Informatique Appliquée au Calcul Scientifique

Cours 6

Droite des moindres carrés

1) Inégalité de Cauchy-Schwarz
2) Droite des moindres carrés
3) Calcul pratique de la droite de régression
4) Introduction aux méthodes de l'optimisation

François Dubois, 17 novembre 2004, édition 03 août 2006, 12 pages.

IACS (6) $\boxed{\text{Droite des moindres carrés}}$

(1) Inégalité de Cauchy-Schwarz.

- On désigne par N un entier ≥ 1, $x = (x_j)_{1\leq j\leq N}$ un ensemble de N nombres réels et
$\xi = (\xi_j)_{1\leq j\leq N}$ une seconde famille de N réels.

Prop (1) On a l'inégalité suivante, dite de Cauchy-Schwarz :

$$(1) \quad \left(\sum_{j=1}^{N} x_j \xi_j\right)^2 \leq \left(\sum_{j=1}^{N} x_j^2\right)\left(\sum_{j=1}^{N} \xi_j^2\right).$$

Preuve de la proposition (1)

- Soit θ un nombre réel et

$$(2) \quad f(\theta) = \sum_{j=1}^{N} (x_j + \theta \xi_j)^2, \quad \theta \in \mathbb{R}.$$

on remarque immédiatement que $f(\cdot)$ est une somme de carrés, donc, un carré étant toujours positif dans le corps des nombres réels,

$$(3) \quad f(\theta) \geq 0, \quad \forall \theta \in \mathbb{R}$$

- On remarque qu'on peut aussi écrire f selon les puissances de θ, et que c'est un polynome de degré deux :

(4) $f(\theta) = a\theta^2 + 2b\theta + c$, $a = \sum_{j=1}^{N} \xi_j^2$, $b = \sum_{j=1}^{N} x_j \xi_j$, $c = \sum_{j=1}^{N} x_j^2$.

- Compte tenu de l'inégalité (3), le polynome $f(\theta)$ est toujours ≥ 0, donc il a au mieux, une racine double et son discriminant $\Delta = b^2 - ac$ est donc négatif ou nul:

(5) $\quad b^2 \leq ac$.

Quand on remplace a, b, c par leurs valeurs (voir la relation (4)), on obtient exactement l'inégalité (1) énoncée

$\square$

Prop ② Cas d'égalité dans Cauchy-Schwarz.

On désigne toujours par $x = (x_j)_{1 \leq j \leq N}$ et $\xi = (\xi_j)_{1 \leq j \leq N}$ deux familles de nombres réels
Si la relation (1) est une égalité, c'est à dire

(6) $\left(\sum_{j=1}^{N} \xi_j x_j \right)^2 = \left(\sum_{j=1}^{N} x_j^2 \right) \left(\sum_{j=1}^{N} \xi_j^2 \right)$,

alors les vecteurs x et ξ sont proportionnels:

(7) $\quad \exists \lambda \in \mathbb{R}, \ \forall j \in \{1, \dots, N\}, \quad x_j = \lambda \xi_j$

Preuve de la proposition ②

Si la relation (6) a lieu, alors $b^2 = ac$ dans la relation (5) et $f(\theta)$ de la relation (4) est en fait un carré ; $f(\theta) = a\left(\theta + \frac{b}{a}\right)^2$ qui s'annule pour

$\theta = -b/a$. Mais si on écrit $f(-b/a)$ à l'aide de la relation (2), on lit : $\sum_{j=1}^{N} \left(x_j - \frac{b}{a}\xi_j\right)^2 = 0$.
Si l'un des nombres $(x_j - \frac{b}{a}\xi_j)$ est non nul, alors la somme $\sum_{j=1}^{N} (x_j - \frac{b}{a}\xi_j)^2$ est strictement positive, ce qui contredit la propriété qu'on vient d'établir. Donc tous les $x_j - \frac{b}{a}\xi_j$ sont nuls pour tout j. Ceci montre la relation (7), avec $\lambda = \frac{b}{a}$. $\square$

- On peut réécrire les relations précédentes avec les moyennes. Si $(x_j)_{j=1,\dots,N}$ désigne une famille quelconque de nombres réels, on pose

(8) $$\langle x \rangle = \frac{1}{N} \sum_{j=1}^{N} x_j .$$

on a donc $\langle x\xi \rangle = \frac{1}{N} \sum_{j=1}^{N} x_j \xi_j$,

$$\langle x^2 \rangle = \frac{1}{N} \sum_{j=1}^{N} x_j^2 \quad \text{et} \quad \langle \xi^2 \rangle = \frac{1}{N} \sum_{j=1}^{N} \xi_j^2 .$$

En divisant les deux membres par N^2, l'inégalité (1) peut se réécrire sous la forme

(9) $$\langle x\xi \rangle^2 \leqslant \langle x^2 \rangle \langle \xi^2 \rangle .$$

- On peut utiliser la relation (9) avec $\xi_j = 1 \ \forall j$. On a alors $\langle \xi^2 \rangle = 1$ et $\langle x\xi \rangle = \langle x \rangle$. On a alors dans ce cas particulier

(10) $$\langle x \rangle^2 \leqslant \langle x^2 \rangle .$$

Dans le cas où la relation (10) est une égalité, ou encore dans le cas d'application de la proposition 2, avec $\xi_j \equiv 1$. On a donc l'existence de $\lambda \in \mathbb{R}$ de sorte que $x_j = \lambda$ $\forall j$ et alors tous les x_j sont égaux. Nous venons de démontrer la proposition suivante.

Prop 3 Soit $x = (x_j)_{1 \leq j \leq N}$ une famille de réels dont au moins deux sont différents : $\exists\, i \neq j$ tq $x_i \neq x_j$. alors

(11) $$\langle x^2 \rangle > \langle x \rangle^2 .$$

2) Droite des moindres carrés.

- On se donne toujours un entier N et des réels $(x_j)_{1 \leq j \leq N}$ qu'on suppose tous distincts. On peut les ordonner pour fixer les idées :

(12) $$x_1 < x_2 < \dots < x_j < x_{j+1} < \dots < x_N .$$

On se donne aussi une famille "quelconque" de réels $(y_j)_{1 \leq j \leq N}$

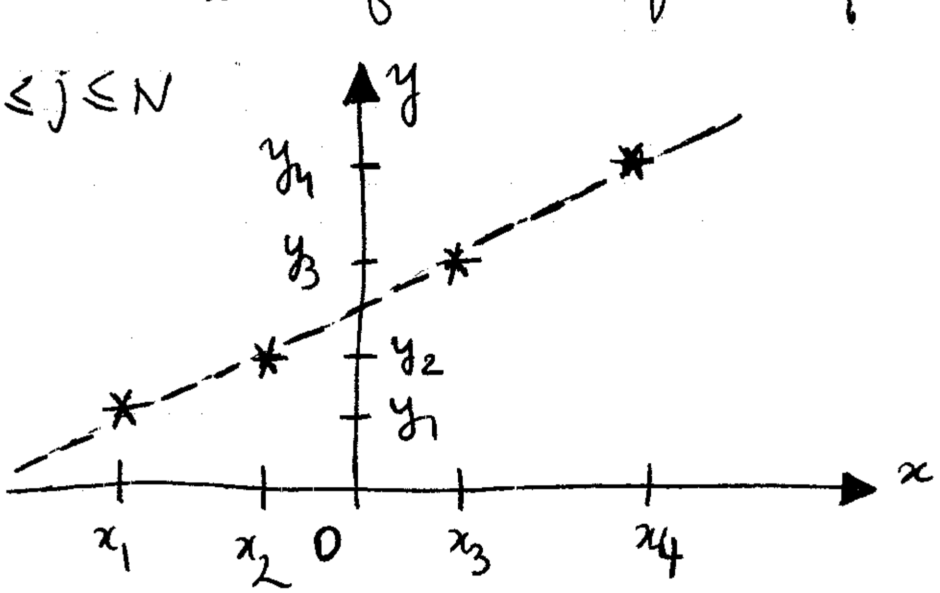

Figure 1

Nous imaginions une configuration comme celle de la figure 1 où les points $M_j(x_j, y_j)_{1\leq j\leq N}$ sont "presque" alignés. Nous aimerions faire passer une droite d'équation

$$(13) \qquad y = \alpha x + \beta \quad , \quad x \in \mathbb{R}$$

"pas trop loin" des points M_j.

• Nous observons d'abord que si les points M_j sont alignés, alors la relation (13) a lieu pour tous les x_j et les y_j ; on a alors $y_j - \alpha x_j - \beta = 0$ pour tout j, relation qu'on peut encore écrire $(y_j - \alpha x_j - \beta)^2 = 0$ pour tout j, soit de manière équivalente $\sum_{j=1}^{N} (y_j - \alpha x_j - \beta)^2 = 0$. En effet, si cette dernière relation a lieu, la somme de carrés (de nombres réels) est nulle et <u>tous</u> les termes de la somme sont nuls. Si les points M_j ne sont <u>pas</u> alignés, la somme précédente n'est pas nulle, mais on va essayer de choisir α et β (la pente et l'ordonnée à l'origine) de sorte que l'écart quadratique $J(\alpha,\beta)$ défini par

$$(14) \qquad J(\alpha,\beta) = \frac{1}{N} \sum_{j=1}^{N} (y_j - \alpha x_j - \beta)^2$$

soit minimal, d'où le nom de la méthode de "moindres carrés" car la fonction $J(\alpha,\beta)$ est une somme de carrés.

Th ① Droite des moindres carrés

On suppose $N \geqslant 2$ et les réels $(x_j)_{1\leq j \leq N}$ choisis de sorte que la relation (12) a lieu.
Il existe un unique couple (a,b) de sorte que

(15) $J(\alpha,\beta) \geqslant J(a,b)$, $\forall (\alpha,\beta) \in \mathbb{R}^2$.

On a de plus

(16) $$a = \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\langle x^2\rangle - \langle x\rangle^2}, \quad b = \langle y\rangle - a\langle x\rangle$$

avec les notations introduites en (8) :

(17) $$\begin{cases} \langle x\rangle = \frac{1}{N}\sum_{j=1}^{N} x_j, & \langle y\rangle = \frac{1}{N}\sum_{j=1}^{N} y_j, \\ \langle x^2\rangle = \frac{1}{N}\sum_{j=1}^{N} x_j^2, & \langle xy\rangle = \frac{1}{N}\sum_{j=1}^{N} x_j y_j. \end{cases}$$

La droite "$y = ax+b$", avec a,b calculés à la relation (16) est celle qui parmi toutes les droites de la forme "$y = \alpha x+\beta$", minimise l'écart quadratique moyen $J(\alpha,\beta) = \frac{1}{N}\sum_{j=1}^{N}(y_j - \alpha x_j - \beta)^2$.

Preuve du théorème ①

- Comme $J(\alpha,\beta)$ est une moyenne de somme de carrés, on a toujours $J(\alpha,\beta) \geqslant 0$. On va essayer de l'écrire sous la forme d'une somme de <u>deux</u> carrés seulement. On remarque d'abord que la proposition 3, jointe au fait que tous les x_j sont différents, entraîne

(18) $\langle x^2\rangle - \langle x\rangle^2 > 0$

et l'expression (16) a bien un sens !

- On a ensuite, en développant le carré de la somme de trois termes :

$$J(\alpha,\beta) = \frac{1}{N} \sum_{j=1}^{N} \left(y_j^2 + \alpha^2 x_j^2 + \beta^2 - 2\alpha x_j y_j - 2\beta y_j + 2\alpha\beta x_j \right)$$

$$= \beta^2 - 2\beta\langle y\rangle + 2\alpha\beta\langle x\rangle - 2\alpha\langle xy\rangle + \alpha^2\langle x^2\rangle + \langle y^2\rangle$$

$$= \left(\beta - \langle y\rangle + \alpha\langle x\rangle\right)^2 - \left(\alpha\langle x\rangle - \langle y\rangle\right)^2 + \alpha^2\langle x^2\rangle - 2\alpha\langle xy\rangle + \langle y^2\rangle$$

$$= \left(\langle y\rangle - \alpha\langle x\rangle - \beta\right)^2 + \left(\langle x^2\rangle - \langle x\rangle^2\right)\alpha^2 - 2\alpha\left(\langle xy\rangle - \langle x\rangle\langle y\rangle\right) + \langle y^2\rangle - \langle y\rangle^2$$

$$(19)\quad \left\{ \begin{aligned} J(\alpha,\beta) = &\left(\langle y\rangle - \alpha\langle x\rangle - \beta\right)^2 + \left(\langle x^2\rangle - \langle x\rangle^2\right)\left(\alpha - \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\langle x^2\rangle - \langle x\rangle^2}\right)^2 \\ &+ \langle y^2\rangle - \langle y\rangle^2 - \frac{\left(\langle xy\rangle - \langle x\rangle\langle y\rangle\right)^2}{\langle x^2\rangle - \langle x\rangle^2}. \end{aligned} \right.$$

- Regardons l'expression (19). On a d'abord le premier terme qui est un carré, donc ≥ 0. Puis le second, qui est égal à un nombre >0 (c'est $\langle x^2\rangle - \langle x\rangle^2$) multiplié par $(\alpha - a)^2$, en posant $a = \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\langle x^2\rangle - \langle x\rangle^2}$ comme à la relation (16) ! Il reste un terme constant. Introduisons $b = \langle y\rangle - a\langle x\rangle$. On a alors

$$(20)\quad J(a,b) = \langle y^2\rangle - \langle y\rangle^2 - \frac{\left(\langle xy\rangle - \langle x\rangle\langle y\rangle\right)^2}{\langle x^2\rangle - \langle x\rangle^2}$$

car cette valeur particulière de (α,β) annule tous les carrés. On a donc, compte tenu des relations

(19) et (20) :

(21) $J(\alpha,\beta) = \left(\beta + \alpha\langle x\rangle - \langle y\rangle\right)^2 + \left(\langle x^2\rangle - \langle x\rangle^2\right)(\alpha - a)^2 + J(a,b)$,

ce qui établir la relation (15) car $(\beta + \alpha\langle x\rangle - \langle y\rangle)^2 \geqslant 0$ et $(\alpha - a)^2 \geqslant 0$, $\langle x^2\rangle - \langle x\rangle^2 > 0$. De plus, si $J(\alpha,\beta) = J(a,b)$, ces deux carrés sont nuls et $\alpha = a$, $\beta = \langle y\rangle - \alpha\langle x\rangle = \langle y\rangle - a\langle x\rangle = b$ compte tenu de (16). Cette remarque achève la démonstration du théorème 1. $\square$

3 Calcul pratique de la droite de régression.

- Nous proposons ici d'écrire l'algorithme très simple qui permet de calculer explicitement les coefficients (a, b) de la droite de régression si les nombres x_j et y_j sont donnés.

- Calcul des valeurs moyennes

```
xm = 0 , ym = 0, x2m = 0, xym = 0
boucle sur j, de 1 à N
xm = xm + x(j)
ym = ym + y(j)
x2m = x2m + x(j)**2
xym = xym + x(j)* y(j)
fin de boucle.
xm = xm /N, ym = ym/N
x2m = x2m /N, xym = xym /N.
```

- Utilisation de la relation (16)

```
a = (xym - xm * ym)/(x2m - xm**2)
b = ym - a * xm.
```

(4) Introduction aux méthodes de l'optimisation

- Nous avons présenté au paragraphe 2 une méthode élémentaire pour calculer les coefficients (a,b) de la droite de regression. En effet, elle ne demande de ne connaître que la factorisation $x^2+2\mu x=(x+\mu)^2-\mu^2$, avec des valeurs de x et μ parfois un peu longues à reporter. Pourtant, cette méthode n'est pas la plus populaire. Nous en donnons une seconde dans ce paragraphe, un peu moins simple conceptuellement, mais qui conduit à des calculs plus simples.

(Th) (2) Condition nécessaire d'optimum

Soit $\mathbb{R}^2 \ni (\alpha,\beta) \longmapsto J(\alpha,\beta)\in\mathbb{R}$ une fonction continûment dérivable du couple (α,β). Soit $(a,b)\in\mathbb{R}^2$ un point de minimum de la fonction J:

(22) $J(\alpha,\beta) \geqslant J(a,b), \quad \forall(\alpha,\beta)\in\mathbb{R}^2$.

on a alors nécessairement

(23) $\dfrac{\partial J}{\partial\alpha}(a,b)=0, \quad \dfrac{\partial J}{\partial\beta}(a,b)=0$.

Preuve du théorème (2)

- Rappelons que pour une fonction de deux variables, $\frac{\partial J}{\partial \alpha}(\alpha, \beta)$ est la dérivée de la fonction "partielle" $\mathbb{R} \ni x \mapsto J(x, \beta) \in \mathbb{R}$ obtenue en fixant le second argument et $\frac{\partial J}{\partial \beta}(\alpha, \beta)$ est la dérivée de la fonction partielle $\mathbb{R} \ni y \mapsto J(\alpha, y) \in \mathbb{R}$ qui fixe le premier argument
- Si (a, b) vérifie (22), on a en particulier $J(a+t, b) \geqslant J(a, b)$, $\forall t \in \mathbb{R}$. Avec $t > 0$, on en déduit $\frac{1}{t}(J(a+t, b) - J(a, b)) \geqslant 0$ et en prenant la limite de cette expression pour t tendant vers 0, $\frac{\partial J}{\partial \alpha}(a, b) \geqslant 0$. Mais on a aussi $J(a-t, b) \geqslant J(a, b)$ pour tout $t \in \mathbb{R}$. Avec $t > 0$ en particulier, $\frac{1}{t}(J(a, b) - J(a-t, b)) \leqslant 0$ et en passant à la limite pour $t \to 0$, $\frac{\partial J}{\partial \alpha}(a, b) \leqslant 0$. Donc $\frac{\partial J}{\partial \alpha}(a, b) = 0$. La seconde inégalité $\frac{\partial J}{\partial \beta}(a, b) = 0$ s'obtient en échangeant le rôle des variables. La preuve est analogue et est laissée au lecteur. $\square$

- Nous utilisons maintenant le théorème 2 avec la fonction J donnée à la relation (14). Si la relation (15) a lieu, alors la relation (23) a lieu. On dérive donc avec soin la relation (14) par rapport à α et β.

$$\frac{\partial J}{\partial \alpha} = \frac{1}{N} \sum_{j=1}^{N} \frac{\partial}{\partial \alpha} \left[\alpha x_j + \beta - y_j \right]^2$$

$$= \frac{1}{N} \sum_{j=1}^{N} 2 (\alpha x_j + \beta - y_j) x_j$$

$$= 2\left(\alpha \langle x^2 \rangle + \beta \langle x \rangle - \langle xy \rangle\right)$$

$$\frac{\partial J}{\partial \beta} = \frac{1}{N} \sum_{j=1}^{N} \frac{\partial}{\partial \beta} \left(\alpha x_j + \beta - y_j\right)^2$$

$$= \frac{2}{N} \sum_{j=1}^{N} (\alpha x_j + \beta - y_j)$$

$$= 2\left(\alpha \langle x \rangle + \beta - \langle y \rangle\right).$$

- Si (a,b) vérifie $\frac{\partial J}{\partial \alpha}(a,b) = \frac{\partial J}{\partial \beta}(a,b) = 0$, on doit donc avoir

$$(24) \quad \begin{pmatrix} \langle x^2 \rangle & \langle x \rangle \\ \langle x \rangle & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \langle xy \rangle \\ \langle y \rangle \end{pmatrix}$$

ce qui pose le problème sous la forme d'un <u>système linéaire</u> de deux équations à deux inconnues. La matrice de ce système est <u>symétrique</u>. Elle est même <u>définie positive</u>. En effet, pour $y_j \equiv 0$ pour tout j, on a :

$$J(\alpha,\beta) = \frac{1}{N} \sum_{j=1}^{N} (\alpha x_j + \beta)^2 = \alpha^2 \langle x^2 \rangle + 2\alpha\beta \langle x \rangle + \beta^2$$

$$= (\alpha \ \ \beta) \begin{pmatrix} \langle x^2 \rangle & \langle x \rangle \\ \langle x \rangle & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \geqslant 0$$

car $J(\alpha,\beta) \geqslant 0$. De plus, le déterminant de cette matrice "de moindres carrés" vaut $\langle x^2\rangle - \langle x\rangle^2$ qui est <u>strictement positif</u> compte tenu de la relation (18).

- La résolution du système linéaire (18(24) n'offre alors pas de difficulté. On multiplie la seconde ligne par $\langle x\rangle$ et on la retranche de la première. Il vient

$$\left(\langle x^2\rangle - \langle x\rangle^2\right)\ a = \langle xy\rangle - \langle x\rangle\langle y\rangle$$

ce qui constitue la première relation de (16). La seconde est identique à la seconde équation.

- Si cette approche nous a permis de <u>calculer</u> sans peine les valeurs de a et b qui sont associées à la droite de régression, elle ne démontre <u>pas</u> la relation de minimum (22), qui fait alors l'objet d'une étude spécifique. Nous renvoyons le lecteur aux ouvrages sur l'optimisation, par exemple celui de P.G. Ciarlet ou celui de J.C. Culioli.

FD, 17 nov 04