i}^2 + 2\sum_{i<j} \partial^2_{i,j} f(a)\, h_i h_j.$$

• **Dérivées directionnelles à l'ordre 2**
Soit $h \in \mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}$ tel que le segment $[a; a+h]$ soit inclus dans $\mathscr{O}$. On pose pour tout $t \in [0;1]$,

$$g_{a,h}(t) = f(a + th).$$

En reprenant le résultat du premier semestre sur les dérivées directionnelles, $g_{a,h}$ est une fonction de classe $\mathscr{C}^1$ sur $[0;1]$ avec pour tout $t \in [0;1]$,

$$g'_{a,h}(t) = \big\langle \nabla f(a+th), h \big\rangle = \sum_{i=1}^n h_i \partial_i f(a+th) \quad \text{où} \quad h = (h_1, \ldots, h_n)\,.$$

À l'aide de la seconde expression, on vérifie que $g_{a,h}$ est de classe $\mathscr{C}^2$ sur $[0;1]$ avec

$$g''_{a,h}(t) = q_{a+th}(h).$$

En particulier $$g''_{a,h}(0) = q_a(h).$$

Développement limité d'ordre 2

Théorème 14 (développement limité d'ordre 2)

Soient $\mathscr{O}$ un ouvert de $\mathbb{R}^n$, $f : \mathscr{O} \to \mathbb{R}$ de classe $\mathscr{C}^2$.
Alors pour tout $a \in \mathscr{O}$, il existe un voisinage $\mathscr{V}$ de $0_{\mathbb{R}^n}$ dans $\mathbb{R}^n$, une fonction $\varepsilon : \mathscr{V} \to \mathbb{R}$ telle que

$$\forall\, h \in \mathscr{V}, \qquad f(a+h) = f(a) + \langle \nabla f(a), h\rangle + \frac{1}{2} q_a(h) + \|h\|^2 \varepsilon(h) \quad \text{et} \quad \varepsilon(h) \underset{h\to 0}{\longrightarrow} 0$$

où q_a est la forme quadratique associée à $\nabla^2 f(a)$.

4 Applications à l'optimisation

4.1 Rappels : extrema locaux/globaux

Définition 15 (extrema locaux)

Soient $f : \mathbb{R}^n \to \mathbb{R}$ et $a \in \mathbb{R}^n$.

- On dit que f a un **maximum local** en a s'il existe $r > 0$ tel que
$$\forall x \in \mathbb{R}^n, \qquad \Big(\|x-a\| < r \quad \Rightarrow \quad f(a) \geqslant f(x) \Big).$$
- On dit que f a un **minimum local** en a s'il existe $r > 0$ tel que
$$\forall x \in \mathbb{R}^n, \qquad \Big(\|x-a\| < r \quad \Rightarrow \quad f(a) \leqslant f(x) \Big).$$
- On dit que f a un **extremum local** si f a un maximum local ou un minimum local.

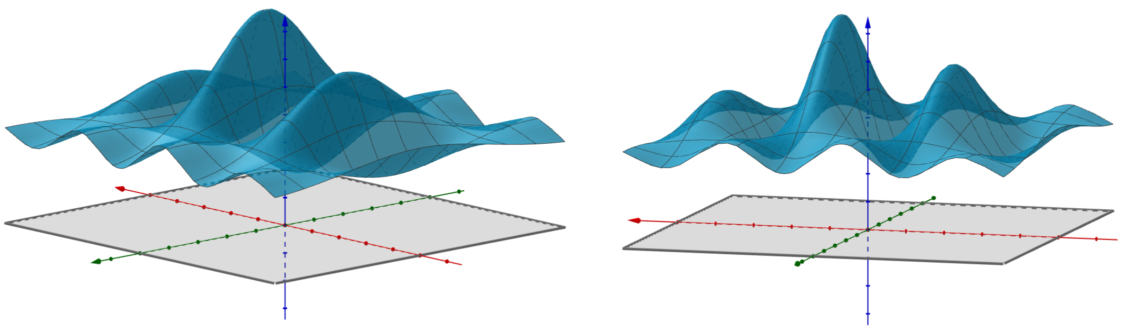
4.2 Condition d'existence d'extremum

Théorème 16 (sur un fermé borné)

Une fonction continue sur une partie fermée bornée admet un maximum global et un minimum global.

Remarque. On peut traduire mathématiquement. Soit $f : \mathcal{F} \subset \mathbb{R}^n \to \mathbb{R}$ continue. Il existe $\alpha, \beta \in \mathbb{R}$ tels que
$$\forall x \in \mathcal{F}, \qquad f(\alpha) \leqslant f(x) \leqslant f(\beta).$$

Ci-dessous, un exemple d'une fonction continue de deux variables définies sur un fermé borné. On vérifie que la fonction est bien bornée et atteint ses bornes.



4.3 Condition nécessaire d'ordre 1

Théorème 17 (condition d'ordre 1)

Soient $f : \mathcal{O} \to \mathbb{R}$ de classe $\mathscr{C}^1$ sur $\mathcal{O}$ et $a \in \mathcal{O}$.

Si
- La partie $\mathcal{O}$ est un ouvert de $\mathbb{R}^n$.
- La fonction f a un extremum local en $a \in \mathbb{R}^n$.

Alors a est un point critique de f, c'est-à-dire $\nabla f(a) = 0$.

Attention. Comme pour le cas d'une variable :

- L'énoncé précédent n'est valable que sur un ouvert.
- La réciproque est fausse. On parle alors de **point col** ou **point selle**.

Exercice 5

Exemple

1. Donner le domaine de définition de f définie par $f(x,y) = x^{\ln(x)} + y^{\ln(y)}$.
2. Vérifier que f est de classe $\mathscr{C}^1$ et préciser les dérivées partielles.
3. Montrer que f a, au plus, un extremum. Est-ce un minimum, un maximum ?

4.4 Condition suffisante d'ordre 2

Théorème 18 (condition d'ordre 2)

Si a est un point critique de f :

- **Si** $\mathrm{Sp}\left(\nabla^2 f(a)\right) \subset \mathbb{R}_+^*$, **alors** f admet un minimum local en a.
- **Si** $\mathrm{Sp}\left(\nabla^2 f(a)\right) \subset \mathbb{R}_-^*$, **alors** f admet un maximum local en a.
- **Si** $\mathrm{Sp}\left(\nabla^2 f(a)\right)$ contient deux réels non nuls de signes distincts, **alors** f n'admet pas d'extremum en a.

Remarque. Les réciproques sont fausses. Par exemple, la fonction définie sur $\mathbb{R}^2$ par $f(x,y) = x^4 + y^4$ admet bien un minimum au point $(0,0)$, pourtant la matrice hessienne est nulle en $(0,0)$ et le spectre est réduit à $\{0\}$ qui n'est pas inclus dans $\mathbb{R}_+^*$.

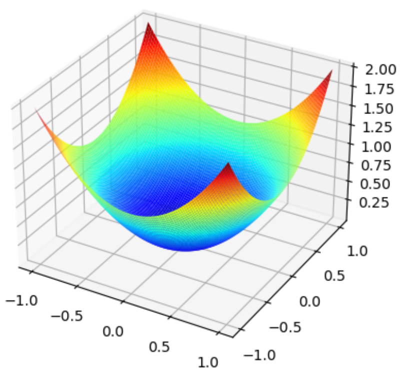
Illustration dans $\mathbb{R}^2$

Posons pour $\alpha, \beta \in \mathbb{R}$, $f_{\alpha,\beta}(x,y) = \alpha x^2 + \beta y^2$. La matrice Hessienne en point critique $(0,0)$ est

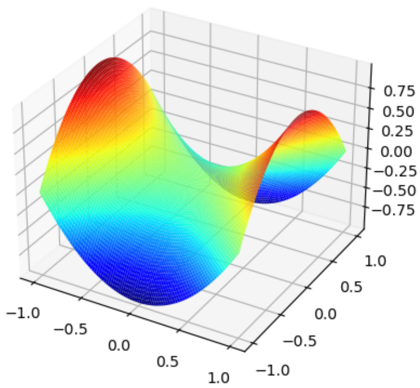
$$\mathrm{H}_{\alpha,\beta} = \begin{bmatrix} 2\alpha & 0 \\ 0 & 2\beta \end{bmatrix}.$$

Trois cas sont bien à distinguer :

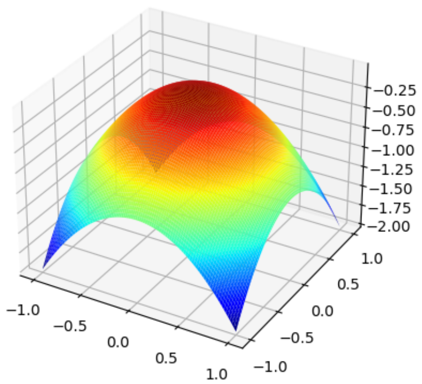
I. $\alpha, \beta \in \mathbb{R}_*^+$,



II. α, β de signes opposés,



III. $\alpha, \beta \in \mathbb{R}_*^-$.



Exemples

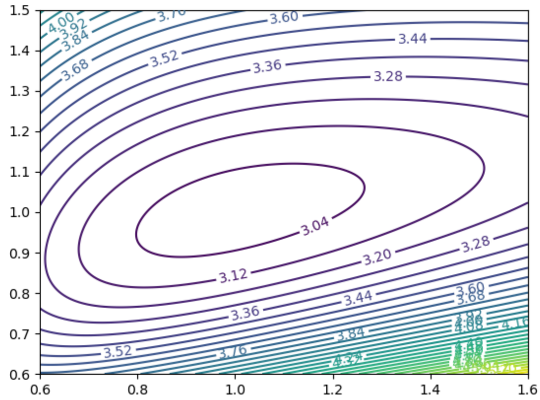
On considère la fonction f définie sur l'ouvert $\mathbb{R}_+^* \times \mathbb{R}_+^*$ par :

$$\forall (x,y) \in \mathbb{R}_+^* \times \mathbb{R}_+^*, \qquad f(x,y) = \frac{x}{y^2} + y^2 + \frac{1}{x}.$$

Justifier que la fonction f admet un minimum local en $(1,1)$. Illustrations par les lignes de niveau :

Editeur

```
import matplotlib.pyplot as plt

def f(x,y):
    return x/y**2+y**2+1/x

x=np.linspace(0.4,1.6,200)
y=np.linspace(0.5,1.5,200)
X , Y = np.meshgrid(x,y)
Z = f(X,Y)

graphe = plt.contour(X,Y,Z,20)
# Le 20 pour 20 lignes de niveau
plt.show()
```



Exercice 6

✦ ✎ Soit f la fonction de classe $\mathscr{C}^2$ définie sur $\mathscr{O} = \mathbb{R}_+^* \times \mathbb{R}$ par :

$$\forall (x, y) \in \mathbb{R}_+^* \times \mathbb{R}, \qquad f(x, y) = x\left((\ln x)^2 + 2y^2\right).$$

1. Démontrer qu'il existe dans $\mathscr{O}$ un unique point col pour f.
2. Est-ce que f admet un maximum global ? un minimum global ?

• Complétons l'étude de la dimension 2 avec *les notations de Monge* (hors programme).
Soit f une fonction de classe $\mathscr{C}^2$ sur un ouvert $\mathscr{O}$ de $\mathbb{R}^2$. Pour $a \in \mathbb{R}^2$, on notera H_a la matrice hessienne de l'application au point a.

$$\mathrm{H}_a = \begin{bmatrix} r & s \\ s & t \end{bmatrix}.$$

Exercice 7

✦✦ ✎ Supposons que $a \in \mathscr{O}$ est un point critique de f. Justifier que

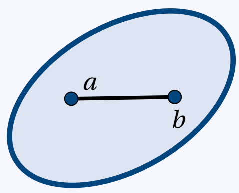
– si $rt - s^2 < 0$, alors $f(a)$ est un point col.
– si $rt - s^2 > 0$, alors $f(a)$ est un extremum local. Dans ce cas, si $r < 0$, alors $f(a)$ est un maximum local et, si $r > 0$, alors $f(a)$ est un minimum local.

4.5 Convexité, condition suffisante d'extremum global

Définition 19 (partie convexe)

Une partie C de $\mathbb{R}^n$ est dite **convexe** si, pour tout couple (a, b) d'éléments de C, le segment $[a; b]$ est tout entier inclus dans C. Autrement dit, C est convexe lorsque

$$\forall\, a, b \in \mathrm{C}, \quad \forall \lambda \in [0, 1], \quad \lambda a + (1 - \lambda) b \in \mathrm{C}.$$



Exemples. $\mathbb{R}^n$, $\mathscr{B} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ sont des parties convexes. Si $(\mathrm{I}_i)_{i \in [\![1;n]\!]}$ désignent des intervalles de $\mathbb{R}$, alors $\mathrm{I}_1 \times \ldots \times \mathrm{I}_n$ est une partie convexe.

Théorème 20 (convexité, extremum global)

Soient $\mathscr{O}$ est un ouvert convexe de $\mathbb{R}^n$, $f : \mathscr{O} \to \mathbb{R}$ et a,un point critique de f.

- **Si** pour tout $x \in \mathscr{O}$, $\mathrm{Sp}\left(\nabla^2 f(x)\right) \subset \mathbb{R}^+$, **alors** f admet un minimum global en a.
- **Si** pour tout $x \in \mathscr{O}$, $\mathrm{Sp}\left(\nabla^2 f(x)\right) \subset \mathbb{R}^-$, **alors** f admet un maximum global en a.

Exercice 8

✦✦ Preuve
On conserve les mêmes notations et hypothèses. À l'aide des résultats préliminaires sur la fonction $g_{a,h}$ et d'une formule de Taylor, justifier que

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \int_0^1 (t-u) q_{a+uh}(h) \, \mathrm{d}u.$$

Conclure sur le théorème.

Exemple. Soit $f(x,y,z) = 11x^2 + 3y^2 + 3z^2 - 10xy + 10xz - 6yz + 1$.
⇀ La fonction f est de classe $\mathscr{C}^2$ sur $\mathbb{R}^3$ et on vérifie que la matrice hessienne en (x,y,z) est

$$\nabla^2 f(x,y,z) = 2 \begin{bmatrix} 11 & -5 & 5 \\ -5 & 3 & -3 \\ 5 & -3 & 3 \end{bmatrix}.$$

⇀ Utilisons Python pour obtenir rapidement le spectre :

Editeur

```
import numpy as np
H= 2*np.array
     ([[11,-5,5],[-5,3,-3],[5,-3,3]])
Sp=np.linalg.eigvals(H)
```

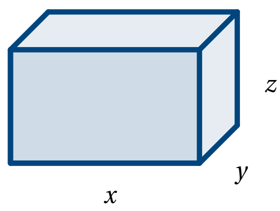
Console

```
>>> # script executed
>>> Sp
array([32.,  2.,  0.])
```

⇀ On vérifie que f admet un point critique donné par $a = (0,1,1)$.
⇀ Concluons : en tout point $(x,y,z) \in \mathbb{R}^3$, la matrice hessienne a un spectre inclus dans $\mathbb{R}^+$. On peut donc en déduire que f admet en a un minimum global.

4.6 Exemple détaillé sur un ouvert

On souhaite minimiser la surface d'un pavé de volume 1. Notons x, y, z les longueurs des cotés. L'aire de la surface a pour expression $2(xy + xz + yz)$. Le volume vaut $xyz = 1$ et $z = 1/(xy)$. Finalement, on cherche à minimiser sur $\mathscr{O} = \mathbb{R}_*^{+2}$, la fonction f définie par



$$f(x,y) = xy + xz + yz = xy + \frac{1}{x} + \frac{1}{y}.$$

Précisons que f n'admet pas de maximum global puisque : $f(x,1) = x + \frac{1}{x} + 1 \underset{x \to +\infty}{\longrightarrow} +\infty$.

Plan d'étude :

- Les dérivées partielles
- Étude du ou des points critiques
- Étude locale au niveau du point critique
- Étude globale
- Une inégalité de convexité

Exercices

Exercice 9. ✧

1. **a)** Préciser l'ensemble de définition de f définie par $f(x,y) = \dfrac{x}{x+y}$.

   b) Est-ce un ouvert, un fermé, ni l'un ni l'autre ?

2. Préciser les lignes de niveau de f.

Exercice 10. ✦ ✎

1. Justifier que l'équation $x - \ln(x) = 2$ d'inconnue $x \in \mathbb{R}_*^+$ a deux solutions α, β avec $\alpha \in]0;1[$ et $\beta \in [2;4]$.
2. On considère maintenant la fonction f de classe $\mathscr{C}^2$ sur l'ouvert $\mathscr{O} =]0,+\infty[^2$ définie par :

$$\forall (x,y) \in]0,+\infty[^2, \qquad f(x,y) = \frac{x^2}{2} - xy - 2x + e^y.$$

   a) Calculer les dérivées partielles d'ordre 1 de f.

   b) Montrer que la fonction f admet exactement deux points critiques : $A = \big(\alpha, \ln(\alpha)\big)$ et $B = \big(\beta, \ln(\beta)\big)$.

   c) Pour tout $\lambda \in \mathbb{R}$, expliciter $\det\big(\nabla^2 f(A) - \lambda I_2\big)$. En déduire la nature local du point critique A.

   d) Même question pour B.

Exercice 11. ✦✦

1. Donner le plus grand ouvert $\mathscr{O}$ telle que l'expression $f(x,y) = x^y - y^x$ soit bien définie.
2. Justifier que f admet un unique point critique et préciser sa nature (extremum/point selle).

Exercice 12. ✦✦ ✎ On définit la partie $\mathscr{B} = \Big\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leqslant 8\Big\}$ et la fonction f sur $\mathbb{R}^2$ par :

$$\forall (x,y) \in \mathbb{R}^2, \qquad f(x,y) = x^4 + y^4 - 2(x-y)^2.$$

1. Déterminer les points critiques de f. Préciser la valeur de f en ces points.
2. **a)** Justifier que pour tout $(x,y) \in \mathbb{R}^2$, $\quad f(x,y) \geqslant \dfrac{1}{2}\Big(x^2 + y^2\Big)^2 - 4\Big(x^2 + y^2\Big)$.

   b) 🔍 En déduire que si $(x,y) \notin \mathscr{B}$, alors $f(x,y) \geqslant 0$.

   c) Justifier que f admet un minimum m_0 sur $\mathscr{B}$.

3. En déduire que m_0 est un minimum global de f. Préciser sa valeur.

Exercice 13. ✦✦✦ *Oraux HEC 2008*

Soit f la fonction définie par :

$$\forall (x;y) \in \mathbb{R}^2, \qquad f(x,y) = x^3 + y^3 - 3xy + 1.$$

1. Calculer les dérivées partielles d'ordre 1 et d'ordre 2 de f.
2. Déterminer les points critiques de f.
3. La fonction f a-t-elle des extrema locaux ? globaux ?

Exercice 14. ✦✦ ✎ *D'après ESCP 01*

Dans cet exercice, n désigne un entier naturel supérieur ou égal à 2 . On note I_n la matrice identité de $\mathscr{M}_n(\mathbb{R})$ et J_n la matrice de $\mathscr{M}_n(\mathbb{R})$ dont tous les éléments valent 1.

1. Déterminer le rang de J_n et en déduire ses valeurs propres. La matrice J_n est-elle diagonalisable ?

   Dans toute la suite, on considère la fonction f_n définie sur $\mathbb{R}^n$ par :

$$f_n(x_1,\ldots,x_n) = \left(\sum_{i=1}^n x_i\right) \exp\left(-\sum_{i=1}^n {x_i}^2\right).$$

2. Montrer que f_n est de classe $\mathscr{C}^2$ sur $\mathbb{R}^n$.
3. Montrer que f_n possède deux points critiques a et $b = -a$, avec a dont les coordonnées sont positives.
4. Justifier que la hessienne de f_n en a est $H_n(a) = \dfrac{-2}{\sqrt{2ne}}(nI_n + J_n)$.
5. Établir que f_n possède un extremum local en a. Quelle est sa nature ? Donner sa valeur. On admet que f_n possède un extremum local de nature et de valeur opposées en b.

6. **a)** Étudier la fonction h définie sur $\mathbb{R}^+$, par : $\forall t \geqslant 0,\ h(t) = t\mathrm{e}^{-t^2}$.

b) Montrer que, pour tout $(x_1, \dots, x_n)$ de $\mathbb{R}^n$,

$$\left(\sum_{k=1}^{n} x_k\right)^2 \leqslant n \sum_{k=1}^{n} {x_k}^2.$$

c) Déduire des deux questions précédentes que f_n admet en a et b des extremums globaux.

Compléments

Exercice 15. ✦✦✦ Soit $f : \mathbb{R}^n \to \mathbb{R}$ de classe $\mathscr{C}^1$. On note $\mathscr{S}$ la sphère unité de $\mathbb{R}^n$ et $\mathscr{B}$ la boule unité ouverte :

$$\mathscr{S} = \left\{x \in \mathbb{R}^n \mid \|x\| = 1\right\}, \qquad \mathscr{B} = \left\{x \in \mathbb{R}^n \mid \|x\| < 1\right\}.$$

On suppose que f est constante sur $\mathscr{S}$. Démontrer l'existence de $a \in \mathscr{B}$ tel que $\nabla f(a) = 0_{\mathbb{R}^n}$.

Exercice 16. ✦ ✎ **Dérivation d'une intégrale à paramètre**
On considère les fonctions f, définie sur $\mathscr{F} = [0, +\infty[\times [0, +\infty[$ et g, définie sur $[0; +\infty[$ par

$$f(x,t) = \mathrm{e}^{-t^2}\sqrt{1+xt} \quad \text{et} \quad g(x) = \int_0^{+\infty} f(x,t)\,\mathrm{d}t.$$

1. **a)** Montrer que pour tout $\alpha \in \mathbb{R}_*^+$, l'intégrale $\displaystyle\int_0^{+\infty} t^{\alpha}\mathrm{e}^{-t^2}\,\mathrm{d}t$ est convergente.

b) En déduire que la fonction g est bien définie sur $[0; +\infty[$.

c) Justifier que g est croissante sur $[0, +\infty[$.

2. **a)** Vérifier que f est de classe $\mathscr{C}^2$ sur $\mathscr{F}$ et préciser $\partial_1 f(x,t)$ et $\partial^2_{1,1} f(x,t)$.

b) Montrer que pour tout $(x,t) \in \mathscr{F}$,

$$\left|\partial^2_{1,1} f(x,t)\right| \leqslant \frac{t^2}{4}\mathrm{e}^{-t^2}.$$

c) Soit $x_0 \in [0; +\infty[$, montrer que, pour $(x,t) \in \mathscr{F}$

$$\left|f(x,t) - f(x_0,t) - (x - x_0)\,\partial_1 f(x_0,t)\right| \leqslant \frac{t^2}{8}\mathrm{e}^{-t^2}\,|x - x_0|^2.$$

d) En déduire l'existence d'une constate $\mathrm{C} \in \mathbb{R}^+$ telle que pour tout $x_0 \in [0; +\infty[$,

$$\left|g(x) - g(x_0) - (x - x_0)\int_0^{+\infty} \partial_1 f(x_0,t)\,\mathrm{d}t\right| \leqslant \mathrm{C}\,|x - x_0|^2.$$

e) Conclure en montrant que g est dérivable sur $[0; +\infty[$ et que g' est définie par

$$\forall x \in [0; +\infty[, \qquad g'(x) = \int_0^{+\infty} \partial_1 f(x,t)\,\mathrm{d}t.$$

f) Retrouver le sens de variation de g.

Exercice 17. ✦✦ ✎ *EML Lyon 2007*
On considère l'application $f : [0; +\infty[\to \mathbb{R}$ définie par

$$f(x) = \begin{cases} \frac{\ln(1+x)}{x} & \text{si } x > 0 \\ 1 & \text{si } x = 0. \end{cases}$$

1. Vérifier que f est de classe $\mathscr{C}^1$ sur $\mathbb{R}^+$.
2. On définit les fonctions $\mathrm{F} : [0; +\infty[\to \mathbb{R}$ et $\mathrm{G} :]0; +\infty[^2 \to \mathbb{R}$ par

$$\mathrm{F}(x) = \int_0^x f(t)\,\mathrm{d}t \quad \text{et} \quad \mathrm{G}(x,y) = \mathrm{F}(xy) - \mathrm{F}(x) - \mathrm{F}(y).$$

a) Montrer que G est de classe $\mathscr{C}^2$ sur $]0; +\infty[^2$, puis exprimer, pour tout $(x,y) \in]0; +\infty[^2$, les dérivées partielles premières et secondes de G en (x,y) en fonction de x, y, $f(x)$, $f(y)$, $f(xy)$, $f'(x)$, $f'(y)$, $f'(xy)$.

b) Établir que G admet $(1,1)$ comme unique point critique.

c) Est-ce que G admet un extremum local ?

Exercice 18. ✦✦ **Maximum de vraisemblance - cas continu avec deux paramètres**
Soient $m \in \mathbb{R}$ et $v \in \mathbb{R}_*^+$ deux paramètres inconnues. Soit X une variable aléatoire suivant une loi normale $\mathscr{N}(m; v)$. On note $f_{m,v}$ la densité continue sur $\mathbb{R}$ de X.
Soient $x_1, \dots, x_n$ des réels fixés. On définit les fonctions L et ℓ sur l'ouvert $\mathbb{R} \times \mathbb{R}_*^+$ par

$$\forall (m,v) \in \mathbb{R} \times \mathbb{R}_*^+, \qquad \mathrm{L}(m,v) = \prod_{k=1}^{n} f_{m,v}(x_k) \quad \text{et} \quad \ell(m,v) = \ln\big(\mathrm{L}(m,v)\big).$$

1. Expliciter $f_{m,v}$, puis $\ell(m,v)$ pour $(m,v) \in \mathbb{R} \times \mathbb{R}_*^+$.

2. La fonction ℓ est de classe $\mathscr{C}^2$, préciser les dérivées partielles d'ordre 1 et 2.

3. En déduire l'unique point critique de ℓ. *On pourra introduire les notations*

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{et} \quad \sigma_x^2 = \frac{1}{n}\left(\sum_{i=1}^{n} (x_i - \overline{x})^2\right).$$

4. Conclure en montrant que L admet un maximum local atteint en un unique point $(\hat{m}, \hat{v})$.

5. Proposer des estimateurs pour m et v en partant de $(\hat{m}, \hat{v})$.

Exercice 19. ✦✦ ✎ Soient $n \in \mathbb{N} \setminus \{0;1\}$ et la fonction f définie sur $\mathbb{R}^n$ par

$$\forall\, (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n, \qquad f(x_1, x_2, \ldots, x_n) = \sum_{k=1}^{n} {x_k}^2 + \left(\sum_{k=1}^{n} x_k\right)^2 - \sum_{k=1}^{n} x_k.$$

1. Montrer que f est de classe $\mathscr{C}^2$ sur $\mathbb{R}^n$ et préciser les dérivées partielles d'ordre 1 et 2 de f sur $\mathbb{R}^n$.

2. **a)** Déterminer le seul point critique $a = (a_1, a_2, \ldots, a_n)$ de f sur $\mathbb{R}^n$.

b) Vérifier que la hessienne de f en ce point s'écrit sous la forme $\mathrm{H} = 2(\mathrm{I}_n + \mathrm{J})$ où J désigne la matrice de $\mathscr{M}_n(\mathbb{R})$ dont tous les coefficients sont égaux à 1.

c) Déterminer un polynôme annulateur de degré 2 de J. Que dire sur les valeurs propres de J ?

d) En déduire que $f(a)$ est un minimum local de f.

3. 🔍 Est-ce que $f(a)$ est un minimum global ?

Exercice 20. ✦✦✦ **Problème des moindres carrés**
Soient $x = (x_i)_{i \in [\![1;n]\!]}$, $y = (y_i)_{i \in [\![1;n]\!]}$. On définit les moyennes empiriques et variance empirique par :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \sigma_x^2 = \frac{1}{n}\left(\sum_{i=1}^{n} (x_i - \overline{x})^2\right) = \frac{1}{n}\left(\sum_{i=1}^{n} {x_i}^2\right) - \overline{x}^2$$

On suppose $\sigma_x \sigma_y \neq 0$ et on définit aussi la covariance et le coefficient de corrélation empiriques de x et y par :

$$\mathrm{Cov}(x,y) = \frac{1}{n}\left(\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})\right) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i y_i\right) - \overline{x} \cdot \overline{y}, \quad \rho(x,y) = \frac{\mathrm{Cov}(x,y)}{\sigma_x \sigma_y}.$$

Soit F la fonction définie sur $\mathbb{R}^2$ par :

$$\forall\, (a,b) \in \mathbb{R}^2, \qquad \mathrm{F}(a,b) = \sum_{i=1}^{n} \left(a x_i + b - y_i\right)^2.$$

1. Justifier que F est de classe $\mathscr{C}^2$ sur $\mathbb{R}^2$ et préciser les dérivées partielles d'ordre 1 et 2, puis vérifier que la matrice hessienne de F peut s'exprimer sous la forme :

$$\forall\, (a,b) \in \mathbb{R}^2, \qquad \nabla^2 \mathrm{F}(a,b) = \frac{2}{n}\begin{bmatrix} \sigma_x^2 + \overline{x}^2 & \overline{x} \\ \overline{x} & 1 \end{bmatrix}.$$

2. **a)** Justifier que F admet un unique point critique $(\hat{a}, \hat{b})$ que l'on exprimera en fonction de $\overline{x}$, $\overline{y}$, σ_x^2 et $\mathrm{Cov}(x,y)$.

b) Vérifier que $\mathrm{F}(\hat{a}, \hat{b}) = n\sigma_y^2 \left(1 - \rho^2(x,y)\right)$.

3. Montrer que $\mathrm{F}(\hat{a}, \hat{b})$ correspond à un minimum global de F.

4. Montrer que l'on a : $|\rho(x,y)| \leqslant 1$. Que dire de x et y lors du cas d'égalité $|\rho(x,y)| = 1$?

Exercice 21. ✦✦✦ **Existence d'une valeur propre pour une matrice symétrique et quotient de Rayleight**
Soit $n \in \mathbb{N}^*$, on note $\langle \cdot, \cdot \rangle$ le produit scalaire canonique sur $\mathbb{R}^n$. On identifie dans la suite, vecteurs de $\mathbb{R}^n$ et matrice de $\mathscr{M}_{n,1}(\mathbb{R})$. Soit S une matrice symétrique de $\mathscr{M}_n(\mathbb{R})$, on pose

$$\mathrm{R_S} : \left\{ \begin{array}{ccc} \mathbb{R}^n \setminus \{0_n\} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \dfrac{\langle \mathrm{S}x, x \rangle}{\|x\|^2}. \end{array} \right.$$

1. **a)** Justifier que la restriction de $\mathrm{R_S}$ à $\mathscr{S} = \left\{x \in \mathbb{R}^n \mid \|x\| = 1\right\}$ admet un minimum et un maximum.

b) 🔍 En remarquant que, pour tous $\alpha \in \mathbb{R}^*$, $x \in \mathbb{R}^n \setminus \{0_n\}$, $\mathrm{R_S}(x) = \mathrm{R_S}(\alpha x)$, montrer que l'application $\mathrm{R_S}$ admet un minimum et un maximum global sur $\mathbb{R}^n \setminus \{0_n\}$.

2. On note $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{R}^n$, et pour tout $x \in \mathbb{R}^n$, $\mathrm{D}(x) = \|x\|^2$ et $\mathrm{N}(x) = \langle \mathrm{S}x, x \rangle$.

a) Vérifier que D et N sont de classe $\mathscr{C}^1$ avec

$$\partial_1 \mathrm{D}(x) = 2\,\langle x, e_1 \rangle \quad \text{et} \quad \partial_1 \mathrm{N}(x) = 2\,\langle \mathrm{S}x, e_1 \rangle.$$

b) En déduire que R_S est de classe $\mathscr{C}^1$ sur $\mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}$ et que le gradient en x est donné par :

$$\forall\, x \in \mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}, \qquad \nabla R_S(x) = \frac{2}{\|x\|^2}\big(Sx - R_S(x)\,x\big).$$

c) Soit x_0 un point où le maximum de R_S est atteint. Montrer que x_0 est vecteur propre de S.

Exercice 22. ✦✦ ✎ Soient l'ouvert $\mathscr{O} = \left(\mathbb{R}_+^*\right)^n$ et la fonction f définie sur $\mathscr{O}$ par

$$f(x_1,\ldots,x_n) = \left(\sum_{k=1}^n x_k\right)\left(\sum_{k=1}^n \frac{1}{x_k}\right).$$

1. Vérifier que f est de classe $\mathscr{C}^2$ sur l'ouvert $\mathscr{O}$. Préciser les dérivées partielles d'ordre 1.

2. 🔧 Déterminer les points critiques de f et donner la valeur de f en ces points.

3. **a)** Montrer que si a est un point critique de f alors il existe $\alpha \in \mathbb{R}_*^+$ tel que :

$$\nabla^2 f(a) = \alpha(nI_n - J)$$

où J la matrice de taille n dont tous les coefficients sont égaux à 1.

b) Déterminer les valeurs propres de $nI - J$. Peut-on en déduire la nature des points critiques de f ?

L'objectif de la suite est de prouver que f admet un minimum global au point critique en utilisant des inégalités plus ou moins classiques.

4. *Première méthode : inégalité de Cauchy-Schwarz*
Expliciter l'inégalité de Cauchy-Schwarz pour le produit scalaire canonique sur $\mathbb{R}^n$. Conclure.

5. *Deuxième méthode : inégalité arithmético-géométrique*
On démontre que pour une fonction concave $f : I \to \mathbb{R}$, on a

$$\forall\,(a_1,a_2,\cdots,a_n) \in I^n,\ \forall\,(\lambda_1,\lambda_2,\cdots,\lambda_n) \in [0;1]^n,\ \text{tels que } \sum_{i=1}^n \lambda_i = 1,\quad f\left(\sum_{k=1}^n \lambda_k a_k\right) \geqslant \sum_{k=1}^n \lambda_k f(a_k).$$

a) À partir de la fonction logarithme, montrer que

$$\forall a_1,a_2,\cdots,a_n \in \mathbb{R}_*^+, \qquad \frac{1}{n}\sum_{i=1}^n a_i \geqslant \sqrt[n]{\prod_{i=1}^n a_i}.$$

b) 🔧 Conclure sur la nature du point critique.

6. *Troisième méthode : inégalité de Tchebychev pour les sommes*

a) Soient $n \in \mathbb{N}^*$ et des réels $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_n$ et $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n$. En étudiant le signe de la somme double $\sum_{i,j\in[\![1;n]\!]}\left(a_i - a_j\right)\left(b_i - b_j\right)$, montrer que

$$\frac{1}{n}\sum_{i=1}^n a_i b_i \leqslant \left(\frac{1}{n}\sum_{i=1}^n a_i\right)\left(\frac{1}{n}\sum_{i=1}^n b_i\right).$$

b) 🔧 Conclure.

CHAPITRE 18

Estimations

Le hasard est le plus grand romancier du monde ; pour être fécond, il n'y a qu'à l'étudier.

HONORÉ DE BALZAC
Avant-propos à la Comédie humaine.

1 Modéliser, estimer, tester, prédire

On considère un phénomène aléatoire et on s'intéresse à une variable aléatoire X qui pourrait le décrire. On suppose que la loi de probabilité de X n'est pas complètement spécifiée et appartient à une famille de lois dépendante d'un paramètre θ appartenant à un ensemble Θ. Le paramètre θ est une quantité inconnue, fixée dans toute l'étude, que l'on cherche à déterminer ou à encadrer.

Le problème de l'*estimation ponctuelle* consiste alors à préciser la valeur du paramètre θ (ou plus généralement d'une fonction $g(\theta)$) à partir d'un échantillon de données $x_1, \ldots, x_n$ obtenues en observant n fois le phénomène. Cette fonction du paramètre représentera en général une valeur caractéristique de la loi inconnue comme son espérance, sa variance, son étendue [1], etc.

Exemples.

- *Exemple 1. Nombre de buts en Ligue 1.*

⇀ Voici le nombre de buts par journée lors de la dernière saison :

26, 34, 29, 31, 28, 25, 32, 30, 25, 26, 29, 29, 29, 30, 21, 29, 29, 27,

27, 15, 26, 35, 30, 23, 23, 22, 21, 19, 23, 32, 38, 30, 26, 35, 30, 36, 30, 37.

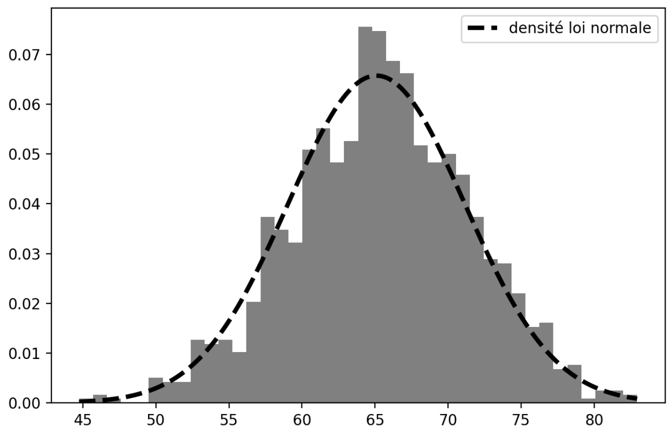
Ces nombres constituent *l'échantillon de données.*
⇀ On pose un *modèle* en supposant que le nombre de buts lors d'une journée est une variable aléatoire X qui suit une loi de Poisson. Notons ce paramètre λ.
⇀ Pour donner une valeur à λ, (on dira *estimer*), on peut utiliser le fait que X admet une espérance $\mathbf{E}(X) = \lambda$. Ainsi, on peut préciser λ en prenant la valeur moyenne du nombre de buts (à savoir ici $\lambda = 2.81$).
⇀ Une fois le modèle posé, on peut le *tester* s'il donne des résultats cohérents et ainsi l'utiliser pour faire de la *prédiction.*

- *Exemple 2. Durée de vie d'un composant électrique.*

- *Exemple 3.*

Un producteur d'œufs de poules souhaite anticiper sa production pour l'année suivante en analysant sa production actuelle. Ci-contre, l'histogramme des différents poids (en grammes) des œufs récoltés dans l'année.

1. Si $X(\Omega) = [a; b]$, l'étendue est définie par la quantité $b - a$.



Console

```
# L représente la liste des
    différents poids

>>> len(L)
51243

>>> np.mean(L) # poids moyen
65.115928

>>> np.std(L)  # écart-type
6.0736541402592824
```

2 Estimation ponctuelle

2.1 Définitions et exemples

Dans la suite, on se fixe :

- Un espace probabilisable $(\Omega, \mathscr{A})$.
- Un espace des paramètres Θ qui est une partie de $\mathbb{R}^n$.
  On suppose de plus que pour chaque paramètre $\theta \in \Theta$, il existe une probabilité $\mathbf{P}_\theta$ définie sur $(\Omega, \mathscr{A})$.
- Une application X qui est bien une variable aléatoire sur tous les espaces probabilisés $(\Omega, \mathscr{A}, \mathbf{P}_\theta)$ lorsque $\theta \in \Theta$.

Notation. Sous réserve d'existence :

- $\mathbf{E}_\theta(X)$ désigne l'espérance de X pour la probabilité $\mathbf{P}_\theta$.
- $\mathbf{V}_\theta(X)$ est la variance de X pour la probabilité $\mathbf{P}_\theta$.

Définition 21 (échantillon)

Soient $X : \Omega \to \mathbb{R}$ une variable aléatoire définie sur l'espace probabilisable $(\Omega, \mathscr{A})$ et $n \in \mathbb{N}^*$. On appelle **n-échantillon** de la loi de X tout vecteur aléatoire $(X_1, \dots, X_n)$ tel que :

- Les applications $X_1, \dots, X_n$ sont des variables aléatoires sur $(\Omega, \mathscr{A})$.
- Les variables $X_1, \dots, X_n$ sont indépendantes relativement à la probabilité $\mathbf{P}_\theta$ et de même loi que X pour tout $\theta \in \Theta$.

Vocabulaire.
- On dit aussi que la loi de X est la loi parente (ou encore loi mère) de l'échantillon.
- On note souvent $(X_1, \dots, X_n)$ *i.i.d* pour signaler que les variables sont indépendantes et identiquement distribuées (c'est-à-dire de même loi).

En pratique, un échantillon de données $x_1, \dots, x_n$ est la réalisation de n variables aléatoires $X_1, \dots, X_n$. L'objectif de l'estimation ponctuelle est alors de préciser le paramètre θ (ou une fonction $g(\theta)$) qui « explique au mieux » les valeurs de l'échantillon.

Définition 22 (estimateur)

- On appelle **estimateur de** θ toute variable aléatoire de la forme $\varphi(X_1, \dots, X_n)$, où $(X_1, \dots, X_n)$ est un n-échantillon et φ, une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$.

- Plus généralement, pour $g : \Theta \to \mathbb{R}$, une fonction, un **estimateur de g**(θ) est une variable aléatoire de la forme $\varphi(X_1, \dots, X_n)$ où $(X_1, \dots, X_n)$ est un n-échantillon.

Remarques.

- Un estimateur ne peut pas dépendre de θ puisque c'est la valeur que l'on souhaite déterminer.
- Estimer ponctuellement $g(\theta)$ par le réel $\varphi(x_1,\dots,x_n)$ où $\varphi(X_1,X_2,\dots,X_n)$ est un estimateur de $g(\theta)$ et $(x_1,\dots,x_n)$ est une réalisation de l'échantillon $(X_1,\dots,X_n)$, c'est décider d'accorder à $g(\theta)$ la valeur $\varphi(x_1,\dots,x_n)$.

Exemples.

- Soit $(X_1,\dots,X_n)$ un n-échantillon de la loi de X. Il est souvent utile de considérer **l'estimateur de la moyenne empirique** donné par :

$$\overline{X_n} = \frac{X_1+X_2+\cdots+X_n}{n} \qquad (\bullet)$$

Mais on peut inventer toute une gamme d'estimateurs :

$$T_n = \max_{1\leqslant k\leqslant n} X_k, \quad U_n = \min_{1\leqslant k\leqslant n} X_k, \quad V_n = \ln(1+|U_n|), \quad W_n = \max\left\{k\in[\![1;n]\!] \mid X_k = T_n\right\}, \quad Y_n = 1, \quad \text{etc.}$$

- Créons un échantillon d'une loi uniforme $[0;\theta]$.

On peut essayer de retrouver la valeur de θ à partir de deux estimateurs

$$2\overline{X_n} \quad \text{et} \quad T_n = \max_{k\in[\![1;n]\!]} X_k.$$

On obtient deux estimations de θ par :

```
>>> 2*sum(Ech)/500
0.4386345472617729
>>> max(Ech)
0.4289769082767542
```

Editeur

```
# Fixons une valeurs au paramètre "inconnu"
theta=0.43

# création de l'échantillon
n=500 # taille de l'échantillon
Ech=theta*np.random.rand(n)


np.round(Ech,2)
# pour n'afficher que les 2 premières décimales
array([0.18, 0.25, 0.28, 0.26, 0.05, 0.14, 0.15, 0.02,
    0.15, 0.35, 0.2 ,
        0.25, 0.41, 0.17, 0.39, 0.32, 0.22, 0.21, 0.12,
            0.14, 0.35, 0.18,
        0.16, 0.23, 0.1 , 0.19, 0.28, 0.3 , 0.31, 0.23, 0.4
            , 0.08, 0.18 ...
```

2.2 Biais, convergence et comparaison des estimateurs

Définition 23 (biais d'un estimateur)

Soit T_n un estimateur de $g(\theta)$ tel que tout $\theta\in\Theta$, T_n admet une espérance pour la probabilité $\mathbf{P}_\theta$.

- On définit le **biais** de T_n en $g(\theta)$ par $b_\theta(T_n) = \mathbf{E}_\theta(T_n) - g(\theta)$.
- Si pour tout $\theta\in\Theta$, $b_\theta(T_n) = 0$, on dit l'estimateur est **sans biais**. Sinon, on dit que l'estimateur est biaisé.

Exemple. Avec les notations de la définition et si X admet une espérance θ, l'estimateur de la moyenne empirique $\overline{X_n}$ est un estimateur sans biais de θ.

Remarque. On peut donner une définition moins contraignante : un estimateur est **asymptotiquement sans biais** si $b_\theta(T_n) \underset{n\to\infty}{\longrightarrow} 0$. Par exemple, la variable M_n définie précédemment est un estimateur asymptotiquement sans biais.

Exercice 23

✦✦ ✎ **Estimateur de la variance**

Soit la variable X admet un espérance μ et une variance σ^2. On pose

$$T_n = \frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2 \quad \text{et} \quad V_n = \frac{1}{n}\sum_{k=1}^{n}\left(X_k-\overline{X_n}\right)^2.$$

1. On suppose, dans cette question, que μ connu et on cherche à estimer σ^2 qui est donc inconnue. Montrer que T_n est un estimateur sans biais de σ^2.
2. On suppose maintenant que μ est aussi inconnu.
   a) Montrer que V_n est un estimateur asymptotiquement sans biais de σ^2 et calculer le biais de cet estimateur.
   b) Construire, à partir de V_n, un estimateur sans biais de σ^2.

Définition 24 (estimateur convergent)

Soit $(T_n)_{n\in\mathbb{N}}$, une suite d'estimateurs de $g(\theta)$.
On dit que la suite $(T_n)_{n\in\mathbb{N}}$ est **convergente** si pour tout $\theta\in\Theta$, la suite $(T_n)_{n\in\mathbb{N}}$ converge en probabilité vers la variable aléatoire presque sûrement constante $g(\theta)$. Autrement dit

$$\forall\,\theta\in\Theta,\quad \forall\varepsilon\in\mathbb{R}_*^+,\qquad \mathbf{P}_\theta\left(\left|T_n-g(\theta)\right|\geqslant\varepsilon\right)\underset{n\to\infty}{\longrightarrow}0.$$

Vocabulaire. Par abus de langage, on dit simplement que T_n est un estimateur convergent de $g(\theta)$.

Exercice 24

✦ **Exemples**

1. Considérons X une variable aléatoire dont la loi est uniforme sur $[0;\theta]$, où le paramètre θ est inconnu. Vérifier que l'estimateur $T_n=2\overline{X}_n$ est convergent de θ.
2. Soit X une variable à densité donnée pour $\theta\in\mathbb{R}_*^+$ par

$$\forall\,x\in\mathbb{R},\qquad f_\theta(x)=\frac{2}{\theta}\left(1-\frac{x}{\theta}\right)\mathbf{1}_{[0;\theta]}(x).$$

   a) Pour un échantillon $(X_1,\ldots,X_n)$ de même loi que X, montrer que $3\overline{X_n}$ est un estimateur sans biais de θ.
   b) Vérifier que cet estimateur est convergent de θ.

Remarque. La notion de convergence des estimateurs ne donne aucune assurance pratique que la valeur prise par un estimateur à partir de l'échantillon de données sera assez proche de la vraie valeur du paramètre. On quantifie la qualité des estimateurs par la notion de **risque quadratique**. Cette notion est maintenant hors-programme mais on pourra consulter l'exercice 32, p.25, pour des précisions.

Proposition 25 (composition et estimateurs)

Soit $(T_n)_{n\in\mathbb{N}}$, une suite d'estimateurs de $g(\theta)$.

Si
- La suite $(T_n)_{n\in\mathbb{N}}$ est une suite d'estimateurs convergente de $g(\theta)$.
- La fonction f est continue sur $\mathbb{R}$.

Alors $\left(f(T_n)\right)_{n\in\mathbb{N}}$ est une suite d'estimateurs convergente de $f\left(g(\theta)\right)$.

Exercice 25

✦ ✎ Soit $(X_1,X_2,\ldots,X_n)$ un échantillon de la loi normale $\mathcal{N}(0,\sigma^2)$.

1. Pour tout $i\in[\![1;n]\!]$, calculer l'espérance de la variable aléatoire $|X_i|$.
2. En déduire un estimateur sans biais et convergent de σ.

Exercice 26

✦✦ **Biais et composition**

Soit $(X_1,X_2,\ldots,X_n)$ un échantillon de la loi de Poisson de paramètre λ inconnu. On sait que la moyenne empirique $\overline{X}_n=\sum_{k=1}^{n}X_k/n$ est un estimateur sans biais et convergent de λ. On cherche à estimer $e^{-\lambda}$.
Est-ce que l'estimateur $T_n=e^{-\overline{X}_n}$ est un estimateur sans biais de $e^{-\lambda}$? asymptotiquement sans biais ? convergent ?

Proposition 26 (condition suffisante de convergence)

Soit $(T_n)_{n\in\mathbb{N}}$, une suite d'estimateurs de $g(\theta)$.

Si $\quad\mathbf{E}(T_n)\underset{n\to\infty}{\longrightarrow}g(\theta)\quad$ et $\quad\mathbf{V}(T_n)\underset{n\to\infty}{\longrightarrow}0.$

Alors $(T_n)_{n\in\mathbb{N}}$ est une suite d'estimateurs convergente de $g(\theta)$.

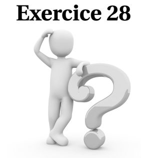
Exercice 27

✦ Prouver cet énoncé à partir de l'inégalité de Markov.

Remarque. On peut dire qu'un estimateur sans biais Y_n est meilleur qu'un autre estimateur sans biais Z_n si $\mathbf{V}(Y_n) \leqslant \mathbf{V}(Z_n)$ pour tout $n \in \mathbb{N}$. Le fait d'être meilleur se matérialise dans l'inégalité de Bienaymé-Tchebychev :

$$\forall \varepsilon \in \mathbb{R}_*^+, \qquad \mathbf{P}_\theta\left(|T_n - g(\theta)| \geqslant \varepsilon\right) \leqslant \frac{\mathbf{V}(Y_n)}{\varepsilon^2} \leqslant \frac{\mathbf{V}(Z_n)}{\varepsilon^2}.$$

Exercice 28



✧ ✎ Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes et suivant toutes la loi $\mathscr{B}(p)$. Le paramètre p est inconnu, on cherche à l'estimer. Pour cela, on pose

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad T_n = \frac{2}{n(n+1)}\sum_{i=1}^{n} i X_i.$$

1. Montrer que $\overline{X_n}$ et T_n sont deux estimateurs sans biais de p.
2. Calculer et comparer les variances de $\overline{X_n}$ et de T_n.
3. Montrer que $\overline{X_n}$ et T_n sont deux estimateurs convergents de p.

Comparaison des estimateurs sans biais - simulation Python

Reprenons le cas d'une variable X suivant une loi uniforme sur $[0;\theta]$ et des deux estimateurs sans biais :

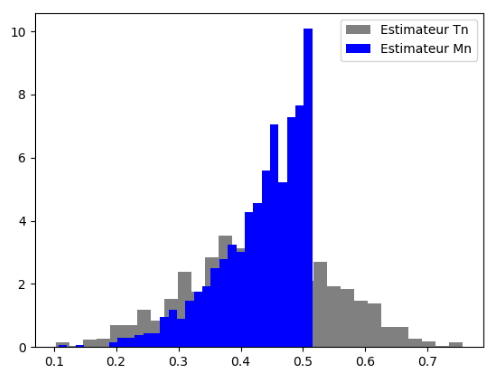
$$\overline{X_n} = \frac{2}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad \widetilde{M_n} = \frac{n+1}{n}\max(X_1,\ldots,X_n).$$

Simulons un grand nombre d'échantillons de données et affichons les réalisations de ces deux estimateurs à l'aide des histogrammes.

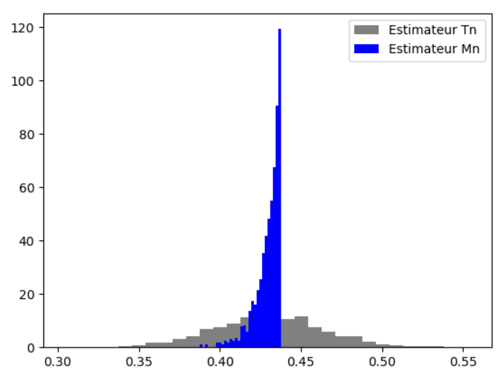
Editeur

```
theta=0.4292081919392057
# choix du paramètre "inconnu"

def estimateurT(n):
    E=theta*np.random.rand(n)
    return 2*np.sum(E)/n

def estimateurM(n):
    return ((n+1)/n)*max(theta*np.random.rand
        (n))

# affichage des histogrammes

def histogrammes(n):
    LM=np.zeros(1000)
    LT=np.zeros(1000)
    for i in range(1000):
        LM[i]=estimateurM(n)
        LT[i]=estimateurT(n)

    plt.hist(LM,30)
    plt.hist(LT,30)
    plt.show()
```

```
>>> histogrammes(5)
```




```
>>> histogrammes(50)
```



Le meilleur estimateur sans biais est celui qui a la variance la plus faible. En effet, c'est celui où le risque que l'échantillon de données donnent une valeur loin de l'espérance (qui est dans ce cas le paramètre à estimer) est le plus faible.

Exemple : Un exemple détaillé : questions sensibles lors d'un sondage d'opinion

3 Estimation par intervalle de confiance

S'il existe des critères pour juger des qualités d'un estimateur ponctuel T_n de $g(\theta)$, aucune certitude ne peut être apportée quant au fait que l'estimation donnée par l'échantillon de données soit une « bonne » valeur du paramètre $g(\theta)$. L'estimation par intervalle de confiance permet de trouver un intervalle aléatoire qui contienne $g(\theta)$ avec une probabilité minimale donnée. Dans tout ce paragraphe, $(U_n)_{n\in\mathbb{N}^*}$ et $(V_n)_{n\in\mathbb{N}^*}$ désigneront deux suites d'estimateurs de $g(\theta)$ telles que pour tous $\theta \in \Theta$ et $n \in \mathbb{N}^*$,

$$\mathbf{P}_\theta\big([U_n \leqslant V_n]\big) = 1.$$

3.1 Intervalle de confiance

Définition 27 (intervalle de confiance)

Soient $\alpha \in]0,1[$, U_n et V_n deux estimateurs de $g(\theta)$ tels que pour tout $\theta \in \Theta$

$$\mathbf{P}_\theta\Big(\ g(\theta) \in [U_n ; V_n]\ \Big) \geqslant 1-\alpha.$$

On dit que l'intervalle $[U_n, V_n]$ est un **intervalle de confiance** de $g(\theta)$ avec un risque d'au plus α ou au niveau de confiance au moins égal à $1-\alpha$.

Remarque. En pratique, on part d'un échantillon de données $x_1, x_2, \ldots, x_n$. On calcule les valeurs

$$u_n = U_n(x_1, x_2, \ldots, x_n) \quad \text{et} \quad v_n = V_n(x_1, x_2, \ldots, x_n).$$

Ainsi on construit un intervalle aléatoire $[u_n, v_n]$ dans lequel $g(\theta)$ à une probabilité supérieure à $1-\alpha$ de s'y trouver.

Remarque. Dans les conditions usuelles, on considère des niveaux de confiance de 95% ou 99% (soit $\alpha = 0,05$ ou $\alpha = 0,01$).

Estimation par intervalle de confiance en utilisant l'inégalité de Bienaymé-Tchebychev

Méthode

Comment obtenir un intervalle de confiance à partir de l'inégalité de Bienaymé-Tchebychev?

Soit T_n, une variable aléatoire admettant une variance, l'inégalité s'écrit

$$\forall \varepsilon \in \mathbb{R}_*^+, \qquad \mathbf{P}_\theta\Big(|T_n - \mathbf{E}_\theta(T_n)| \geqslant \varepsilon\Big) \leqslant \frac{\mathbf{V}_\theta(T_n)}{\varepsilon^2}.$$

Comme $\Big[\,|T_n - \mathbf{E}_\theta(T_n)| < \varepsilon\Big] \subset \Big[\,|T_n - \mathbf{E}_\theta(T_n)| \leqslant \varepsilon\Big]$, on obtient par passage au complémentaire

$$\mathbf{P}_\theta\Big(|T_n - \mathbf{E}_\theta(T_n)| \leqslant \varepsilon\Big) \geqslant 1 - \frac{\mathbf{V}_\theta(T_n)}{\varepsilon^2}.$$

Si T_n est un estimateur sans biais de $g(\theta)$ (c'est-à-dire $\mathbf{E}_\theta(T_n) = g(\theta)$) et si on peut trouver un entier n tel que $\frac{\mathbf{V}_\theta(T_n)}{\varepsilon^2} \leqslant \alpha$, on peut récrire l'inégalité

$$\mathbf{P}_\theta\big(T_n - \varepsilon \leqslant g(\theta) \leqslant T_n + \varepsilon\big) \geqslant 1-\alpha.$$

L'intervalle de confiance est alors $[T_n - \varepsilon ; T_n + \varepsilon]$, il est de longueur 2ε.

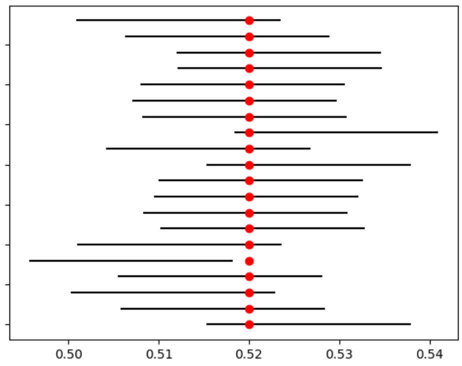
Exemple. *Estimation du paramètre p d'une loi de Bernoulli.*
On trouve un intervalle de confiance de p à un niveau de confiance $1-\alpha$ avec

$$\left[\overline{X}_n - \frac{1}{2\sqrt{n\alpha}}, \overline{X}_n + \frac{1}{2\sqrt{n\alpha}}\right].$$

Simulation Python. Le code suivant crée plusieurs séries de données et construit l'intervalle aléatoire associé à chaque série de données.

Editeur

```
p=0.52
# le paramètre "inconnu" à estimer

n=10000 # Nombre de sondés
alpha=0.20 # niveau de risque

NbreTest=20
for i in range(NbreTest):
    # Création de l'échantillon de données
    Ech=np.random.rand(n)<p
    # Calcul des bords de l'intervalle
    u=np.mean(Ech)-1/(2*(n*alpha)**(1/2))
    v=np.mean(Ech)+1/(2*(n*alpha)**(1/2))
    plt.plot([u,v],[i,i])
    plt.plot([p],[i],'o')
plt.show()
```



On constate que dans la majorité des cas, le paramètre inconnu est bien dans l'intervalle aléatoire construit à partir de l'échantillon.

Estimation par intervalle de confiance de la moyenne d'une loi normale d'écart-type connu

Soit $(X_1, X_2, \dots, X_n)$ un n-échantillon issu d'une loi normale $\mathcal{N}(\mu;\sigma^2)$. On suppose σ connu mais l'espérance μ est inconnue et on cherche à l'estimer. On considère $\overline{X_n}$ la moyenne empirique de l'échantillon. D'après la loi faible des grands nombres, nous savons que c'est un estimateur sans biais et convergent de μ. De plus, par les règles de stabilité par combinaisons linéaires des lois normales indépendantes :

$$\overline{X}_n \hookrightarrow \mathcal{N}\left(\mu;\frac{\sigma^2}{n}\right) \quad \text{et} \quad Y_n = \sqrt{n}\cdot\frac{\overline{X}_n-\mu}{\sigma} \hookrightarrow \mathcal{N}(0;1).$$

Notons Φ, la fonction de répartition de la loi normale centrée réduite et posons t_α positif tel que $t_\alpha = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$. Avec ce choix

$$\mathbf{P}(-t_\alpha \leqslant Y_n \leqslant t_\alpha) = \Phi(t_\alpha) - \Phi(-t_\alpha) = 2\Phi(t_\alpha) - 1 = 1-\alpha \qquad \text{car } \Phi(t_\alpha) + \Phi(-t_\alpha) = 1.$$

En revenant à $\overline{X}_n$, on trouve

$$\mathbf{P}\left(\overline{X}_n - t_\alpha\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \overline{X}_n + t_\alpha\frac{\sigma}{\sqrt{n}}\right) = \mathbf{P}\left(-t_\alpha \leqslant \frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}} \leqslant t_\alpha\right) = \mathbf{P}(-t_\alpha \leqslant Y_n \leqslant t_\alpha) = 1-\alpha$$

On vient de montrer que

$$\left[\overline{X}_n - t_\alpha\frac{\sigma}{\sqrt{n}}\,;\,\overline{X}_n + t_\alpha\frac{\sigma}{\sqrt{n}}\right]$$

est un intervalle de confiance de μ avec un niveau de confiance égal à $1-\alpha$.

3.2 Intervalle de confiance asymptotique

Définition 28 (intervalle de confiance asymptotique)

Soient $\alpha \in]0;1[$, U_n et V_n deux estimateurs de $g(\theta)$ tels que pour tout $\theta \in \Theta$

$$\mathbf{P}_\theta\Big(g(\theta) \in [U_n;V_n]\Big) \geqslant 1-\alpha_n \quad \text{et} \quad \alpha_n \underset{n\to\infty}{\longrightarrow} \alpha.$$

On dit que l'intervalle $[U_n;V_n]$ est un **intervalle de confiance asymptotique** de $g(\theta)$ avec un risque d'au plus α ou au niveau de confiance au moins égal à $1-\alpha$.

Intervalle de confiance asymptotique du paramètre d'une loi de Bernoulli

Soit $(X_1, X_2, \dots, X_n)$ un n-échantillon issu d'une loi de Bernoulli de paramètre p. Par indépendance

$$n\overline{X_n} = \sum_{i=1}^{n} X_i \hookrightarrow \mathcal{B}(n;p).$$

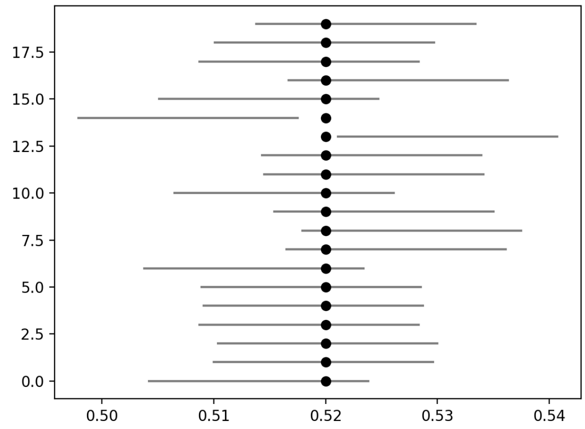
Par le théorème central limite, on montre qu'un intervalle de confiance asymptotique de p au niveau de confiance $1-\alpha$ est

$$\left[\overline{X_n} - \frac{t_\alpha}{2\sqrt{n}}\,;\,\overline{X_n} + \frac{t_\alpha}{2\sqrt{n}}\right].$$

Simulation Python. Le code suivant crée plusieurs séries de données et construit l'intervalle aléatoire obtenu par l'intervalle de confiance asymptotique associé à chaque série de données.

Editeur

```
p=0.52 # le paramètre "inconnu" à estimer

n=10000 # Nombre de sondés
alpha=0.05; talpha=1.96

NbreTest=20
for i in range(NbreTest):
    Ech=np.random.rand(n)<p
    u=np.mean(Ech)-talpha/(2*(n)**(1/2))
    v=np.mean(Ech)+talpha/(2*(n)**(1/2))
    plt.plot([u,v],[i,i])
    plt.plot([p],[i],'o')
plt.show()
```



Remarque. Les intervalles obtenus sont bien plus précis que ceux obtenus par l'inégalité de Bienaymé-Tchebychev. En effet, les longueurs des intervalles obtenus par le théorème central limite sont plus petites que celles obtenues par l'inégalité de Bienaymé-Tchebychev. Par contre, on s'attend à ce que la probabilité que le paramètre ne soit pas dans l'intervalle construit à partir des données (c'est le cas ici du sixième test en partant du haut) soit plus grande.

Exercices

Estimation ponctuelle

Exercice 29. ✧ **Estimation d'une variance**
Soient $n \in \mathbb{N}^*$, $p \in]0;1[$ et une variable aléatoire $X \hookrightarrow \mathscr{B}(n;p)$. Démontrer qu'il existe deux réels $\alpha_n, \beta_n \in \mathbb{N}^*$ (indépendants de p) tels que la variable aléatoire $\alpha_n X + \beta_n X^2$ soit un estimateur sans biais de $\mathbf{V}(X)$.

Exercice 30. ✦ ✎ Soient $\theta \in]0,1[$ un paramètre inconnu et X une variable aléatoire de loi définie par

$$\forall\, k \in \mathbb{N}, \qquad \mathbf{P}(X = k) = (k+1)(1-\theta)^2\theta^k.$$

1. Justifier que X admet un moment d'ordre 1 et 2. Préciser l'espérance $\mathbf{E}(X)$.

Soit $(X_1,\ldots,X_n)$ un échantillon de même loi que X. On pose $\quad S_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i.$

2. Expliciter une fonction bijective $g :]0;1[\to \mathbb{R}_*^+$ telle que $\mathbf{E}(S_n) = g(\theta)$.

3. En déduire que $g^{-1}(S_n)$ est un estimateur convergent de θ.

Exercice 31. ✦ ✎ **Estimation d'un paramètre d'une loi uniforme**
Soient $n \in \mathbb{N}^*$ et $\theta \in \mathbb{R}_*^+$. Soient X une variable aléatoire suivant une loi uniforme sur $[0;\theta[$ et $(X_1,\ldots,X_n)$ un échantillon de même loi que X.

1. Montrer que pour tout $n \in \mathbb{N}^*, T_n = \frac{2}{n}(X_1 + \cdots + X_n)$ est un estimateur sans biais de θ.

2. Considérons $Y_{\min}$ et $Y_{\max}$ définies par :

$$Y_{\min} = \min_{i \in [\![1;n]\!]} X_i \quad \text{et} \quad Y_{\max} = \max_{i \in [\![1;n]\!]} X_i.$$

Pour chacune de ces deux variables, préciser la fonction de répartition, une densité de probabilité, l'espérance et la variance.
On pourra remarquer que $Y_{\min}$ *et* $\theta - Y_{\max}$ *ont même loi.*

3. Posons $T'_n = \frac{n+1}{n} Y_{\max}$. Justifier que T'_n est un estimateur sans biais de θ.

4. Quel est le meilleur estimateur de θ entre T'_n et T_n ?
Un estimateur sans biais est d'autant meilleur que sa variance est faible.

5. Posons $T''_n = Y_{\min} + Y_{\max}$. Montrer sans calculs superflus que $\mathbf{V}\left(T''_n\right) \leqslant 4\,\mathbf{V}(Y_{\max})$.

6. En déduire que T''_n est un estimateur sans biais de θ meilleur que T_n pour n suffisamment grand.

7. Est-ce que les estimateurs T_n, T'_n et T''_n sont convergents ?

Exercice 32. ✦ ✎ **Risque quadratique**
Soit T_n un estimateur de $g(\theta)$. On suppose que pour tout $\theta \in \Theta, T_n$ admet un moment d'ordre 2 pour la probabilité $\mathbf{P}_\theta$. Dans ce cas, on appelle risque quadratique de T_n en $g(\theta)$ et on note $r_\theta(T_n)$ le réel :

$$r_\theta(T_n) = \mathbf{E}_\theta\left(\left(T_n - g(\theta)\right)^2\right).$$

1. Justifier que $r_\theta(T_n) = \left(b_\theta(T_n)\right)^2 + \mathbf{V}_\theta(T_n)$.

2. **a)** Justifier que si le risque quadratique de T_n tend vers 0 quand n tend vers l'infini, alors T_n est un estimateur convergent de $g(\theta)$.

b) En déduire que, si T_n est un estimateur asymptotiquement sans biais de $g(\theta)$ (c'est-à-dire $\mathbf{E}(T_n) \underset{n\to\infty}{\longrightarrow} g(\theta)$) et si la variance de T_n tend vers 0 lorsque n tend vers l'infini, alors T_n est un estimateur convergent de $g(\theta)$.

3. Pour comparer deux estimateurs T_n et U_n de $g(\theta)$, on peut calculer leur risque quadratique. Si on a :

$$\forall\, \theta \in \Theta, \qquad r_\theta(T_n) \leqslant r_\theta(U_n),$$

alors on dira que T_n est un meilleur estimateur de $g(\theta)$ que U_n.

a) *Exemple 1.*
Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes suivant la loi exponentielle $\mathscr{E}(\lambda)$. On pose

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad T_n = \frac{1}{n+1}\sum_{i=1}^{n} X_i.$$

Comparer ces deux estimateurs de $\theta = 1/\lambda$.

b) *Exemple 2.*
Soient $T_{n,0}$ et $T_{n,1}$ deux estimateurs de θ, sans biais et indépendants. Pour tout a réel, on pose

$$T_{n,a} = aT_{n,1} + (1-a)T_{n,0}.$$

i) Est-ce que $T_{n,a}$ est un estimateur sans biais de θ ?

ii) Pour quelle valeur de a, l'estimateur est-il le meilleur ?

Exercice 33. ✦✦ **Estimateur sans biais de l'écart-type σ d'une loi normale centrée** *extrait HEC 2005*
Soit X une variable aléatoire qui suit une loi normale centrée et d'écart-type σ, le paramètre réel inconnu σ, est strictement positif.

1. Montrer que la variable aléatoire $T = \frac{X^2}{2\sigma^2}$ suit une loi γ de paramètre $1/2$. En déduire la valeur de $\Gamma(1/2)$.

2. Pour n entier naturel non nul, on considère un n-échantillon $(X_1, X_2, \dots, X_n)$ constitué de variables indépendantes et de même loi que X.

   a) On désigne par S_n la variable aléatoire $S_n = \sum\limits_{i=1}^{n} \frac{{X_i}^2}{2\sigma^2}$. Quelle est la loi de probabilité de S_n ?

   b) En déduire que la variable aléatoire Y_n définie par $Y_n = \frac{1}{n}\sum\limits_{i=1}^{n} {X_i}^2$ est un estimateur sans biais de σ^2.

Exercice 34. ✦✦ ✎ **Estimation des paramètres d'une loi de Pareto**
Soient $(b,\theta) \in \mathbb{R}^{+*} \times \mathbb{R}$. On suppose que X est une va à densité qui prend ses valeurs dans $[\theta; +\infty[$, dont une densité f est définie par :

$$\forall x \in \mathbb{R}, \quad f(x) = \begin{cases} 0 & \text{si } x < \theta \\ \frac{1}{b}\exp\left(-\frac{x-\theta}{b}\right) & \text{si } x \geqslant \theta. \end{cases}$$

Soient $n \in \mathbb{N} \setminus \{0;1\}$ et $(X_1, \dots, X_n)$ un n-échantillon de même loi que X. On pose :

$$\overline{X_n} = \frac{X_1 + \cdots + X_n}{n} \quad \text{et} \quad T_n = \min(X_1, \dots, X_n).$$

1. Soit $k \in [\![1;n]\!]$. Reconnaitre la loi de $Y_k = (X_k - \theta)/b$.

2. On pose $\overline{Y_n} = \frac{Y_1 + \cdots + Y_n}{n}$ et $U_n = \min(Y_1, \dots, Y_n)$. Calculer $\mathbf{E}\left(\overline{Y_n}\right)$ et $\mathbf{E}(U_n)$.

3. Exprimer les variables aléatoires $\overline{X_n}$ et T_n en fonction des variables $\overline{Y_n}$ et U_n. En déduire $\mathbf{E}\left(\overline{X_n}\right)$ et $\mathbf{E}(T_n)$.

4. Déterminer un estimateur sans biais $\widehat{\theta}_n$ de θ et un estimateur sans biais $\widehat{b}_n$ de b sous la forme de combinaisons linéaires de $\overline{X_n}$ et T_n.

Exercice 35. ✦✦ **Estimation et loi de Poisson** *d'après oral ESCP 2022, sujet 35*
Soit $X_1, \dots, X_n$ un n-échantillon de variables aléatoires indépendantes qui suivent la loi de Poisson de paramètre $\lambda \in \mathbb{R}_*^+$ inconnu. On cherche dans cet exercice à estimer $e^{-\lambda}$.
Pour tout $k \in [\![1,n]\!]$, on note Y_k la fonction indicatrice de l'événement $[X_k = 0]$. Pour tout $n \in \mathbb{N}^*$, on pose :

$$\overline{Y_n} = \frac{1}{n}\sum_{k=1}^{n} Y_k \quad \text{et} \quad S_n = \sum_{k=1}^{n} X_k$$

1. **a)** Déterminer la loi de Y_k.

   b) Montrer que $\overline{Y_n}$ est un estimateur sans biais de $e^{-\lambda}$.

   c) $\overline{Y_n}$ est-il un estimateur convergent de $e^{-\lambda}$?

2. Pour $j \in \mathbb{N}$, on pose $\varphi(j) = \mathbf{P}_{[S_n=j]}(X_1 = 0)$. Calculer $\varphi(j)$.

3. On pose à présent $T_n = \varphi(S_n)$.

   a) Montrer que T_n est un estimateur sans biais de $e^{-\lambda}$.

   b) T_n est-il un estimateur convergent de $e^{-\lambda}$?

4. Comparer les variances des deux estimateurs T_n et $\overline{Y}_n$.

Estimations par intervalles de confiance

Exercice 36. ✦✦ **Calcul d'un intervalle de confiance**
Soit $(X_1, \dots, X_n)$ un échantillon de la loi exponentielle de paramètre $\lambda > 0$ inconnu. On pose

$$Y_n = \min(X_1, \dots, X_n).$$

1. Montrer que Y_n suit une loi exponentielle dont on précisera le paramètre.

2. Soit $\alpha \in]0; 2/3[$. Déterminer deux réels a_n et b_n tels que

$$\mathbf{P}(n\lambda Y_n \leqslant a_n) = \frac{\alpha}{2} = \mathbf{P}(n\lambda Y_n \geqslant b_n).$$

3. Justifier que $\left[\frac{n\mathrm{Y}_n}{a_n};\frac{n\mathrm{Y}_n}{b_n}\right]$ est un intervalle de confiance de $\frac{1}{\lambda}$ au niveau de confiance d'au moins $1-\alpha$.

Exercice 37. ✦✦ ✎ **Calcul d'un intervalle de confiance II**
Soit $(\mathrm{X}_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes et de même loi exponentielle $\mathscr{E}(\lambda)$ où le paramètre $\lambda>0$ est inconnu. On pose $\overline{\mathrm{X}_n}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{X}_i$.

1. Justifier que $\lambda\sqrt{n}\overline{\mathrm{X}_n}-\sqrt{n}$ converge en loi vers une variable aléatoire de loi $\mathscr{N}(0,1)$.

2. Soit $\alpha\in\,]0,1[$. Notons t_α l'unique réel tel que $\Phi(t_\alpha)=1-\frac{\alpha}{2}$. Montrer que

$$\left[\left(1-\frac{t_\alpha}{\sqrt{n}}\right)\frac{1}{\overline{\mathrm{X}_n}},\left(1+\frac{t_\alpha}{\sqrt{n}}\right)\frac{1}{\overline{\mathrm{X}_n}}\right]$$

est un intervalle de confiance asymptotique de λ au niveau de risque α.

Lien entre estimation et optimisation

Exercice 38. ✦
Soit $(\mathrm{X}_1,\mathrm{X}_2,\dots,\mathrm{X}_n,\mathrm{Y}_1,\mathrm{Y}_2,\dots,\mathrm{Y}_m)$ un échantillon de variables aléatoires indépendantes, de même loi de Bernoulli de paramètre inconnu $p\in\,]0;1[$. On pose :

$$\overline{\mathrm{X}_n}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{X}_i,\quad \overline{\mathrm{Y}_m}=\frac{1}{m}\sum_{i=1}^{m}\mathrm{Y}_i\quad\text{et}\quad \mathrm{Z}=a\overline{\mathrm{X}_n}+b\overline{\mathrm{Y}_m},\quad\text{où}\quad (a,b)\in\mathbb{R}^2.$$

Pour quelles valeurs de a et b, Z est-il le meilleur estimateur sans biais de p ?

Exercice 39. ✦ **Estimation et optimisation sous contrainte**
Soient $n\in\mathbb{N}^*$ et $\theta\in\mathbb{R}_*^+$. Soient X une variable aléatoire d'espérance $\mathbf{E}(\mathrm{X})=\theta\neq0$, de variance $\mathbf{V}(\mathrm{X})=1$ et $(\mathrm{X}_1,\dots,\mathrm{X}_n)$ un échantillon de même loi que X. On pose $\overline{\mathrm{X}_n}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{X}_i$.

1. Montrer que $\overline{\mathrm{X}_n}$ est un estimateur sans biais de θ et préciser sa variance.

2. Soit $(\alpha_1,\dots,\alpha_n)\in\mathbb{R}^n$. On note $\mathrm{Y}_n=\sum_{i=1}^{n}\alpha_i\mathrm{X}_i$.

- **a)** Donner une condition nécessaire et suffisante sur $\alpha_1,\dots,\alpha_n$ pour que Y_n soit un estimateur sans biais de θ. On suppose dans la suite que cette condition est vérifiée.
- **b)** Calculer $\mathrm{Cov}\left(\overline{\mathrm{X}_n},\mathrm{Y}_n\right)$. En déduire $\mathbf{V}\left(\overline{\mathrm{X}_n}\right)\leqslant\mathbf{V}(\mathrm{Y}_n)$.
- **c)** Donner un choix de paramètres $\alpha_1,\dots,\alpha_n$ afin que la variance de Y_n soit minimale ?

Exercice 40. ✦ ✎ **Maximum de vraisemblance - cas discret**

Soit $\theta\in\mathbb{R}_*^+$. Pour tout k de $\mathbb{N}$, on pose : $\qquad p_k=\alpha\left(\frac{\theta}{1+\theta}\right)^k\quad$ où $\quad\alpha\in\mathbb{R}$.

1. Expliciter α de sorte que $(p_k)_{k\in\mathbb{N}}$ définisse une loi de probabilité.
Dans la suite, on considère une variable aléatoire X dont la loi est donnée par

$$\mathrm{X}(\Omega)=\mathbb{N}\quad\text{et}\quad\forall k\in\mathbb{N},\quad \mathbf{P}(\mathrm{X}=k)=p_k.$$

2. Reconnaître la loi de $\mathrm{X}+1$. En déduire l'espérance et la variance de X.

3. *Estimation de* θ *par maximum de vraisemblance*
Soient $(\mathrm{X}_1,\mathrm{X}_2,\dots,\mathrm{X}_n)$ un échantillon de même loi que X et $x_1,x_2,\dots,x_n$ appartenant à $\mathrm{X}(\Omega)$. On définit les fonctions L et ℓ par

$$\forall\theta\in\mathbb{R}_+^*,\quad \mathrm{L}(\theta)=\prod_{k=1}^{n}\mathbf{P}\left(\mathrm{X}_k=x_k\right)\quad\text{et}\quad \ell(\theta)=\ln\left(\mathrm{L}(\theta)\right).$$

L'objectif est de choisir la valeur de θ où est atteint le maximum (s'il existe) de L.

- **a)** Exprimer $\ell(\theta)$ à l'aide de $s_n=\sum_{k=1}^{n}x_k$. En déduire que ℓ admet un maximum.
- **b)** Justifier que L admet un maximum atteint en un unique point que l'on exprimera sous la forme $\theta=f(x_1,\dots,x_n)$ où f est une fonction de $\mathbb{R}^n\to\mathbb{R}$.

4. On pose maintenant $\mathrm{T}_n=f(\mathrm{X}_1,\dots,\mathrm{X}_n)$. Vérifier que T_n est un estimateur convergent de θ.

Exercice 41. ✦✦ **Loi exponentielle translatée**
Soient $\theta\in\mathbb{R}$ et Y une variable aléatoire suivant une loi exponentielle de paramètre 1. On considère la variable aléatoire $\mathrm{X}=\mathrm{Y}+\theta$ de densité f_θ définie par :

$$\forall x\in\mathbb{R},\qquad f_\theta(x)=\exp\left(-(x-\theta)\right)\mathbf{1}_{[\theta,+\infty[}(x).$$

1. Préciser l'espérance et la variance de X.

2. 🔍 Soit $(\mathrm{X}_1,\dots,\mathrm{X}_n)$ un échantillon de même loi que X. Vérifier que $\mathrm{T}_n=\overline{\mathrm{X}_n}-1$ est un estimateur convergent de θ.

3. *Estimation par le maximum de vraisemblance*
Soient $x_1, \ldots, x_n$ des réels fixés. On définit la fonction L sur $\mathbb{R}$ par

$$\forall\, \theta \in \mathbb{R}, \qquad \mathrm{L}(\theta) = \prod_{k=1}^{n} f_\theta(x_k).$$

a) i) Que de dire de $\mathrm{L}(\theta)$ si l'un des réels x_k est inférieur strictement à θ ?

ii) Dans le cas où $\theta \leqslant \min(x_1, \ldots, x_n)$, expliciter $\ln\big(\mathrm{L}(\theta)\big)$.

iii) En déduire que la fonction L admet un maximum atteint uniquement en $\widehat{\theta} = \min(x_1, \ldots, x_n)$.

b) Ce résultat justifie le choix de considérer l'estimateur : $\mathrm{S}_n = \min(\mathrm{X}_1, \ldots, \mathrm{X}_n)$.
On a vu que si $(\mathrm{Y}_1, \ldots, \mathrm{Y}_n)$ est un échantillon de même loi que Y alors $\min(\mathrm{Y}_1, \ldots, \mathrm{Y}_n)$ suit une loi exponentielle $\mathscr{E}(n)$. En déduire la loi de S_n. Justifier que S_n est un estimateur convergent de θ.

c) Quel est le meilleur estimateur entre T_n et S_n ?
On regardera quel estimateur à la risque quadratique $r_\theta(\mathrm{T}_n) = \big(b_\theta(\mathrm{T}_n)\big)^2 + \mathbf{V}_\theta(\mathrm{T}_n)$ *le plus petit.*

4. *Intervalle de confiance*
Soit $\alpha \in]0;1[$. On pose $\gamma_{n,\alpha} = |\ln(\alpha)|/n$. Montrer que

$$\mathbf{P}(\mathrm{S}_n - \gamma_{n,\alpha} \leqslant \theta \leqslant \mathrm{S}_n) = 1 - \alpha.$$

CHAPITRE 19

Statistiques

> *La statistique est aujourd'hui un fait social total : elle règne sur la société, régente les institutions et domine la politique. Un vêtement de courbes, d'indices, de graphiques, de taux recouvre l'ensemble de la vie. L'éducation disparaît derrière les enquêtes PISA, l'université derrière le classement de Shanghai, les chômeurs derrière la courbe du chômage... La statistique devait refléter l'état du monde, le monde est devenu un reflet de la statistique.*
>
> OLIVIER REY
> Mathématicien et philosophe contemporain

Que ce soit à l'issue d'une enquête, par observation directe ou à la suite d'une expérimentation, nous disposons à présent d'un ensemble de données qu'il convient maintenant de traiter. C'est-à-dire, les organiser, les représenter et en déterminer les éléments caractéristiques comme la moyenne, la variance ou l'écart-type. Tel est l'enjeu de la statistique descriptive.

1 Rappels en statistiques univariées

Soit Ω un ensemble.

- Ω désigne **la population** qui fera l'objet de l'étude.
- Les éléments $\omega_i \in \Omega$ sont les **individus**.
- Une **variable statistique** (ou **caractère**) X sur la population Ω est une application $X : \Omega \to F$. Dans le cas où F est une partie de $\mathbb{R}$, on dit que le caractère est **quantitatif**. Sinon, on dit que le caractère est qualitatif.

  Remarque. Dans la suite, on ne considère que les caractères quantitatifs.

- Si Ω est un ensemble fini, le nombre d'éléments de Ω, noté $\mathrm{Card}(\Omega)$ est **l'effectif de la population**.
- Quand il n'est pas possible d'étudier chaque individu de la population, on étudie seulement les individus d'une partie finie E de la population Ω. Dans ce cas, la partie E est appelée **échantillon**. Le nombre d'individus de l'échantillon, $\mathrm{Card}(E)$, est **la taille de l'échantillon**. Ainsi, si N désigne la taille, on peut écrire $E = \{e_1, e_2, \cdots, e_N\}$ où les e_i sont distincts deux à deux.

On distinguera deux types de variables statistiques.

- Si l'ensemble des valeurs prises par la variable, noté $X(\Omega)$, est un ensemble fini, on dit que la variable quantitative est **discrète**.
- Dans le cas contraire, on dit que la variable quantitative est **continue.**

1.1 Cas discret

Soit X une variable statistique discrète et $E = \{e_1, e_2, \cdots, e_N\}$ un échantillon.

- La donnée du N-uplet des observations $x = \big(X(e_1), X(e_2), \cdots, X(e_N)\big)$ définit une **série statistique**.
- L'échantillon étant un ensemble fini, l'ensemble des valeurs prises par la variable X sur E est aussi fini. On peut l'écrire $X(E) = \{m_1, m_2, \cdots, m_p\}$ où $m_1 < m_2 < \cdots < m_p$ où chaque m_i est une **valeur** ou **modalité.**

- **L'effectif d'une modalité** m_i est le nombre d'individu de E pour lequel le caractère prend la modalité m_i. C'est-à-dire $n_i = \text{Card}\{e \in E \mid X(e) = m_i\}$. Si N est la taille de l'échantillon, $N = \sum\limits_{i=1}^{p} n_i$.
- **La fréquence d'une modalité** m_i est la quantité $f_i = \dfrac{\text{Effectif de la modalité}}{\text{Taille de l'échantillon}} = \dfrac{n_i}{N}$. On a alors $\sum\limits_{i=1}^{p} f_i = 1$.
- **La fréquence cumulée d'une modalité** m_i est la somme de toutes les fréquences des modalités qui lui sont inférieures. Autrement dit, $F_i = \sum_{m_j \leqslant m_i} f_j$.

Remarque. Donner une série statistique est équivalent à la donnée du couple (m, n) où

- $m = (m_1, m_2, \cdots, m_p)$ est le p-uplet constitué des modalités,
- $n = (n_1, n_2, \cdots, n_p)$ est le p-uplet constitué des effectifs.

Il est alors commode de représenter une série statistique à l'aide d'un tableau :

Modalités	m_1	m_2	...	m_p	Total
Effectifs	n_1	n_2	...	n_p	N

1.2 Cas continu

On découpe l'ensemble des valeurs possibles $X(\Omega)$ en un certain nombre d'intervalles. Notons $p \in \mathbb{N}^*$ le nombre d'intervalles choisis. Ces intervalles doivent être deux à deux disjoints, et leur réunion est égale à (ou contient) l'ensemble des valeurs possibles. Plus précisément, en notant $a_1, a_2, \ldots, a_{p+1} \in \mathbb{R}$, les bornes de tous ces intervalles avec $a_1 < a_2 < \cdots < a_{p+1}$, on a

$$X(\Omega) \subset \bigcup_{i=1}^{p} [a_i ; a_{i+1}[.$$

L'intervalle $[a_i ; a_{i+1}[$ est une **classe**.

Remarques. On peut ainsi définir l'effectif d'une classe et sa fréquence. On peut aussi utiliser ces définitions pour une variable discrète lorsque l'ensemble des valeurs prises est trop important.

2 Paramètres

2.1 Paramètres de position

Définition 29 (La moyenne empirique)

Soit $x = (x_i)_{1 \leqslant i \leqslant N}$, une série statistique. On définit la **moyenne** de la série par

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i.$$

Si $(m_i)_{1 \leqslant i \leqslant p}$, $(n_i)_{1 \leqslant i \leqslant p}$ et $(f_i)_{1 \leqslant i \leqslant p}$ désignent respectivement les modalités, les effectifs et les fréquences, on a aussi

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{p} n_i \, m_i = \sum_{i=1}^{p} f_i \, m_i$$

La moyenne empirique s'obtient avec Python par la commande `np.mean(x)`.

Définition 30 (La médiane d'une série statistique)

Soit $x = (x_i)_{1 \leqslant i \leqslant N}$ une série statistique ordonnées suivant l'ordre croissant. On définit la **médiane** de la série statistique par

- x_p si N est impair avec $p = (N+1)/2$,
- $\frac{x_p + x_{p+1}}{2}$ si N est pair avec $p = N/2$.

Autrement dit, la médiane un nombre réel m_e tel que le nombre d'individus pour lesquels X prend une valeur inférieure ou égale à m_e soit égal au nombre d'individus pour lesquels X prend une valeur supérieure ou égale à m_e.

2.2 Paramètres de dispersion

Définition 31 (La variance empirique et l'écart-type)

Soit x, une série statistique.
- **La variance** de la série est le réel, noté s_x^2, défini par

$$s_x^2 = \frac{1}{\mathrm{N}}\left(\left(x_1-\overline{x}\right)^2+\left(x_2-\overline{x}\right)^2+\ldots+\left(x_\mathrm{N}-\overline{x}\right)^2\right) = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left(x_i-\overline{x}\right)^2.$$

- **L'écart-type**, s_x d'une série statistique est défini comme la racine carrée de la variance.

Remarques.
- Si les modes $m_1, m_2, \cdots, m_p$ de la série sont donnés avec leurs effectifs $n_1, n_2, \cdots, n_p$ ou leurs fréquences $f_1, f_2, \ldots, f_p$, alors

$$s_x^2 = \frac{1}{\mathrm{N}}\sum_{i=1}^{p} n_i\left(m_i-\overline{x}\right)^2 = \sum_{i=1}^{p} f_i\left(m_i-\overline{x}\right)^2.$$

- Une variance est toujours un nombre positif. L'écart-type est donc bien défini.
- Une série a une variance nulle si et seulement si toutes les valeurs de la série sont identiques.

Les commandes Python sont `np.var(x)`, pour la variance et `np.std(x)`, pour l'écart-type.

En pratique, on utilise la formule suivante pour calculer la variance.

Théorème 32 (Formule de Koenig-Huygens)

Soit x une série statistique de modalités $(m_i)_{1\leqslant i\leqslant p}$, d'effectifs $(n_i)_{1\leqslant i\leqslant p}$ et de variance s_x^2, alors

$$s_x^2 = \left(\frac{1}{\mathrm{N}}\sum_{i=1}^{p} n_i {m_i}^2\right) - \overline{x}^2.$$

Remarque. Si x^2 désigne la série dont chaque donnée est ${x_i}^2$. On démontre que $\frac{1}{\mathrm{N}}\sum_{i=1}^{p} n_i {m_i}^2$ est la moyenne $\overline{(x^2)}$. La formule de formule de Koenig-Huygens devient alors

$$s_x^2 = \overline{(x^2)} - \left(\overline{x}\right)^2.$$

3 Statistiques bivariées

3.1 Premières définitions

Définition 33 (La covariance empirique)

Soient $x = (x_i)_{1\leqslant i\leqslant \mathrm{N}}$ et $y = (y_i)_{1\leqslant i\leqslant \mathrm{N}}$, deux séries statistiques. On définit la **covariance** des deux séries par

$$\mathrm{Cov}(x,y) = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right).$$

On définit aussi le **coefficient de corrélation empirique** par $\rho_{x,y} = \dfrac{\mathrm{Cov}(x,y)}{\sigma_x\sigma_y} \in [-1;1]$.

Proposition 34 (Calcul de la covariance)

Soient $x = (x_i)_{1\leqslant i\leqslant N}$ et $y = (y_i)_{1\leqslant i\leqslant N}$, deux séries statistiques. On a

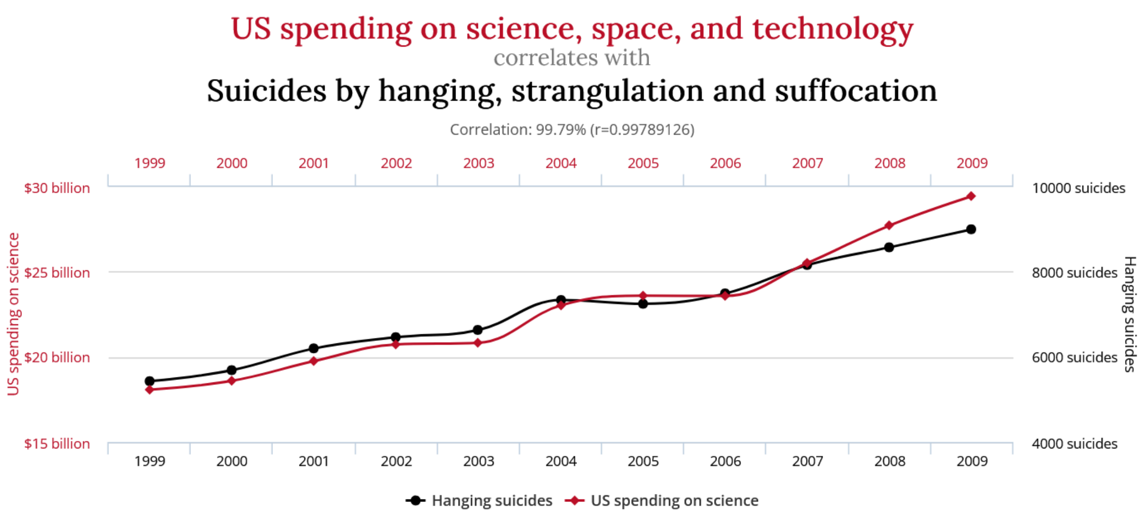
$$\mathrm{Cov}(x, y) = \overline{x * y} - \overline{x} \cdot \overline{y}$$

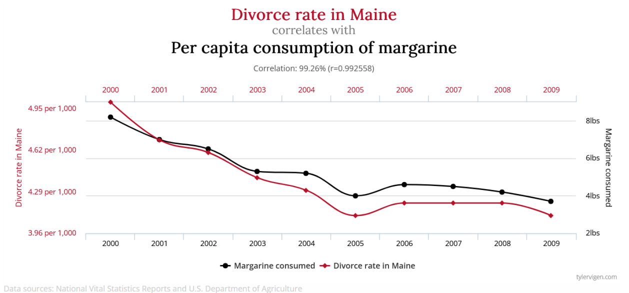
où $\overline{x * y}$ désigne la moyenne de la série $x * y = (x_i \cdot y_i)_{1\leqslant i\leqslant N}$.
Dès lors, la commande Python pour obtenir la covariance empirique est

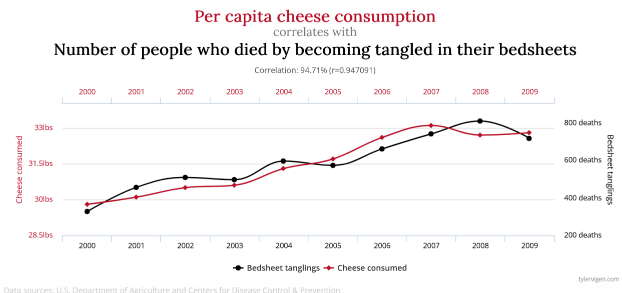
```
np.mean(x*y)-np.mean(x)*np.mean(y)
```

Corrélation n'est pas causalité

Quelques exemples issus du site : *http ://tylervigen.com/spurious-correlations.*







3.2 Nuage de points, droite de régression

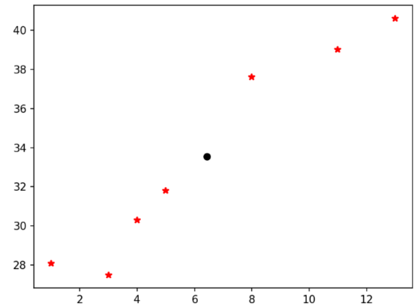
Dans la suite, x, y désigne deux séries statistiques. La question est de savoir dans quelle mesure l'une des deux, dite **expliquée**, dépend de l'autre, dite **explicative**.

Définition 35 (Nuage de points)

Soient $x = (x_i)_{1\leqslant i\leqslant N}$ et $y = (y_i)_{1\leqslant i\leqslant N}$, deux séries statistiques.
Le **nuage de points** associé à ces deux séries est l'ensemble des points M_i du plan de coordonnées (x_i, y_i) où $i \in [\![1, N]\!]$.

Remarque. Le point $(\overline{x}, \overline{y})$ est le **point moyen** du nuage.

- Voici le code pour tracer un nuage de points associé à deux séries statistiques de même effectif, ainsi que sont point moyen.

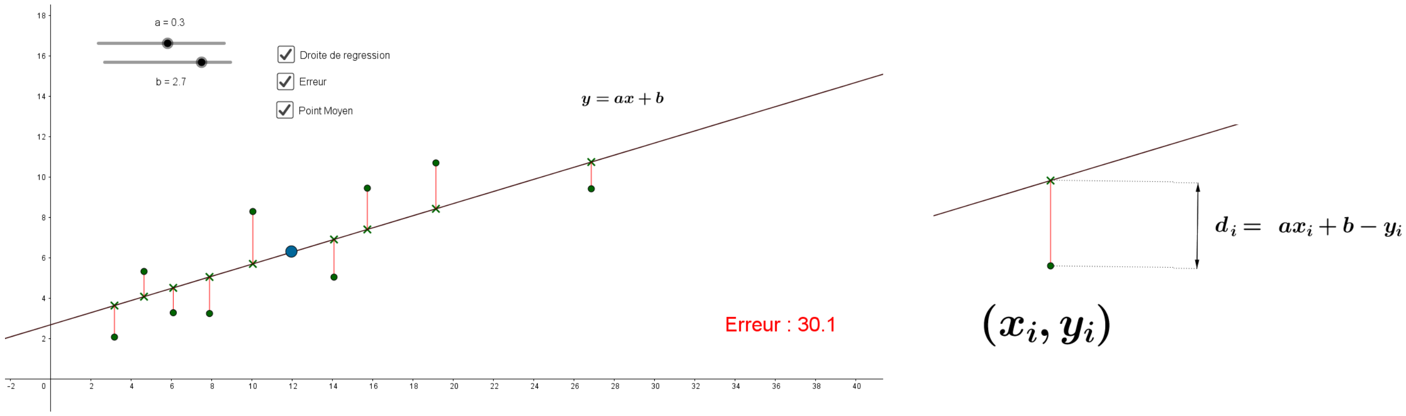


Editeur

```
import matplotlib.pyplot as plt
import numpy as np

# Exemples avec deux séries
x=np.array([1,3,4,5,8,11,13])
y=np.array([28.1,27.5,30.3,31.8,37.6,39,40.6])


# Tracé du nuage de points et du point moyen

plt.plot(x,y,'r*')
plt.plot(np.mean(x),np.mean(y),'ko')
```

Soit n un entier supérieur à 2.
Considérons n points de $\mathbb{R}^2$, $(x_1, y_1), \dots, (x_n, y_n)$ non alignés verticalement. On cherche la droite qui « approxime » au mieux ces n points. Si on note $y = ax + b$, l'équation d'une droite, on cherche à minimiser l'erreur

$$\mathrm{E}_r = \sum_{i=1}^{n} {d_i}^2 = \sum_{i=1}^{n} \left(ax_i + b - y_i\right)^2.$$



Point de vue algèbre linéaire

Traduisons matriciellement le problème. Posons.

$$\mathrm{X} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \mathrm{A} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \quad \text{et} \quad \mathrm{B} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \text{de sorte que} \quad \mathrm{AX} - \mathrm{B} = \begin{bmatrix} ax_1 + b - y_1 \\ ax_2 + b - y_2 \\ \vdots \\ ax_n + b - y_n \end{bmatrix}.$$

Si on considère le produit scalaire canonique sur $\mathscr{M}_{n,1}(\mathbb{R})$ et la norme associée

$$\mathrm{E}_r = \|\mathrm{AX} - \mathrm{B}\|^2.$$

D'après les résultats sur les projecteurs et le théorème des moindres carrés, on a montré qu'il existe une unique solution X_0 donnée par l'équation ${}^t\mathrm{AAX}_0 = {}^t\mathrm{AB}$. On obtient alors ses composantes a et b :

$$a = \rho(x, y)\sqrt{\frac{\mathbf{V}(y)}{\mathbf{V}(x)}} \quad \text{et} \quad b = \overline{y} - a\overline{x}.$$

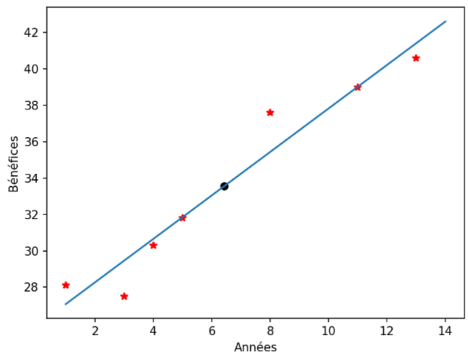
L'équation de la droite est alors :

$$y - \overline{y} = \rho(x, y)\frac{\sigma(y)}{\sigma(x)}(x - \overline{x}).$$

Exemple. On peut compléter l'exemple Python précédent en affichant la droite de régression. On comprend aussi sur cet exemple que la droite de régression peut être utiliser pour faire de la prédiction en prolongeant la droite.
Précisons aussi qu'il existe des commandes Python plus directes pour afficher la droite de régression.

Editeur

```
# Tracé de la droite de régression

plt.xlabel('Années')
plt.ylabel('Bénéfices')
COVxy=np.mean(x*y)-np.mean(x)*np.mean(y)
a=COVxy/np.std(x)**2
b=np.mean(y)-a*np.mean(x)
t=np.linspace(1,14,2)
plt.plot(t,a*t+b)
```

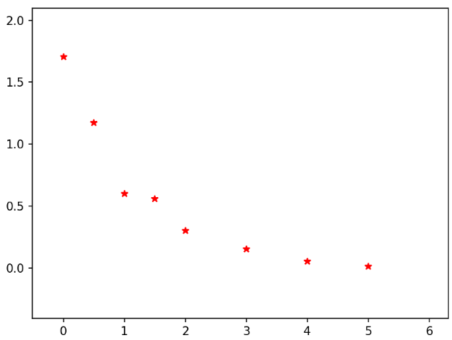


Remarque. On vérifie que le point moyen appartient à la droite de régression.

Point de vue fonctions de plusieurs variables (voir exercice 20, p.15)

3.3 Un exemple d'ajustement linéaire

Il arrive parfois que le nuage ne s'accorde pas à une droite mais à une courbe d'une fonction classique.

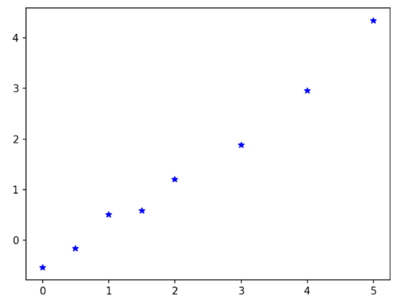


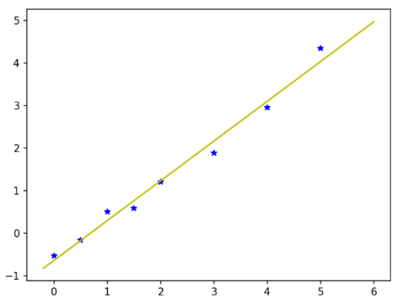
On peut alors supposer que le nuage s'organise autour de la courbe d'équation $y = f(ax+b)$ pour une certaine fonction f. Si f est bijective, on introduit la série $z = f^{-1}(x)$. Prenons l'exemple des séries :

Editeur

```
x=np.array([0,0.5,1,1.5,2,3,4,5])
y=np.array([1.707, 1.173, 0.601, 0.558,
    0.299,0.152, 0.052, 0.013])

plt.plot(x,y,'r*')
```

Dans l'exemple, on va tester avec la fonction f définie sur $\mathbb{R}$ par $f(t) = \exp(-t)$.
On procède alors à une régression linéaire sur les séries x, z pour estimer les réels a et b.





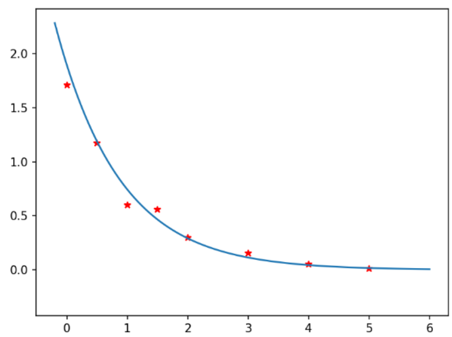
Console

```
>>> a
0.9349559793558717
>>> b
-0.6381866319796354
```

Testons le résultat, en comparant la courbe obtenue avec le nuage de points.

Editeur

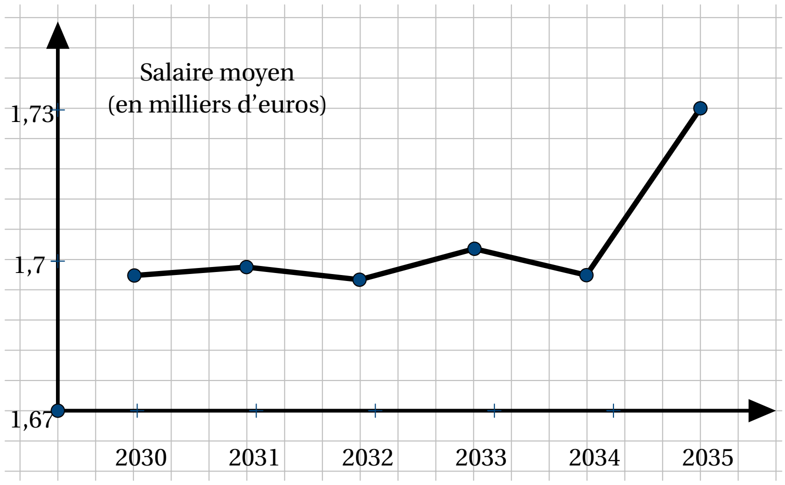
```
z=-np.log(y)

COVxz=np.mean(x*z)-np.mean(x)*np.mean(z)
a=COVxz/np.std(x)**2
b=np.mean(z)-a*np.mean(x)

plt.plot(x,y,'r*')
t=np.linspace(-0.2,6,100)
plt.plot(t,np.exp(-(a*t+b)))
plt.show()
```



Quelques biais statistiques

Exercice 42. ✦✦ Chercher l'erreur ou le biais dans les raisonnements suivants.

1. Dans la classe d'ECG, on compte qu'en moyenne les élèves ont 1.7 frères et sœurs. Ce qui donne 2.7 enfants par femme alors que le nombre d'enfants par femme n'est que de 1.8 dans la population française. Donc les enfants de familles nombreuses font plus facilement des études.
2. L'espérance de vie des professeurs de mathématiques est de 82 ans, soit 3 ans de plus que l'espérance de la population française. Donc les mathématiques sont bonnes pour la santé !
3. Lors de la seconde Guerre Mondiale, la Royal Air Force souhaite améliorer le taux de retour de ses bombardiers partis frapper les positions Allemandes. Les ingénieurs de la R.A.F décident d'étudier la localisation des impacts de balles sur les avions à leur retour de mission, puis de renforcer le blindage sur les zones les plus atteintes.
4. En 1936, le démocrate Franklin D. Roosevelt et le Républicain Alfred M. Landon concoururent pour la présidence des États-Unis. Juste avant l'élection, le journal *Literary Digest* réalisa un sondage de grande ampleur : 10 millions de bulletins de sondage sont distribués aux abonnés du magazine et à des gens figurant au bottin téléphonique. Après plus de 2 millions de réponses, Alfred Landon est annoncé président des États-Unis !
5. Comme le montre le graphe suivant, la nouvelle politique économique démarrée en 2034 a eu des effets considérables sur le salaire moyen des français.



6. Deux traitements contre les calculs rénaux présentent les résultats en fonction de la taille des calculs

petits calculs		gros calculs	
Traitement 1	Traitement 2	Traitement 1	Traitement 2
81 succès sur 87	234 succès sur 270	192 succès sur 263	55 succès sur 80

En regroupant les résultats, on obtient

Traitement 1	Traitement 2
78% (273 succès sur 350)	83% (289 succès sur 350)

Le second traitement est donc le plus efficace car il a le plus gros pourcentage de réussite.

7. Dans une université, à l'examen de la licence de biologie, les filles ont mieux réussi que les garçons et à l'examen de la licence de physique, les filles ont, là encore, mieux réussi que les garçons. Pourtant, en regroupant les résultats des deux licences, on découvre que les garçons ont mieux réussi que les filles. Il y a donc une erreur dans les résultats de la licence en faveur des garçons.

Les faits sont têtus, il est plus facile de s'arranger avec les statistiques.

Mark Twain

CHAPITRE 20

Optimisation sous contraintes linéaires

La science naît du jour où des erreurs, des échecs, des surprises désagréables, nous poussent à regarder le réel de plus près.

RENÉ THOM
Mathématicien français, fondateur de la théorie des catastrophes (1923-2002)

1 Position du problème

Soient $n \in \mathbb{N} \setminus \{0;1\}$ et $p \in \mathbb{N}^*$ avec $p < n$. Considérons de plus $g_1, g_2, \ldots, g_p$, p formes linéaires sur $\mathbb{R}^n$. Autrement dit, pour chaque $i \in [\![1;p]\!]$, il existe $\lambda_{i,1}, \lambda_{i,2}, \ldots, \lambda_{i,n}$ tels que

$$\forall\, x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n, \quad g_i(x) = \sum_{k=1}^{n} \lambda_{i,k} x_k.$$

Notons $\mathscr{C}$, l'ensemble des solutions du système linéaire

$$\mathscr{S} \quad \left\{ \begin{array}{ccc} g_1(x) & = & b_1 \\ \vdots & \vdots & \vdots \\ g_p(x) & = & b_p, \end{array} \right.$$

où l'inconnue est un élément x de $\mathbb{R}^n$. Soit $\mathscr{H}$ l'ensemble des solutions du système homogène associé :

$$\mathscr{S}_0 \quad \left\{ \begin{array}{ccc} g_1(x) & = & 0 \\ \vdots & \vdots & \vdots \\ g_p(x) & = & 0. \end{array} \right.$$

Comme $\mathscr{H} = \bigcap_{i=1}^{p} \operatorname{Ker}(g_i)$, on constate que $\mathscr{H}$ est un sous-espace vectoriel de $\mathbb{R}^n$. Si $\mathscr{C}$ n'est pas vide et si $x_0 \in \mathscr{C}$ alors on peut écrire $\mathscr{C} = \{x_0 + h \in \mathbb{R}^n \mid h \in \mathscr{H}\}$. On dit alors que $\mathscr{C}$ est un espace affine.

Définition 36 (extremum sous contrainte)

Soit $f : \mathscr{O} \to \mathbb{R}$ où $\mathscr{O}$ est une partie de $\mathbb{R}^n$.
- On dit que f admet un **extremum global sous la contrainte** $\mathscr{C}$ en un point a si la restriction de f à $\mathscr{C}$ admet un extremum global en a.
- On dit que f admet un **extremum local sous la contrainte** $\mathscr{C}$ en un point a si la restriction de f à $\mathscr{C}$ admet un extremum local en a.

Remarque. La fonction f admet un maximum global sous la contrainte $\mathscr{C}$ en a si et seulement si :

$$\forall\, x \in \mathscr{C}, \qquad f(x) \leqslant f(a).$$

Exemple. Une société de production d'épingles possède deux usines avec des fonctionnements différents. La première usine produit x tonne(s) d'épingles pour un coût $C_1(x)$ et la seconde produit la même quantité avec un coût $C_2(x)$ (en euros). Une analyse des coûts donne : $C_1(x) = 3+4x$ et $C_2(x) = 4x^2+2$. Le coût total d'une production d'une quantité x_1 par la première usine et x_2, pour la seconde est

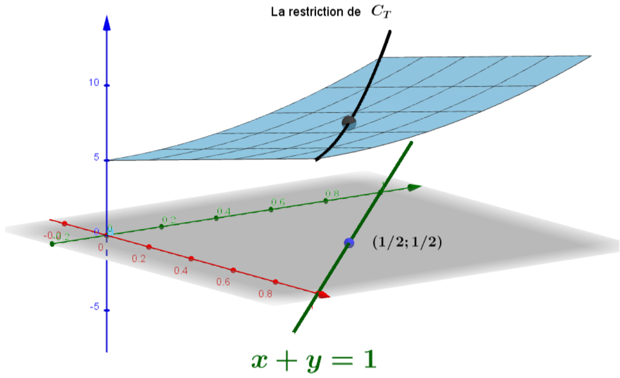
$$C_T(x_1, x_2) = C_1(x_1) + C_2(x_2) = 5 + 4x_1 + 4{x_2}^2.$$

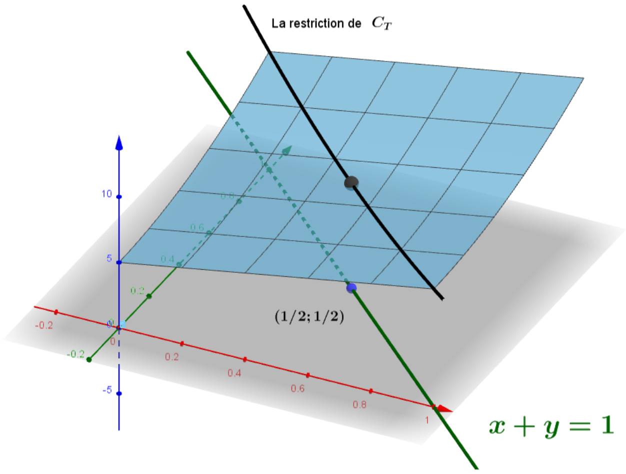
Pour une production d'une tonne, il faut rajouter la contrainte linéaire $g(x_1, x_2) = x_1 + x_2 = 1$. Si on souhaite minimiser le coût d'une tonne d'épingles en répartissant la production entre les deux usines, on cherche le minimum global de la fonction C_T sous la contrainte $\mathscr{C} = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 = 1\}$.

2 Recherche d'extrema sous contraintes linéaires

2.1 Par substitution

L'exemple des épingles. Ci-dessous, la surface représentative de C_T et la courbe représentative de la restriction à $\mathscr{C}$.



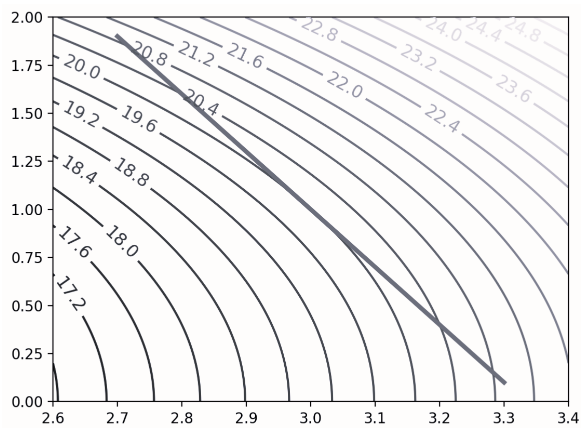


Exercice 43

1. Déterminer le minimum de $f : (x, y) \in \mathbb{R}^2 \mapsto x^2 + y^2 + 10$ sous la contrainte
$3x + y = 10$.
2. Tester la pertinence du résultat à l'aide du code Python ci-dessous.

Editeur

```
x=np.linspace(2.6,3.4,200)
y=np.linspace(0,2,200)
X , Y = np.meshgrid(x,y)
Z = X**2+Y**2+10

graphe = plt.contour(X,Y,Z,25)
plt.clabel(graphe,inline=1)
# lignes de niveau

x=np.array([2.7,3.3])
plt.plot(x,-3*x+10)
plt.show()
```



Exercice 44

✧ Soit f la fonction définie sur $\mathbb{R}^3$ par $f(x, y, z) = xe^{x(y^2+z^2+1)}$.

1. En raisonnant par substitution, déterminer le minimum de de f sous la contrainte linéaire : $\begin{cases} x = 1 \\ y + z = 0 \end{cases}$.
2. Est-ce que f admet un maximum sous cette même contrainte ?

2.2 Par du calcul différentiel

Les énoncés

Théorème 37 (condition nécessaire d'ordre 1)

Soient $f : \mathscr{O} \to \mathbb{R}$ où $\mathscr{O}$ est une partie ouverte de $\mathbb{R}^n$ et $a \in \mathscr{O}$.

Si
- f est de classe $\mathscr{C}^1$ sur $\mathscr{O}$.
- f admet un extremum sous la contrainte $\mathscr{C}$ en un point a.

Alors pour tout $h \in \mathscr{H}$, la dérivée de f en a dans la direction h est nulle, c'est-à-dire

$$\forall\, h \in \mathscr{H}, \qquad \langle \nabla f(a), h \rangle = 0.$$

Autrement dit, le gradient $\nabla f(a)$ appartient à $\mathscr{H}^\perp$.

Vocabulaire. Soient f une fonction de classe $\mathscr{C}^1$ sur un ouvert $\mathscr{O}$ de $\mathbb{R}^n$, a un point de $\mathscr{O}$. Si $\nabla f(a)$ appartient à $\mathscr{H}^\perp$, on dit que a est un **point critique de f sous contrainte** $\mathscr{C}$.

Remarque. Avec les notations du début de chapitre, pour tout $i \in [\![1;p]\!]$, g_i est de classe $\mathscr{C}^1$ avec :

$$\forall\, x \in \mathbb{R}^n, \qquad \nabla g_i(x) = (\lambda_1, \lambda_2, \ldots, \lambda_n) = \text{ Cste}.$$

On notera donc simplement ∇g_i (on confond ici la constante avec la fonction). En particulier, on a

$$\forall\, x \in \mathbb{R}^n, \qquad g_i(x) = \langle \nabla g_i, x \rangle.$$

Proposition 38 (famille génératrice de l'orthogonal)

Avec les notations introduites précédemment,

$$\mathscr{H}^\perp = \mathrm{Vect}\left(\nabla g_1, \ldots, \nabla g_p\right)$$

Théorème 39 (condition d'ordre 1)

Soient f une fonction de classe $\mathscr{C}^1$ sur un ouvert $\mathscr{O}$ de $\mathbb{R}^n$ et $a \in \mathscr{O}$.

Si
- f est de classe $\mathscr{C}^1$ sur $\mathscr{O}$.
- f admet un extremum sous la contrainte $\mathscr{C}$ en un point a.

Alors il existe $\left(\lambda_1, \ldots, \lambda_p\right) \in \mathbb{R}^p$ tels que $\nabla f(a) = \sum_{i=1}^{p} \lambda_i \nabla g_i$.

Vocabulaire. Si la famille $\left(\nabla g_1, \ldots, \nabla g_p\right)$ est libre, les réels $\lambda_1, \ldots, \lambda_p$ sont uniques. Ils sont appelés multiplicateurs de Lagrange.

Exercice 45

✦ Déterminer les extrema possibles de la fonction f définie sur $\mathbb{R}^3$ par $f(x, y, z) = xyz$ sous la contrainte

$$\mathscr{C} : \left\{ \begin{array}{rcl} x + y + z & = & 0 \\ x - y + z & = & 2. \end{array} \right.$$

⚠ **Attention.** La condition précédente n'est qu'une condition nécessaire et non suffisante. Elle ne fournit que des candidats pour les extrema et il convient ensuite de poursuivre l'analyse.

- Si l'on recherche un extremum *global*, on peut revenir à la définition et étudier le signe de

$$f(x) - f(a) \quad \text{pour } x \in \mathscr{C}$$

ou, à l'aide du changement de variable $h = x - a$, on regarde le signe de

$$f(a+h) - f(a) \quad \text{pour } h \in \mathscr{H}.$$

- Si on cherche un extremum *local*, on peut revenir au développement limité d'ordre 2. Notons q_a, la forme quadratique de f en a.

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} q_a(h) + \|h\|^2 \varepsilon(h)$$

avec $\varepsilon(h) \underset{h \to 0}{\longrightarrow} 0$ et $h \in \mathscr{H}$. Dans le cas d'un point critique sous la contrainte $\mathscr{C}$, on obtient la simplification :

$$\forall\, h \in \mathscr{H}, \qquad f(a+h) - f(a) = \frac{1}{2} q_a(h) + \|h\|^2 \varepsilon(h) \quad (\text{car } \nabla f(a) \in \mathscr{H}).$$

On regarde ensuite le signe de la forme quadratique (restreinte à $\mathscr{H}$).

→ Si $q_a(h) \geqslant 0$ pour tout $h \in \mathscr{H}$, on a un minimum local.

→ Si $q_a(h) \leqslant 0$ pour tout $h \in \mathscr{H}$, on a un maximum local.

→ Sinon, on ne peut pas conclure.

- Dernier argument : se souvenir qu'il existe toujours un maximum et un minimum pour une fonction continue sur un fermé borné.

Deux exemples d'études globales

- Soient $f : (x, y, z) \in \mathbb{R}^3 \mapsto x^2 + y^2 + z^2 \in \mathbb{R}$ et $g : (x, y, z) \in \mathbb{R}^3 \mapsto x + y + z \in \mathbb{R}$.

Déterminons les extrema globaux de f sous la contrainte $\mathscr{C} : g(x, y, z) = 3$.

→ *Étude du ou des points critiques sous contrainte*

→ *Étude locale au niveau du point critique*

→ *Étude globale*

- On considère

$$f : x = (x_1, \dots, x_n) \in \left(\mathbb{R}^+\right)^n \mapsto \prod_{k=1}^{n} x_k \quad \text{et} \quad g : x \in \mathbb{R}^n \mapsto \sum_{k=1}^{n} x_k.$$

Étudions les extrema de f sous la contrainte $g(x) = c$ où $c \in \mathbb{R}_*^+$.

→ *Étude du ou des points critiques sous contrainte*

→ *Nature du point critique*

Exercice 46

✦ ✎ Soit f définie sur $\mathbb{R}^3$ par $f(x, y, z) = x^2 - 2xy + yz + y - z$.

1. Vérifier que $a = \left(\frac{4}{5}, \frac{3}{5}, \frac{1}{5}\right)$ est l'unique point critique sous la contrainte

$$\begin{cases} 2x - y & = 1 \\ x + z & = 1. \end{cases}$$

2. On a $f(a) = 1/5$. Est-ce un maximum ou minimum global ? ni l'un ni l'autre ?

2.3 Interprétation géométrique avec une seule contrainte linéaire

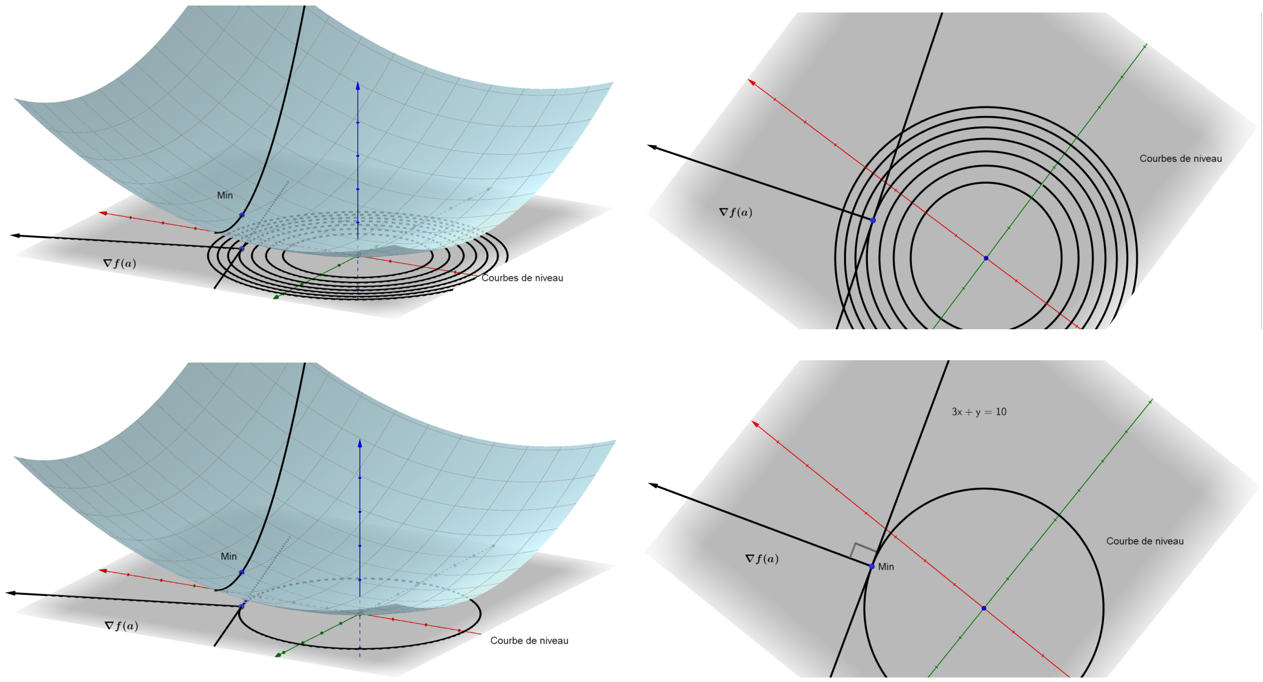
Dans le cas d'une fonction de deux variables avec une seule contrainte linéaire, on cherche à minimiser la restriction de f à une droite d'équation $g(x, y) = \lambda x + \mu y = b$. Le vecteur ∇g est une vecteur normal [1] à la droite $\mathscr{C}$. Or, on sait que :

→ Si un extremum sous contrainte est atteint en a, $\nabla f(a)$ est colinéaire à ∇g. En particulier, $\nabla f(a)$ est un vecteur normal à la droite $\mathscr{C}$.

→ Le vecteur $\nabla f(a)$ est aussi orthogonal à la ligne de niveau $\mathscr{L}_{f(a)}$.

1. Dit autrement, le vecteur est orthogonal aux vecteurs directeurs de la droite.

On en déduit que dans le cas d'un extremum sous contrainte, la droite qui représente la contrainte est tangente à la ligne de niveau. Autrement dit, elle ne peut la franchir.
C'est ce qu'on constate avec l'exemple de la fonction $f : (x, y) \in \mathbb{R}^2 \mapsto x^2 + y^2 + 10$ sous la contrainte $3x + y = 10$.



3 Cas des fonctions définies sur des fermés bornés

Commençons par résumer la démarche avant de l'appliquer à un exemple.

Méthode

Existence et calcul des extrema d'une fonction définie sur un fermé borné
Dans la suite, on considère une fonction f de classe $\mathscr{C}^1$ sur un ensemble fermé borné, noté $\mathscr{K}$.

- *I. Existence*
  On vérifie que l'ensemble de départ de f est bien une partie :
  - fermé, par exemple, en montrant qu'il existe φ, une fonction de $\mathbb{R}^n$ et continue, telle que
  $$\mathscr{K} = \left\{x \in \mathbb{R}^n \,|\, \varphi(x) \leqslant c\right\}.$$
  - bornée, en montrant qu'il existe un réel M tel que pour tout $x \in \mathscr{K}$, $\|x\| \leqslant \mathrm{M}$.

  On sait alors que f admet un maximum global et un minimum global sur $\mathscr{K}$, mais on ne sait pas en quels points ils sont atteints.
- *II. Étude sur l'ouvert*
  Soit $\mathscr{O}$, le plus grand ouvert inclus dans $\mathscr{K}$. Autrement dit, $\mathscr{O}$ correspond à $\mathscr{K}$ privé de ses bords. On étudie alors les points critiques pour avoir des points candidats possibles où les extrema sont atteints.
- *III. Étude sur le bord*
  On étudie maintenant la fonction sur le bord $\mathscr{K} \setminus \mathscr{O}$. On retrouve le cas d'une recherche d'extremum sous contrainte.
- *IV. Conclusion*
  On compare les différents valeurs de f aux points obtenus sur l'ouvert et sur le bord pour déterminer le maximum et le minimum global.

Exemple. Considérons la fonction f définie sur $\mathbb{R}^2$ par $f(x, y) = 3x^2 + 2y^2 - 4x - 3y + 4xy + 3$.
On souhaite étudier le minimum global de la restriction de f à l'ensemble

$$\mathscr{K} = \left\{(x, y) \in \mathbb{R}^2 \,|\, x \geqslant 0,\ y \geqslant 0, x + y \leqslant 1\right\}.$$

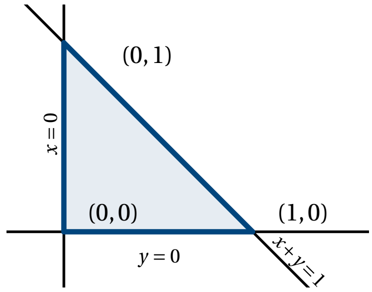
- *I. Existence*
  L'ensemble $\mathscr{K}$ est délimité par les droites d'équation $x = 0$, $y = 0$ et $x + y = 1$. C'est un fermé. Par exemple, c'est l'intersection de trois fermés :

$$\mathbb{R}^+ \times \mathbb{R}, \quad \mathbb{R} \times \mathbb{R}^+ \quad \text{et} \quad \left\{(x, y) \in \mathbb{R}^2 \mid x + y \leqslant 1\right\}.$$

  La partie $\mathscr{K}$ est aussi bornée. Par exemple

$$\mathscr{K} \subset \overline{\mathscr{B}}(0; 1).$$

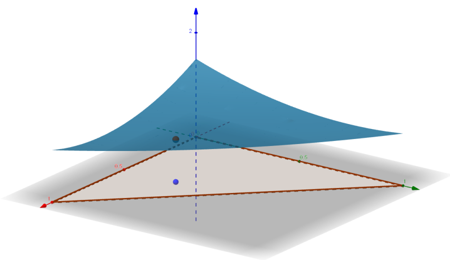
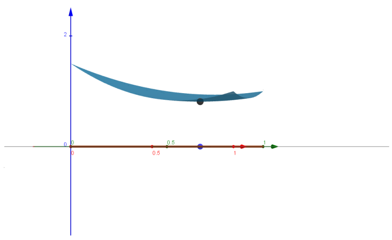
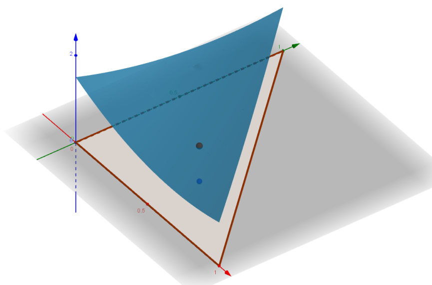
  C'est-à-dire $\quad \forall (x, y) \in \mathscr{K}, \quad \|(x, y)\| \leqslant 1.$
  La fonction f étant continue car polynomiale, il existe donc un minimum sur $\mathscr{K}$.



- *II. Étude sur l'ouvert*
- *III. Étude sur le bord*
- *IV. Conclusion*
  En comparant les valeurs des minima obtenus sur chaque sous-ensembles de $\mathscr{K}$, le minimum global est atteint pour :

$$a = \left(\frac{1}{2}; \frac{1}{4}\right) \quad \text{et} \quad f(a) = 1,675.$$

Illustrons le résultat par plusieurs vues de la surface de f :

Exercice 47

✦✦ Exemple
Soient $\mathrm{D} = [-1; 1]^2$ et f la fonction définie sur D par

$$f(x, y) = (y - x)^2 + 6xy.$$

1. Justifier que la fonction f admet un maximum et un minimum.
2. Vérifier que f a un unique point critique. Préciser sa nature.
3. Donner le maximum et minimum de f.

Exercices

Exercice 48. ✦ ✎ Soit f définie sur $\mathcal{O} = \left(\mathbb{R}_+^*\right)^3$ par $f(x,y,z) = x\ln x + y\ln y + z\ln z$. Pour tout $c \in \mathbb{R}$, on pose

$$\mathscr{C}_c = \left\{(x,y,z) \in \mathcal{O} \mid x+y+z = 3c\right\}.$$

1. Vérifier que f est de classe $\mathscr{C}^1$ sur $\mathcal{O}$, puis montrer que f admet un unique point critique sous la contrainte $\mathscr{C}_c$.

2. Justifier que ce point critique correspond à un minimum local sous la contrainte $\mathscr{C}_c$.

Exercice 49. ✦✦ 🔑 Soient $n \in \mathbb{N} \setminus \{0;1\}$ et f la fonction définie sur $\left(\mathbb{R}_*^+\right)^n$ par

$$f(x_1,\ldots,x_n) = \sum_{k=1}^{n} {x_k}^4.$$

Étudier les extrema locaux et globaux de f sous la contrainte $\sum_{i=1}^{n} x_i = n$.

Exercice 50. ✦✦ ✎ **Variante de l'inégalité arithmético-géométrique**
Soient $n \in \mathbb{N}^*$, $\alpha_1,\ldots,\alpha_n \in \mathbb{R}_*^+$ tels que $\sum_{i=1}^{n} \alpha_i = 1$. On définit en plus les fonctions f et g sur $\left(\mathbb{R}^+\right)^n$ par

$$f(x_1,\ldots,x_n) = \prod_{i=1}^{n} {x_i}^{\alpha_i} \quad \text{et} \quad g(x_1,\ldots,x_n) = \sum_{i=1}^{n} \alpha_i x_i.$$

Posons de plus $\mathscr{K} = \left\{(x_1,\ldots,x_n) \in \left(\mathbb{R}^+\right)^n \mid g(x_1,\ldots,x_n) = 1\right\}$.

1. Déterminer les points critiques de f sur $\left(\mathbb{R}_+^*\right)^n$ sous la contrainte $g(x_1,\ldots,x_n) = 1$.

2. Montrer que f possède un maximum M sur $\mathscr{K}$ et que celui-ci est atteint sur $\mathscr{K} \cap \left(\mathbb{R}_+^*\right)^n$.

3. 🔑 En déduire que

$$\forall\, (x_1,\ldots,x_n) \in \left(\mathbb{R}^+\right)^n, \qquad \prod_{i=1}^{n} {x_i}^{\alpha_i} \leqslant \sum_{i=1}^{n} \alpha_i x_i.$$

Optimisation et probabilité

Exercice 51. ✦✦ **Optimisation et estimateur**
Soit $\lambda \in \mathbb{R}_*^+$, un paramètre inconnu. Soit $(X_1,\ldots,X_n)$ un échantillon d'une loi de Poisson de paramètre λ.
Montrer que parmi toutes les combinaisons linéaires des variables $X_1,\ldots,X_n$, il existe un unique estimateur T_n sans biais de λ et qui est de variance minimale. Reconnaître cet estimateur.

Exercice 52. ✦✦ ✎ Soient deux entiers $n \geqslant 2$ et $r \geqslant 3$. Une expérience aléatoire peut aboutir à r résultats différents R_1, R_2, …, R_r de probabilités respectives $x_1, x_2, \ldots, x_r \in \,]0;1[$. On effectue ensuite n répétitions de l'expérience précédente et, pour tout $i \in [\![1,r]\!]$, on note X_i la variable aléatoire qui vaut 1 si le résultat R_i n'est pas obtenu à l'issue de ces n épreuves et qui vaut 0 sinon.
On désigne par X la variable aléatoire égale au nombre de résultats qui n'ont pas été obtenus à l'issue des n épreuves.

1. **a)** Exprimer la variable X en fonction des variables $X_1, X_2, \ldots, X_r$.

b) Donner la loi de X_i pour tout $i \in [\![1;r]\!]$.

c) En déduire l'espérance de X.

2. En optimisant sous la contrainte $\sum_{i=1}^{r} x_i = 1$, justifier l'existence et donner la valeur du minimum local de $\mathbf{E}(X)$.

3. Dans la suite des questions, on se propose de retrouver le résultat précédent en raisonnant par substitution.

a) Sachant que $\sum_{i=1}^{r} x_i = 1$, écrire $\mathbf{E}(X)$ comme une fonction, notée $f : (]0;1[)^{r-1} \to \mathbb{R}$, des $(r-1)$ variables $x_1,\ldots,x_{r-1}$.

b) Montrer que f est de classe $\mathscr{C}^2$ sur l'ouvert $(]0;1[)^{r-1}$. Expliciter les dérivées partielles d'ordre 1 de f.

c) Vérifier que le seul point critique de f est $a = \left(\frac{1}{r}, \frac{1}{r}, \ldots, \frac{1}{r}\right)$.

d) Soit J la matrice de $\mathscr{M}_{r-1}(\mathbb{R})$ dont tous les éléments sont égaux à 1. Vérifier que la matrice hessienne de f en a s'écrit sous la forme

$$\nabla^2 f(a) = n(n-1)\left(\frac{r-1}{r}\right)^{n-2} (I_{r-1} + J).$$

e) En remarquant que les valeurs propres de J sont 0 et $r-1$, déduire les valeurs propre de $\nabla^2 f(a)$.

f) Justifier que f présente un minimum local au point a. Conclure sur la valeur de $\mathbf{E}(X)$ correspondant à ce minimum local.

Exercice 53. ✦✦✦ ✎ **Entropie dans le cas discret**

Soit h_1 la fonction définie sur $[0,1]$ par : $$h_1(x) = \begin{cases} -x\ln x & \text{si } x \in]0,1] \\ 0 & \text{si } x = 0 \end{cases}.$$

1. Vérifier que h_1 est continue sur $[0;1]$ et donner son graphe.

Soient $n \in \mathbb{N}^*$ et Y, une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbf{P})$ avec $Y(\Omega) = \{y_1, \dots, y_n\}$. On note pour tout $k \in [\![1;n]\!]$, $p_k = \mathbf{P}(Y = y_k)$, puis on définit l'entropie de Y par

$$H(Y) = \sum_{k=1}^{n} h_1(p_k).$$

2. Calculer l'entropie d'une variable aléatoire qui suit une loi uniforme sur $[\![1;n]\!]$.

3. Pour quelle valeur du paramètre p, l'entropie d'une loi de Bernoulli $\mathscr{B}(p)$ est maximale ?

4. L'objectif de la suite est de déterminer parmi les variables aléatoires prenant n valeurs distinctes, celles qui maximise l'entropie. Pour cela, on définit la fonction h_n sur l'ouvert $]0;1[^n$ par

$$\forall x = (x_1, \dots, x_n) \in]0;1[^n, \qquad h_n(x) = \sum_{k=1}^{n} h_1(x_k).$$

a) Justifier que h_n est de classe $\mathscr{C}^2$ sur $]0,1[^n$ et expliciter son gradient ∇h_n et sa matrice hessienne $\nabla^2 h_n$.

b) Vérifier que la fonction h_n admet un unique point critique a sous la contrainte $\sum_{k=1}^{n} x_k = 1$.

c) Soit $x \in]0,1[^n$ tel que $\sum_{k=1}^{n} x_k = 1$ et notons $h = x - a$. Vérifier que, pour tout $t \in [0,1]$, on a $a + th \in]0,1[^n$.

On note alors ψ la fonction définie sur $[0,1]$ par : $\psi(t) = h_n(a + th)$.

d) 🔧 En utilisant la formule de Taylor avec reste intégral à l'ordre 1 pour ψ entre les points 0 et 1, montrer que h_n admet en a un maximum global sous la contrainte $x_1 + \cdots + x_n = 1$. Ce maximum est-il atteint en d'autres points que a ?

5. Parmi les variables aléatoires prenant n valeurs (chacune avec une probabilité non nulle), quelles sont les lois de celles qui ont la plus grande entropie ?

Toute bonne chose a une fin, sauf le saucisson qui en a deux.

PROVERBE DANOIS