



PRÉPASIMPLE

prepasimple.fr

HEC / ESCP

Mathématiques 2

1987

Corrigé de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Corrigé

PARTIE I

1. Minimisation de la somme des distances verticales

(a) Distances aux points projetés

Une projection dans la direction de l'axe des ordonnées conserve l'abscisse. Ainsi :

$$m_1 = (0; b), \quad m_2 = (1; a + b), \quad m_3 = (-2; b - 2a).$$

Les distances sont donc les valeurs absolues des différences d'ordonnées :

$$M_1 m_1 = |b|, \quad M_2 m_2 = |a + b - 1|, \quad M_3 m_3 = |b - 2a|.$$

Par conséquent :

$$f(a; b) = |b| + |a + b - 1| + |2a - b| = |b| + \varphi(a).$$

(b) Étude de la fonction φ

Méthode 1 — Fonctions affines par morceaux

Les deux changements de signe ont lieu en $x = \frac{b}{2}$ et $x = 1 - b$. Leur ordre est donné par :

$$\frac{b}{2} < 1 - b \iff b < \frac{2}{3}.$$

Si $b < \frac{2}{3}$, les expressions successives sont :

$$\varphi(x) = \begin{cases} 1 - 3x & \text{si } x \leq \frac{b}{2}, \\ x + 1 - 2b & \text{si } \frac{b}{2} \leq x \leq 1 - b, \\ 3x - 1 & \text{si } x \geq 1 - b. \end{cases}$$

La courbe est une ligne brisée de pentes successives -3 , 1 et 3 . Ses sommets sont :

$$\left(\frac{b}{2}; 1 - \frac{3b}{2}\right) \quad \text{et} \quad (1 - b; 2 - 3b).$$

Elle est strictement décroissante jusqu'à $\frac{b}{2}$, puis strictement croissante. Son minimum est unique.

Si $b = \frac{2}{3}$, les deux changements de signe coïncident et :

$$\varphi(x) = 3 \left| x - \frac{1}{3} \right|.$$

La courbe comporte deux demi-droites de pentes -3 et 3 et de sommet $\left(\frac{1}{3}; 0\right)$. Son minimum est unique en $x = \frac{b}{2}$.

Si $b > \frac{2}{3}$, on a $1 - b < \frac{b}{2}$ et :

$$\varphi(x) = \begin{cases} 1 - 3x & \text{si } x \leq 1 - b, \\ 2b - 1 - x & \text{si } 1 - b \leq x \leq \frac{b}{2}, \\ 3x - 1 & \text{si } x \geq \frac{b}{2}. \end{cases}$$

Les pentes successives sont -3 , -1 et 3 . Les sommets sont :

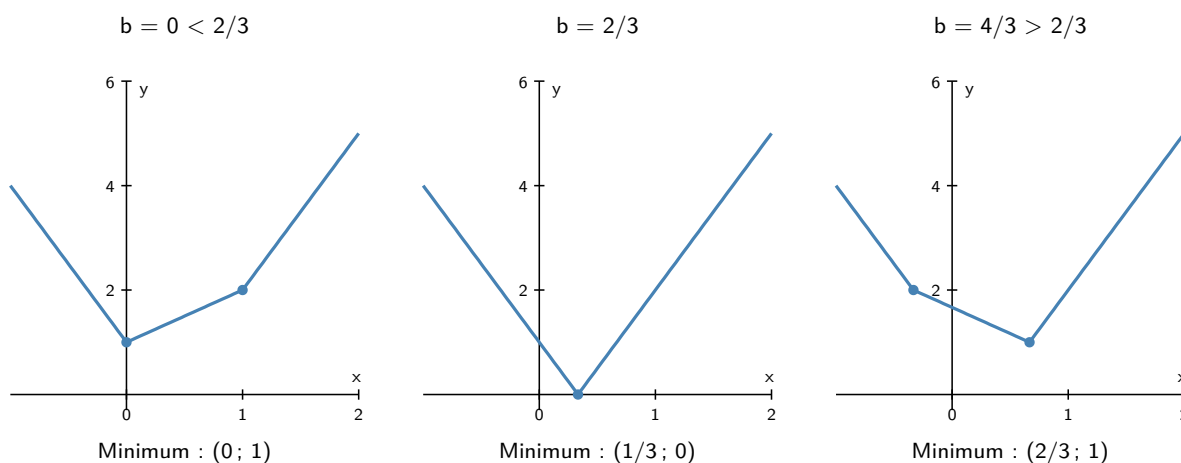
$$(1 - b; 3b - 2) \quad \text{et} \quad \left(\frac{b}{2}; \frac{3b}{2} - 1\right).$$

La fonction est strictement décroissante jusqu'à $\frac{b}{2}$, puis strictement croissante. Dans les trois cas :

$$\min_{x \in \mathbb{R}} \varphi(x) = \varphi\left(\frac{b}{2}\right) = \left|\frac{3b}{2} - 1\right|,$$

et le minimum est atteint uniquement pour $x = \frac{b}{2}$.

Les trois tracés ci-dessous illustrent exactement les trois formes précédentes, avec les valeurs respectives $b = 0$, $b = \frac{2}{3}$ et $b = \frac{4}{3}$. Les formules par morceaux donnent le tracé pour toute autre valeur de b .



Méthode 2 — Inégalité triangulaire

Cette méthode confirme le minimum sans découpage en intervalles. Posons $u = x - \frac{b}{2}$ et $c = \frac{3b}{2} - 1$. Alors :

$$\varphi(x) = 2|u| + |u + c|.$$

L'inégalité triangulaire $|c| = |(u + c) - u| \leq |u + c| + |u|$ donne :

$$\varphi(x) \geq |c| + |u| \geq |c|.$$

Pour $u = 0$, on a l'égalité $\varphi(x) = |c|$. Pour $u \neq 0$, la première minoration entraîne $\varphi(x) > |c|$. Le minimum est donc unique en $x = \frac{b}{2}$. L'étude par morceaux décrit la courbe ; la majoration triangulaire permet de reconnaître rapidement le minimiseur.

(c) Minimum global de f et droite D_1

Méthode 1 — Minimisations successives

D'après la question précédente, pour chaque b , l'unique pente minimisante est $a = \frac{b}{2}$. Il reste à minimiser :

$$F(b) = |b| + \left| \frac{3b}{2} - 1 \right|.$$

Les changements de signe se produisent en 0 et $\frac{2}{3}$:

$$F(b) = \begin{cases} 1 - \frac{5}{2}b & \text{si } b \leq 0, \\ 1 - \frac{1}{2}b & \text{si } 0 \leq b \leq \frac{2}{3}, \\ \frac{5}{2}b - 1 & \text{si } b \geq \frac{2}{3}. \end{cases}$$

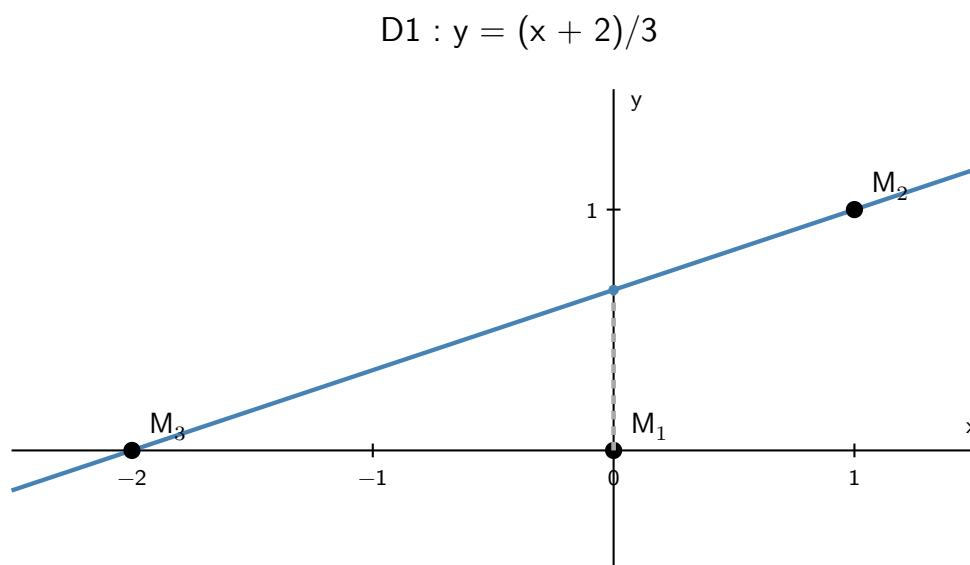
La fonction est strictement décroissante sur $] -\infty; \frac{2}{3}]$, puis strictement croissante sur $[\frac{2}{3}; +\infty[$. Ainsi :

$$(a_1; b_1) = \left(\frac{1}{3}; \frac{2}{3}\right), \quad \min f = \frac{2}{3}.$$

Les deux minimisations sont uniques, donc le couple minimisant est unique. La droite est :

$$D_1 : y = \frac{1}{3}x + \frac{2}{3}.$$

Elle passe par M_2 et M_3 ; seule la distance verticale issue de M_1 est non nulle.



Traits pointillés : distances verticales aux trois observations.

Méthode 2 — Une relation entre les résidus

Posons $e_1 = b$, $e_2 = a + b - 1$ et $e_3 = b - 2a$. Un calcul direct donne :

$$3e_1 - 2e_2 - e_3 = 2.$$

Ainsi :

$$2 \leq 3|e_1| + 2|e_2| + |e_3| \leq 3(|e_1| + |e_2| + |e_3|) = 3f(a; b).$$

Donc $f(a; b) \geq \frac{2}{3}$. Pour atteindre cette borne, il faut notamment que la seconde inégalité soit une égalité. La différence entre ses deux membres est $|e_2| + 2|e_3|$, donc $e_2 = e_3 = 0$. Ces équations donnent $a = \frac{1}{3}$ et $b = \frac{2}{3}$; pour ce couple, $e_1 = \frac{2}{3}$ et la borne est effectivement atteinte. Cette approche donne simultanément la borne et l'unicité.

2. Minimisation de la plus grande distance verticale

(a) Les trois bandes du plan

Les ensembles sont les bandes fermées délimitées respectivement par :

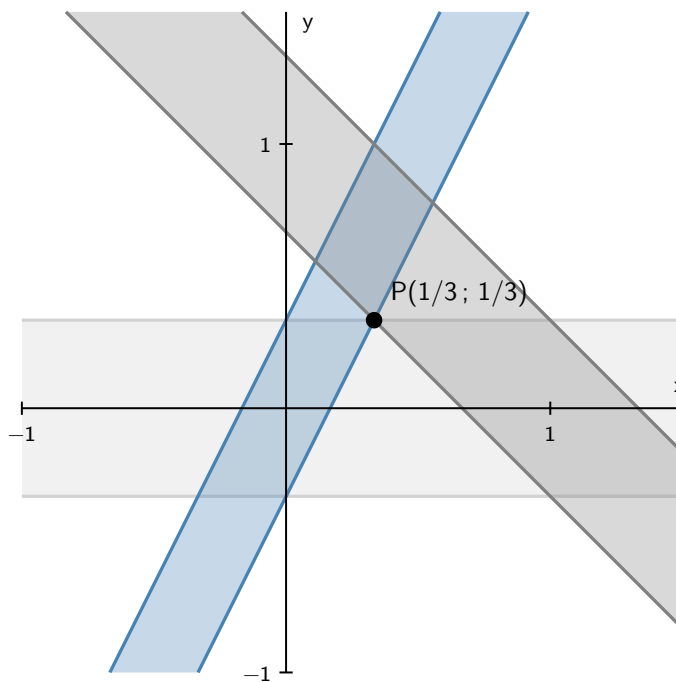
$$E_1 : -\frac{1}{3} \leq y \leq \frac{1}{3},$$

$$E_2 : 2x - \frac{1}{3} \leq y \leq 2x + \frac{1}{3},$$

$$E_3 : \frac{2}{3} - x \leq y \leq \frac{4}{3} - x.$$

La figure représente ces trois bandes dans une même fenêtre; elles se prolongent au-delà de celle-ci. Toutes leurs frontières sont incluses.

Trois bandes fermées dans le plan $(x; y)$



$$E_1 : |y| \leq 1/3$$

$$E_2 : |y - 2x| \leq 1/3$$

$$E_3 : |x + y - 1| \leq 1/3$$

$P(1/3; 1/3)$

L'intersection des trois bandes se réduit au point P.

Pour déterminer l'intersection exactement, soit $(x; y) \in E_1 \cap E_2 \cap E_3$. Les inégalités $y \leq \frac{1}{3}$ et $y - 2x \geq -\frac{1}{3}$ donnent :

$$x \leq \frac{y + \frac{1}{3}}{2} \leq \frac{1}{3}.$$

Par ailleurs, $x + y \geq \frac{2}{3}$ et $y \leq \frac{1}{3}$ donnent :

$$x \geq \frac{2}{3} - y \geq \frac{1}{3}.$$

Ainsi $x = \frac{1}{3}$. Puis $x + y \geq \frac{2}{3}$ impose $y \geq \frac{1}{3}$, donc $y = \frac{1}{3}$. Réciproquement, ce point vérifie les trois contraintes. Finalement :

$$E_1 \cap E_2 \cap E_3 = \left\{ \left(\frac{1}{3}; \frac{1}{3} \right) \right\}.$$

(b) Minimum global de g et droite D_2

Méthode 1 — Utilisation de l'intersection des bandes

On a :

$$g(a; b) = \max(|b|, |b - 2a|, |a + b - 1|).$$

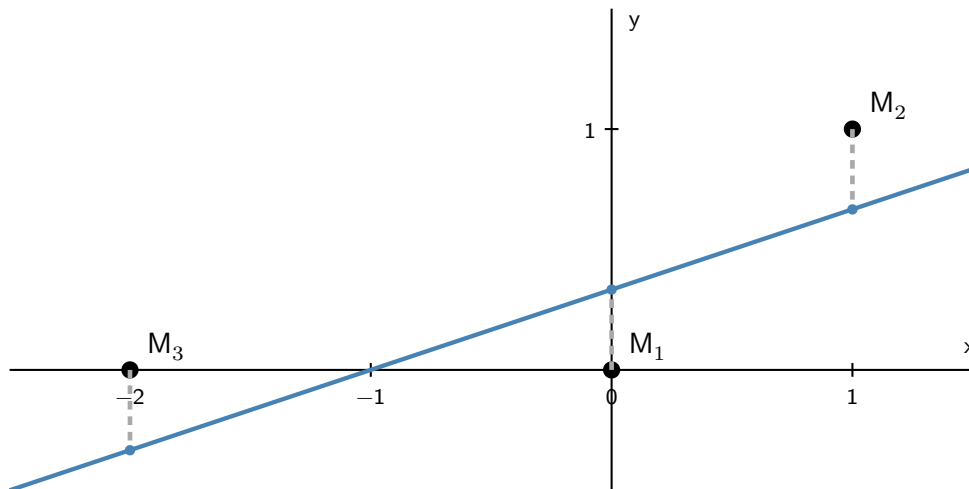
Par conséquent, $g(a; b) \leq \frac{1}{3}$ équivaut à $(a; b) \in E_1 \cap E_2 \cap E_3$. L'unique couple satisfaisant cette condition est $\left(\frac{1}{3}; \frac{1}{3} \right)$, et ses trois distances valent $\frac{1}{3}$.

S'il existait un couple tel que $g(a; b) < \frac{1}{3}$, il appartiendrait aussi à cette intersection et serait donc ce même couple, ce qui contredirait la valeur de g . D'où :

$$(a_2; b_2) = \left(\frac{1}{3}; \frac{1}{3} \right), \quad \min g = \frac{1}{3},$$

$$D_2 : \quad y = \frac{1}{3}x + \frac{1}{3}.$$

$$D2 : y = (x + 1)/3$$



Traits pointillés : distances verticales aux trois observations.

Méthode 2 — Majoration à partir des résidus

Avec les résidus définis en I.1.c, $3e_1 - 2e_2 - e_3 = 2$. Si $t = g(a; b)$, chaque $|e_j|$ est inférieur ou égal à t , donc :

$$2 \leq 3|e_1| + 2|e_2| + |e_3| \leq 6t.$$

Ainsi $g(a; b) \geq \frac{1}{3}$. Pour $t = \frac{1}{3}$, l'égalité dans cette chaîne impose $|e_1| = |e_2| = |e_3| = \frac{1}{3}$ et $e_1 \geq 0, e_2 \leq 0, e_3 \leq 0$. Donc $e_1 = \frac{1}{3}$ et $e_2 = e_3 = -\frac{1}{3}$. On obtient uniquement $b = \frac{1}{3}$ et $a = \frac{1}{3}$. Le couple atteint effectivement la borne. La méthode géométrique représente les contraintes ; la méthode des résidus établit directement le seuil optimal.

3. Ajustement par les moindres carrés verticaux

Question adaptée au programme actuel : le préambule définit la droite de régression par la minimisation de la somme des carrés des résidus, sans supposer connus ses

coefficients.

Méthode 1 — Mise sous forme de sommes de carrés

Les distances calculées en I.1.a donnent :

$$h(a; b) = b^2 + (a + b - 1)^2 + (b - 2a)^2 = 5a^2 + 3b^2 - 2ab - 2a - 2b + 1.$$

Pour a fixé, complétons le carré en b :

$$h(a; b) = 3 \left(b - \frac{a+1}{3} \right)^2 + \frac{2}{3}(7a^2 - 4a + 1).$$

Le coefficient 3 étant strictement positif, le minimum en b est unique et atteint en $b = \frac{a+1}{3}$. De plus :

$$7a^2 - 4a + 1 = 7 \left(a - \frac{2}{7} \right)^2 + \frac{3}{7}.$$

Ainsi :

$$h(a; b) = 3 \left(b - \frac{a+1}{3} \right)^2 + \frac{14}{3} \left(a - \frac{2}{7} \right)^2 + \frac{2}{7}.$$

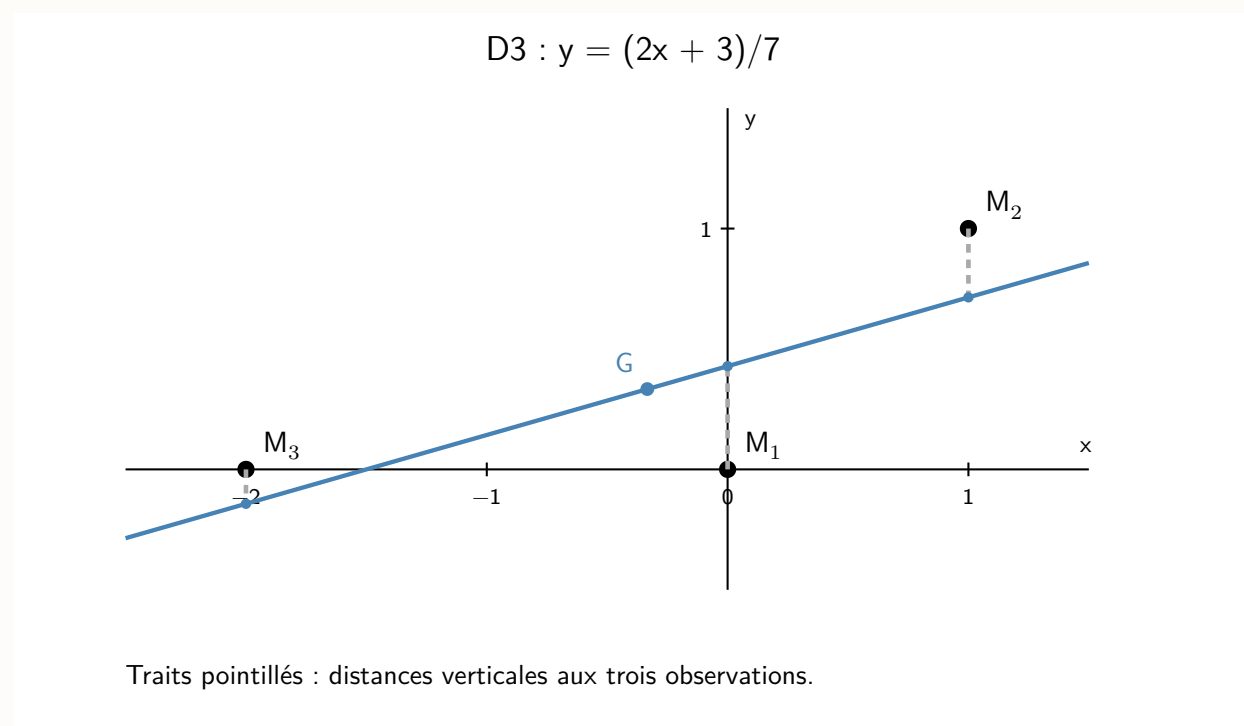
Les deux carrés sont nuls simultanément uniquement pour :

$$(a_3; b_3) = \left(\frac{2}{7}; \frac{3}{7} \right), \quad \min h = \frac{2}{7}.$$

La droite recherchée est donc :

$$D_3 : y = \frac{2}{7}x + \frac{3}{7}.$$

Par définition, c'est la droite de régression de Y par rapport à X pour le nuage des trois observations. Elle passe par le point moyen $G\left(-\frac{1}{3}; \frac{1}{3}\right)$.



Méthode 2 — Gradient et hessienne

La fonction h est polynomiale, donc de classe C^2 sur l'ouvert convexe $\mathbb{R}^2$. Son gradient et sa hessienne sont :

$$\nabla h(a; b) = \begin{pmatrix} 10a - 2b - 2 \\ 6b - 2a - 2 \end{pmatrix}, \quad H = \begin{pmatrix} 10 & -2 \\ -2 & 6 \end{pmatrix}.$$

L'annulation du gradient donne $b = \frac{a+1}{3}$ puis $\frac{28a-8}{3} = 0$. Il existe donc un unique

point critique, $\left(\frac{2}{7}; \frac{3}{7}\right)$.

Pour tout vecteur non nul de coordonnées $(s; t)$, la forme quadratique de la hessienne vérifie :

$$10s^2 - 4st + 6t^2 = 10 \left(s - \frac{t}{5}\right)^2 + \frac{28}{5}t^2 > 0.$$

La matrice H est symétrique. Si λ est une de ses valeurs propres et Z un vecteur propre non nul, le produit scalaire $\langle HZ, Z \rangle = \lambda \|Z\|^2$ est strictement positif. Toutes ses valeurs propres sont donc strictement positives. Comme cette hessienne est constante sur $\mathbb{R}^2$, le critère global sur un ouvert convexe donne un minimum global au point critique.

L'unicité se vérifie aussi sans résultat supplémentaire : pour un accroissement non nul $(s; t)$, le développement exact du polynôme au point critique donne :

$$h\left(\frac{2}{7} + s; \frac{3}{7} + t\right) - h\left(\frac{2}{7}; \frac{3}{7}\right) = 5s^2 - 2st + 3t^2 > 0.$$

Le minimum est donc unique et vaut $\frac{2}{7}$. La première méthode répond directement à la minimisation successive demandée ; la seconde illustre l'usage conforme de la hessienne.

4. Ajustement dans une direction de pente m

(a) Coordonnées des projections et expression de h_m

Le point de la droite passant par $M_j(x_j; y_j)$ et de direction $(1; m)$ a pour coordonnées $(x_j + t_j; y_j + mt_j)$. Son appartenance à D impose :

$$y_j + mt_j = a(x_j + t_j) + b.$$

Puisque $a \neq m$, cette équation possède l'unique solution :

$$t_j = \frac{y_j - ax_j - b}{a - m}.$$

Les coordonnées demandées sont donc :

$$m_1 = \left(\frac{-b}{a - m}; \frac{-mb}{a - m} \right),$$

$$m_2 = \left(\frac{1 - m - b}{a - m}; \frac{a - ma - mb}{a - m} \right),$$

$$m_3 = \left(\frac{2m - b}{a - m}; \frac{m(2a - b)}{a - m} \right).$$

Le vecteur $\overrightarrow{M_j m_j}$ a pour coordonnées $(t_j; mt_j)$, donc sa norme au carré vaut $(1 + m^2)t_j^2$.
En sommant :

$$h_m(a; b) = \frac{1 + m^2}{(a - m)^2} [b^2 + (a + b - 1)^2 + (b - 2a)^2] = \frac{1 + m^2}{(a - m)^2} h(a; b).$$

(b) Cas $m \neq \frac{2}{7}$

Pour $a \neq m$ fixé, le facteur $\frac{1 + m^2}{(a - m)^2}$ est strictement positif et indépendant de b . D'après I.3, l'unique minimum en b est atteint pour :

$$b = \frac{a + 1}{3}.$$

Sa valeur est :

$$h_m \left(a; \frac{a+1}{3} \right) = \frac{2}{3}(1+m^2) \frac{7a^2 - 4a + 1}{(a-m)^2} = \frac{2}{3}(1+m^2)\theta(a).$$

Minimiser h_m revient donc à minimiser θ sur $\mathbb{R} \setminus \{m\}$.

Méthode 1 — Étude des variations

Sur chacun des intervalles $] -\infty; m[$ et $]m; +\infty[$, la fonction est dérivable et :

$$\theta'(a) = \frac{(14a-4)(a-m) - 2(7a^2 - 4a + 1)}{(a-m)^3} = \frac{2((2-7m)a + 2m-1)}{(a-m)^3}.$$

Posons :

$$Q = 7m^2 - 4m + 1 = 7 \left(m - \frac{2}{7} \right)^2 + \frac{3}{7} > 0,$$

$$c = 2 - 7m \neq 0, \quad a_* = \frac{1 - 2m}{2 - 7m}.$$

Alors $\theta'(a) = \frac{2c(a-a_*)}{(a-m)^3}$ et :

$$a_* - m = \frac{Q}{c}.$$

En particulier, $a_* \neq m$, donc ce point est admissible. Les limites sont :

$$\lim_{a \rightarrow -\infty} \theta(a) = \lim_{a \rightarrow +\infty} \theta(a) = 7, \quad \lim_{a \rightarrow m^-} \theta(a) = \lim_{a \rightarrow m^+} \theta(a) = +\infty.$$

Les limites au voisinage de m résultent de la stricte positivité du numérateur en m , égale à Q .

Si $m < \frac{2}{7}$, alors $c > 0$ et $a_* > m$:

- sur $] -\infty; m[$, $\theta' > 0$ et θ croît de la limite 7 vers $+\infty$;
- sur $]m; a_*[$, $\theta' < 0$ et θ décroît de $+\infty$ vers $\theta(a_*)$;
- sur $]a_*; +\infty[$, $\theta' > 0$ et θ croît de $\theta(a_*)$ vers la limite 7.

Si $m > \frac{2}{7}$, alors $c < 0$ et $a_* < m$:

- sur $] -\infty; a_*[$, $\theta' < 0$ et θ décroît de la limite 7 vers $\theta(a_*)$;
- sur $]a_*; m[$, $\theta' > 0$ et θ croît de $\theta(a_*)$ vers $+\infty$;
- sur $]m; +\infty[$, $\theta' < 0$ et θ décroît de $+\infty$ vers la limite 7.

Pour calculer la valeur critique, écrivons

$$7a^2 - 4a + 1 = 7(a - m)^2 + (14m - 4)(a - m) + Q. \text{ Comme } 14m - 4 = -2c \text{ et } a_* - m = \frac{Q}{c} :$$

$$\theta(a_*) = 7 - \frac{2c^2}{Q} + \frac{c^2}{Q} = 7 - \frac{c^2}{Q} = \frac{3}{Q},$$

la dernière égalité utilisant $7Q - c^2 = 3$. De plus, $c \neq 0$ implique $\frac{3}{Q} < 7$. Dans les deux configurations, a_* fournit donc l'unique minimum global.

On en déduit :

$$a_4 = \frac{1 - 2m}{2 - 7m}, \quad b_4 = \frac{a_4 + 1}{3} = \frac{1 - 3m}{2 - 7m},$$

$$\min h_m = \frac{2(1+m^2)}{7m^2 - 4m + 1}.$$

La droite D_4 existe et est unique pour tout $m \neq \frac{2}{7}$.

Méthode 2 — Identité de comparaison avec la valeur minimale

Une identité fournit directement le minimum, tout en laissant à la première méthode l'étude des variations demandée. Pour $a \neq m$:

$$\theta(a) - \frac{3}{Q} = \frac{Q(7a^2 - 4a + 1) - 3(a-m)^2}{Q(a-m)^2}.$$

Or $7Q - 3 = c^2$, $-4Q + 6m = -2c(1 - 2m)$ et $Q - 3m^2 = (1 - 2m)^2$. Ainsi :

$$Q(7a^2 - 4a + 1) - 3(a-m)^2 = (ca - (1 - 2m))^2.$$

Il vient :

$$\theta(a) - \frac{3}{Q} = \frac{(ca - (1 - 2m))^2}{Q(a-m)^2} \geq 0.$$

Puisque $Q > 0$ et $a \neq m$, l'égalité équivaut à $a = a_*$. Ce point est admissible puisque $a_* - m = \frac{Q}{c} \neq 0$. On retrouve exactement le couple et la valeur minimale précédents. Cette identité est particulièrement utile pour prouver le caractère global et l'unicité d'un minimum rationnel.

(c) Cas $m = \frac{2}{7}$

La minimisation en b est inchangée : pour chaque $a \neq \frac{2}{7}$, elle impose $b = \frac{a+1}{3}$. En revanche :

$$\theta(a) = \frac{7 \left(a - \frac{2}{7}\right)^2 + \frac{3}{7}}{\left(a - \frac{2}{7}\right)^2} = 7 + \frac{3}{7 \left(a - \frac{2}{7}\right)^2} > 7.$$

La valeur 7 est approchée lorsque $|a|$ tend vers $+\infty$, mais n'est jamais atteinte. Plus précisément, $\theta'(a) = -\frac{6}{7 \left(a - \frac{2}{7}\right)^3}$: θ est croissante sur $]-\infty; \frac{2}{7}[$ et décroissante sur $]\frac{2}{7}; +\infty[$. Elle n'admet aucun minimum sur son domaine.

Comme $\frac{2}{3}(1 + m^2) = \frac{106}{147}$, on obtient :

$$\inf_{a \neq \frac{2}{7}, b \in \mathbb{R}} h_{2/7}(a; b) = \frac{106}{21}.$$

Cette borne n'est pas atteinte : tout couple admissible donne une valeur strictement supérieure. Il n'existe donc aucune droite D_4 de la forme $y = ax + b$ réalisant le minimum dans ce cas.

Pour comprendre géométriquement cette absence, les droites optimisées en b passent toutes par $G\left(-\frac{1}{3}; \frac{1}{3}\right)$. Lorsque leurs pentes tendent vers l'infini, elles se rapprochent de la droite verticale $x = -\frac{1}{3}$, qui n'appartient pas à la famille $y = ax + b$. Les projections sur cette verticale dans la direction $(1; \frac{2}{7})$ auraient une somme des distances au carré égale à :

$$\left(1 + \frac{4}{49}\right) \sum_{j=1}^3 \left(x_j + \frac{1}{3}\right)^2 = \frac{53}{49} \times \frac{14}{3} = \frac{106}{21}.$$

Ce calcul explique la borne inférieure, sans fournir de minimiseur admissible.

PARTIE II

1. Effet d'une observation sur la corrélation

(a) Coefficient de corrélation des trois observations

Question adaptée au programme actuel : les définitions statistiques nécessaires, avec la convention de variance empirique, sont rappelées au préambule du sujet.

Les sommes portent ici sur les trois points et les variances empiriques utilisent le facteur $\frac{1}{3}$. On a :

$$\bar{x} = \frac{0 + 1 - 2}{3} = -\frac{1}{3}, \quad \bar{y} = \frac{0 + 1 + 0}{3} = \frac{1}{3}.$$

En développant les carrés et en utilisant les définitions des moyennes, on obtient les formules de calcul :

$$V_X = \frac{1}{3} \sum_{j=1}^3 x_j^2 - \bar{x}^2, \quad V_Y = \frac{1}{3} \sum_{j=1}^3 y_j^2 - \bar{y}^2,$$

$$C_{XY} = \frac{1}{3} \sum_{j=1}^3 x_j y_j - \bar{x} \bar{y}.$$

Ainsi :

$$V_X = \frac{5}{3} - \frac{1}{9} = \frac{14}{9} > 0, \quad V_Y = \frac{1}{3} - \frac{1}{9} = \frac{2}{9} > 0,$$

$$C_{XY} = \frac{1}{3} + \frac{1}{9} = \frac{4}{9}.$$

Les variances étant strictement positives, le coefficient de corrélation est défini et vaut :

$$\rho = \frac{\frac{4}{9}}{\sqrt{\frac{14}{9} \times \frac{2}{9}}} = \frac{2}{\sqrt{7}}.$$

(b) Coefficient après ajout de M_0

Pour alléger les calculs, posons $u = x_0$ et $v = y_0$. Les moyennes sur les quatre observations sont :

$$\bar{x} = \frac{u-1}{4}, \quad \bar{y} = \frac{v+1}{4}.$$

Les mêmes formules de calcul, avec le facteur $\frac{1}{4}$, donnent :

$$V_X = \frac{u^2+5}{4} - \frac{(u-1)^2}{16} = \frac{3u^2+2u+19}{16},$$

$$V_Y = \frac{v^2+1}{4} - \frac{(v+1)^2}{16} = \frac{3v^2-2v+3}{16},$$

$$C_{XY} = \frac{uv+1}{4} - \frac{(u-1)(v+1)}{16} = \frac{3uv-u+v+5}{16}.$$

Le coefficient est défini pour tout $(u; v) \in \mathbb{R}^2$, car :

$$3u^2+2u+19 = 3\left(u+\frac{1}{3}\right)^2 + \frac{56}{3} > 0,$$

$$3v^2-2v+3 = 3\left(v-\frac{1}{3}\right)^2 + \frac{8}{3} > 0.$$

Finalement :

$$\rho(M_0) = \frac{3x_0y_0 - x_0 + y_0 + 5}{\sqrt{(3x_0^2 + 2x_0 + 19)(3y_0^2 - 2y_0 + 3)}}.$$

(c) Valeurs accessibles et impossibilité d'une corrélation parfaite

En remplaçant $(u; v)$ par $(5 + k; k)$ puis $(5 + k; -k)$ dans la formule précédente, on obtient :

$$\alpha(k) = \frac{3k^2 + 15k}{\sqrt{(3k^2 + 32k + 104)(3k^2 - 2k + 3)}},$$

$$\beta(k) = \frac{-3k^2 - 17k}{\sqrt{(3k^2 + 32k + 104)(3k^2 + 2k + 3)}}.$$

Les facteurs sous les racines sont strictement positifs pour tout k , comme le montre II.1.b appliquée à ces coordonnées. Par continuité des polynômes, de la racine carrée sur $]0; +\infty[$ et du quotient de fonctions continues à dénominateur non nul, α et β sont continues sur $\mathbb{R}$. De plus :

$$\alpha(0) = \beta(0) = 0.$$

Lorsque $k \rightarrow +\infty$, on peut supposer $k > 0$ et diviser les numérateurs et les facteurs du dénominateur par k^2 :

$$\alpha(k) = \frac{3 + \frac{15}{k}}{\sqrt{\left(3 + \frac{32}{k} + \frac{104}{k^2}\right)\left(3 - \frac{2}{k} + \frac{3}{k^2}\right)}} \rightarrow 1,$$

$$\beta(k) = \frac{-3 - \frac{17}{k}}{\sqrt{\left(3 + \frac{32}{k} + \frac{104}{k^2}\right) \left(3 + \frac{2}{k} + \frac{3}{k^2}\right)}} \rightarrow -1.$$

Soit $r \in]0; 1[$. Puisque $\alpha(k) \rightarrow 1$, il existe $K > 0$ tel que $\alpha(K) > r$. La fonction α est continue sur $[0; K]$, avec $\alpha(0) = 0 < r < \alpha(K)$. Le théorème des valeurs intermédiaires fournit donc $k \in]0; K[$ tel que $\alpha(k) = r$. Le point $M_0 = E_k$ convient.

Soit maintenant $r \in]-1; 0[$. Puisque $\beta(k) \rightarrow -1$, il existe $K > 0$ tel que $\beta(K) < r$. Par continuité sur $[0; K]$ et parce que $\beta(K) < r < \beta(0) = 0$, le même théorème fournit $k \in]0; K[$ tel que $\beta(k) = r$. Le point $M_0 = F_k$ convient.

Pour $r = 0$, le point $M_0 = (5; 0)$ convient. Ainsi toute valeur de $] -1; 1[$ est atteinte.

Pour les valeurs extrêmes, deux méthodes permettent de conclure.

Méthode 1 — Cas d'égalité de Cauchy-Schwarz

Considérons les deux vecteurs de $\mathbb{R}^4$:

$$U = (u - \bar{x}, -\bar{x}, 1 - \bar{x}, -2 - \bar{x}),$$

$$V = (v - \bar{y}, -\bar{y}, 1 - \bar{y}, -\bar{y}).$$

Les variances sont strictement positives, donc U et V sont non nuls. Le coefficient vaut $\frac{\langle U, V \rangle}{\|U\| \|V\|}$. Cauchy-Schwarz donne $|\rho(M_0)| \leq 1$, avec égalité uniquement si les vecteurs sont colinéaires. Comme $U \neq 0$, cette colinéarité imposerait $V = \lambda U$ pour un réel λ .

Chaque observation vérifierait alors $y_j = \lambda x_j + \bar{y} - \lambda \bar{x}$: tous les points seraient sur une même droite. Or $M_1(0; 0)$ et $M_3(-2; 0)$ imposent la droite $y = 0$, à laquelle $M_2(1; 1)$

n'appartient pas. Cette égalité est donc impossible, indépendamment de M_0 .

Méthode 2 — Une identité strictement positive

Posons $A = 3u^2 + 2u + 19$, $B = 3v^2 - 2v + 3$ et $C = 3uv - u + v + 5$. En développant puis en complétant les carrés :

$$\begin{aligned} AB - C^2 &= 8(u^2 + 7v^2 - 4uv + 2u - 6v + 4) \\ &= 8 \left[(u - 2v + 1)^2 + 3 \left(v - \frac{1}{3} \right)^2 + \frac{8}{3} \right] > 0. \end{aligned}$$

Puisque $AB > 0$, il vient $\frac{C^2}{AB} < 1$, c'est-à-dire $|\rho(M_0)| < 1$. Cette vérification algébrique ne nécessite pas d'interpréter le cas d'égalité ; la première méthode explique pourquoi le défaut d'alignement empêche une corrélation parfaite.

Il n'existe donc aucun point donnant $\rho(M_0) = 1$ ou $\rho(M_0) = -1$. L'ensemble exact des valeurs accessibles est $] -1; 1[$.

2. Prescrire la droite de régression après ajout d'un point

Question adaptée au programme actuel : les formules de régression utilisées ci-dessous sont établies à partir de la minimisation, sans les supposer connues.

Pour un nuage fini, notons $\bar{x}$ et $\bar{y}$ ses moyennes et définissons les sommes centrées :

$$S_{xx} = \sum_j (x_j - \bar{x})^2, \quad S_{xy} = \sum_j (x_j - \bar{x})(y_j - \bar{y}), \quad S_{yy} = \sum_j (y_j - \bar{y})^2.$$

Pour une droite de pente p et d'ordonnée à l'origine q , le résidu s'écrit :

$$y_j - px_j - q = (y_j - \bar{y}) - p(x_j - \bar{x}) + (\bar{y} - p\bar{x} - q).$$

Les sommes de $x_j - \bar{x}$ et de $y_j - \bar{y}$ sont nulles. En développant les carrés et en sommant, les termes croisés avec la dernière constante disparaissent :

$$\sum_j (y_j - px_j - q)^2 = S_{yy} - 2pS_{xy} + p^2S_{xx} + N(\bar{y} - p\bar{x} - q)^2.$$

Si $S_{xx} > 0$, la mise sous forme canonique donne :

$$\sum_j (y_j - px_j - q)^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}} + S_{xx} \left(p - \frac{S_{xy}}{S_{xx}} \right)^2 + N(\bar{y} - p\bar{x} - q)^2.$$

Cette somme est minimisée par un unique couple :

$$p = \frac{S_{xy}}{S_{xx}} = \frac{C_{XY}}{V_X}, \quad q = \bar{y} - p\bar{x}.$$

Pour les quatre points du problème, $V_X > 0$ d'après II.1.b, donc cette caractérisation est applicable. En reprenant $u = x_0$, $v = y_0$, elle donne :

$$p = \frac{3uv - u + v + 5}{3u^2 + 2u + 19}, \quad q = \frac{v + 1 - p(u - 1)}{4}.$$

Ainsi $\Delta = D$, où D a pour équation $y = ax + b$, équivaut exactement au système :

$$\begin{cases} 3uv - u + v + 5 = a(3u^2 + 2u + 19), \\ v + 1 = a(u - 1) + 4b. \end{cases}$$

La seconde équation s'écrit $v = au - a + 4b - 1$. En la substituant dans la première :

$$3u(-a + 4b - 1) - u + (au - a + 4b - 1) + 5 - 2au - 19a = 0,$$

soit :

$$4(3b - a - 1)u + 4(b + 1 - 5a) = 0.$$

On obtient donc la caractérisation, valable sans restriction sur a et b :

$$\Delta = D \iff \begin{cases} (3b - a - 1)u = 5a - b - 1, \\ v = au - a + 4b - 1. \end{cases}$$

Toutes les manipulations sont réversibles : le seul dénominateur utilisé est $3u^2 + 2u + 19$, dont la stricte positivité a été vérifiée.

(a) Cas $3b - a - 1 \neq 0$

La première équation possède l'unique solution :

$$x_0 = u = \frac{5a - b - 1}{3b - a - 1}.$$

La seconde fixe alors une unique ordonnée :

$$y_0 = v = a \frac{5a - b - 1}{3b - a - 1} - a + 4b - 1.$$

Ce point est réel et satisfait le système, donc sa droite de régression est bien D .

Réciproquement, tout point donnant $\Delta = D$ doit avoir ces coordonnées. Il existe ainsi exactement un point M_0 .

(b) Cas $3b - a - 1 = 0$, avec $(a; b) \neq \left(\frac{2}{7}; \frac{3}{7}\right)$

La relation $3b - a - 1 = 0$ donne $b = \frac{a+1}{3}$. Pour que la première équation du système puisse être satisfaite, il faudrait :

$$5a - b - 1 = 5a - \frac{a+1}{3} - 1 = \frac{14a-4}{3} = 0.$$

Cela impose $a = \frac{2}{7}$ puis $b = \frac{3}{7}$, couple exclu dans cette question. Le système est donc incompatible et aucun point M_0 ne convient.

Géométriquement, les droites vérifiant $3b - a - 1 = 0$ sont celles qui passent par le point moyen initial $G\left(-\frac{1}{3}; \frac{1}{3}\right)$. Parmi elles, seule la droite de régression initiale pourra rester une droite de régression après ajout d'un point, comme le précise la question suivante.

(c) Cas $(a; b) = \left(\frac{2}{7}; \frac{3}{7}\right)$

Méthode 1 — Résolution du système

Ici, $3b - a - 1 = 0$ et $5a - b - 1 = 0$, donc la première équation du système n'impose aucune restriction sur u . La seconde devient :

$$v = \frac{2}{7}u - \frac{2}{7} + \frac{12}{7} - 1 = \frac{2}{7}u + \frac{3}{7}.$$

Ainsi :

$$\{M_0 : \Delta = D\} = \left\{ \left(u; \frac{2u+3}{7} \right) : u \in \mathbb{R} \right\} = D_3.$$

Chaque point de cette droite convient, y compris le point moyen G et les points d'intersection éventuels avec le nuage initial. Aucune autre position ne convient.

Méthode 2 — Interprétation sans calcul de coefficients

Montrons d'abord la suffisance. Supposons $M_0 \in D_3$. Pour toute droite L de la forme $y = px + q$, la somme des carrés des résidus sur les quatre points est la somme sur les trois anciens points plus le carré du résidu de M_0 . Le second terme est positif ou nul. La somme sur les trois anciens points est minimisée uniquement par D_3 , d'après I.3. Sur D_3 , le résidu ajouté est nul puisque $M_0 \in D_3$. Par conséquent, D_3 minimise encore la somme sur les quatre points, et elle reste l'unique droite de régression.

Pour la nécessité, supposons que la droite de régression des quatre points soit D_3 . Toute droite de régression passe par le point moyen du nuage, comme le montre la formule $q = \bar{y} - p\bar{x}$ établie avant II.2.a. Notons G' le point moyen des quatre observations. Alors $G' \in D_3$. Le point moyen initial G appartient également à D_3 et le lien entre les deux barycentres est :

$$\overrightarrow{GM_0} = 4\overrightarrow{GG'}.$$

Le vecteur $\overrightarrow{GG'}$ est dirigé par D_3 ; le vecteur $\overrightarrow{GM_0}$ aussi, y compris lorsque ces deux vecteurs sont nuls. Ainsi $M_0 \in D_3$.

Cette approche retrouve exactement l'ensemble des points admissibles sans recalculer les coefficients de la nouvelle régression : elle combine la définition par les moindres carrés et la propriété du point moyen.

Relevé des adaptations

- I.3, II.1.a à II.1.c et II.2.a à II.2.c : ajout, au préambule du sujet, des définitions des moyennes, variances et covariance empiriques, du coefficient de corrélation et de la droite de régression par minimisation. La normalisation des variances est explicitée ; les formules de régression sont démontrées avant II.2.a conformément au référentiel, au lieu d'être supposées connues. Les demandes originales et leur numérotation sont conservées.

- II.2.c : remplacement du symbole isolé A dans « $A = D$ » par Δ , symbole de la droite de régression introduit en II.2.a. Cette rectification typographique restaure la notation utilisée dans les deux sous-questions précédentes.

Les tracés demandés en I.1.b, I.1.c, I.2.a, I.2.b et I.3 sont intégrés au corrigé en SVG, à partir des formules exactes. Le PDF original ne comporte aucune figure mathématique à reproduire. Aucune question n'a été supprimée et aucun résultat n'a été admis.