Méthode des moindres carrés selon Gauss

On considère un nuage de points $M_i(x_i,y_i)$ que l'on désire ajuster au mieux par une fonction mathématique f donnée : on recherche les paramètres de f (fonction affine, polynôme, exponentielle, etc.) minimisant la somme des carrés des distances entre y_i et $f(x_i)$. On cherche alors à minimiser la somme :

$$\Delta = \sum_{i=1}^{n} [y_i - f(x_i)]^2$$

La recherche d'un ajustement exponentiel (comme l'évolution d'une population) par une fonction de la forme

$$f(x) = ke^{ax}$$

relève de l'ajustement linéaire des points de coordonnées $(x_i , \ln y_i)$. En effet, si la méthode des moindres carrés fournit la droite d'équation $y = ax + b$, on a en fait :

$$\ln y = ax + b, \text{ soit } y = k.e^{ax} \text{ avec } k = e^b$$

En astronomie, outre l'ajustement décrit ci-après, Gauss appliqua la méthode des moindres carrés à des mesures d'observation conduisant à la résolution de systèmes d'équations linéaires rectangulaires possédant plus d'équations que d'inconnues : il s'agit de rechercher une solution moyenne (ensemble de mesures) la plus vraisemblable en minimisant la somme Σr_i^2 où r_i désigne le résidu, différence entre le 1er membre de la i-ème équation et son terme constant.

Etudions le cas d'une approximation polynomiale de degré p (si p = 1, on est en présence d'un ajustement linéaire). Posons :

$$f(x) = c_o + c_1 x + \ldots + c_p x^p$$

où p est le degré du polynôme d'ajustement et c_k , k variant entre o et p, les coefficients recherchés. Il s'agit ici de minimiser :

$$\Delta(c_o, \ldots, c_p) = [y_i - f(x_i)]^2 = \sum_{i=1}^{n} [y_i - \sum_{k=o}^{p} c_k x_i^k]^2$$

Scanned by CamScanner

Nous admettons ici deux résultats : concernant les fonctions numériques à plusieurs variables, une condition nécessaire (non suffisante) d'extremum en un point $(c_o, \ldots, c_p)$ est que toutes les dérivées partielles de Δ, à savoir $\frac{\partial \Delta}{\partial c_o}, \frac{\partial \Delta}{\partial c_1}, \ldots, \frac{\partial \Delta}{\partial c_p}$ soient nulles en ce point. Dans le cas présent de la méthode des moindres carrés, on obtient effectivement un minimum (on peut le prouver en développant Δ en série de Taylor).

On écrit alors, pour tout k :

$$\frac{\partial \Delta}{\partial c_k}(c_o, \ldots, c_p) = 2\sum_{i=1}^{n}\left([y_i - f(x_i)] \times [-x_i^k]\right) = 0$$

Ce qui conduit au système de p+1 équations : $\sum_{i=1}^{n} x_i^k f(x_i) = \sum_{i=1}^{n} y_i x_i^k$

Notons, pour simplifier : $S_k = \sum_{i=1}^{n} x_i^k$ et $W_k = \sum_{i=1}^{n} y_i x_i^k$.

La première équation (k = 0) s'écrit : $\sum_{i=1}^{n} f(x_i) = \sum_{i=1}^{n} y_i$

c'est à dire : $nc_o + c_1S_1 + c_2S_2 + \ldots + c_pS_p = W_o$

La seconde équation (k = 1) s'écrit :

$$\sum_{i=1}^{n} x_i f(x_i) = \sum_{i=1}^{n} y_i\, x_i$$

c'est à dire :

$$c_oS_1 + c_1S_2 + c_2S_3 + \ldots + c_pS_{p+1} = W_1$$

La k-ème s'écrira :

$$c_oS_{k-1} + c_1S_k + c_2S_{k+1} + \ldots + c_pS_{p+k-1} = W_k$$

et la (p+1)-ème :

$$c_oS_p + c_1S_{p+1} + c_2S_{p+2} + \ldots + c_pS_{2p} = W_p.$$

Le système s'écrit alors matriciellement :

Scanned by CamScanner

$$\begin{pmatrix} n & S_1 & S_2 & \dots & S_p \\ S_1 & S_2 & S_3 & \dots & S_{p+1} \\ S_2 & S_3 & S_4 & \dots & S_{p+2} \\ \dots \\ S_p & S_{p+1} & S_{p+2} & \dots & S_{2p} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \dots \\ c_p \end{pmatrix} = \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ \dots \\ W_p \end{pmatrix}$$

On constate que la matrice du système est symétrique. Si nous notons $a_{i,j}$ le terme générique, on a

- $a_{1,1} = n$;
- $a_{i,j} = S_{i+j-2}$;
- le second membre étant $a_{i,p+2} = W_i$.

On applique à ce système la *méthode du pivot* de Gauss et l'on obtient le programme suivant :

Scanned by CamScanner

Stat : Moindres Carrés

Méthodes des moindres carrés discrets

L'interpolation polynomiale est un outil de peu de valeur pour le traitement des données expérimentales. Par contre il est relativement fréquent d'avoir à ajuster sur des données une loi empirique ou encore d'avoir à déterminer la valeur des paramètres d'une loi théorique de façon à reproduire les résultats d'une expérience. Une façon possible pour aborder de tels problèmes est de chercher les valeurs de paramètres minimisant un critère.

I. Ajustement affine

Comme son chiffre d'affaires d'un mois est connu avec un retard, une entreprise choisit de l'estimer par l'intermédiaire d'une autre grandeur dont la valeur peut-être connue rapidement. Elle choisira celle qui aura, avec le chiffre d'affaires, le plus fort coefficient de corrélation linéaire.

On note :
X : le chiffre d'affaires, mesuré en millions de francs ;
Y la consommation de matière premières, mesurée en tonnes ;
Z les salaires en primes, mesurée en millions de francs.

Les mesures de ces variables ont donné, sur six mois de l'année précédente, les résultats suivants :

NUMERO DU MOIS	1	2	3	4	5	6
X_i	36	40	32	32	41	35
Y_i	0.9	1.2	0.6	0.5	1.4	1
Z_i	3.9	3.7	3.2	3.1	3.6	3.7

1. Représenter graphiquement ces données :
   a. premier graphique : X en abscisses, Y en ordonnées.
   b. second graphique : X en abscisses, Z en ordonnées.
2. Ecrire un programme covar, d'argument deux tableaux X_val, Y_val, qui permet de calculer la covariance de ces deux tableaux. Calculer les coefficients de corrélation linéaire r_1 entre X et Y et r_2 entre X et Z.
3. Déterminer une équation de la droite de régression de Y en X puis de celle de Z en X. Tracer dans un même graphe les couple (X_i,Y_i) (resp. (X_i,Z_i)) et leur droite de régression.
4. Trouver la courbe qui approche au mieux le nuage des points (X_i,Z_i).

Scanned by CamScanner

II. Méthode des moindres carrés

1/ Des points presque alignés (régression linéaire)

Supposons connus **N** points ($(x_i$, $y_i)$, pour i=1…N) et considérons une fonction f(A,B,x) dépendant de **2** paramètres et de la variable x, que l'on cherche à ajuster au mieux sur les données. De façon plus précise, on cherche à déterminer la valeur des paramètres A et B *minimisant* le critère des moindres carrés : $J(A,B)=\sum_{i=1}^{N}\left|f(A,B,x_i)-y_i\right|^2$

Remarquons que ce critère est arbitraire et qu'on peut en concevoir de nombreux autres tout aussi valables. On remarque que J(A,B) ne tient pas compte du signe des écarts et qu'il est *dérivable* par rapport aux paramètres A et B si la fonction f(A,B,x) est dérivable par rapport à ces paramètres.

Une condition *nécessaire* (mais en général non suffisante) pour que le critère soit minimum est que les dérivées de J(A,B) par rapport à A et à B s'annulent.

1. En posant f(A,B,x)=A*x+B, définir les fonctions f(A,B,x) et J(A,B). Calculer les dérivées $\frac{\partial J(A,B)}{\partial A}$ et $\frac{\partial J(A,B)}{\partial B}$.
2. Résoudre le système suivant $\frac{\partial J(A,B)}{\partial A}=0$, $\frac{\partial J(A,B)}{\partial B}=0$. Déduire qu'on peut écrire : $A=\frac{s_0.t_1-s_1.t_0}{s_0.s_2-s_1^2}$ et $B=\frac{s_2.t_0-s_1.t_1}{s_0.s_2-s_1^2}$, tels que : $s_0=N$, $s_1=\sum_{i=1}^{N}x_i$, $s_2=\sum_{i=1}^{N}x_i^2$, $t_0=\sum_{i=1}^{N}y_i$, $t_1=\sum_{i=1}^{N}x_i y_i$.
3. Représenter dans un même graphe le nuage des points (x_i, y_i) et la fonction f(A,B,x)=Ax+b.
4. Ecrire une procédure Moin_care d'argument les tableaux Xvals et Yvals qui réalise la régression linéaire.

2/ Régression polynomiale

On suppose maintenant que la fonction f est un polynôme de degré trois et qu'elle s'écrit : $f(a_0, a_1, a_2, a_3,x)=a_0+a_1.x+a_2.x^2+a_3.x^3$ et $J(a_0, a_1, a_2, a_3)= \sum_{i=1}^{N}\left(f(a_0,a_1,a_2,a_3,x_i)-y_i\right)^2$.

1. Montrer qu'on a : $\frac{\partial J(a_0,a_1,a_2,a_3)}{\partial a_j}=2\sum_{i}^{N}x_i^j(a_0+..+a_3x_i^3-y_i)$ pour j=0…3
2. Montrer qu'en posant $\frac{\partial J(a_0,a_1,a_2,a_3)}{\partial a_j}=0$ pour j=0..3, on obtient un système de 4 équations à 4 inconnues que l'on peut écrire sous la forme S.A=T, avec :

$$S=\begin{pmatrix} N & \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 \\ \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 \end{pmatrix},\ A=\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} \text{ et } T=\begin{pmatrix} \sum y_i \\ \sum x_i.y_i \\ \sum x_i^2.y_i \\ \sum x_i^3.y_i \end{pmatrix}.$$

Scanned by CamScanner

3. En utilisant une boucle for construire les éléments S_i ainsi que les t_i pour i=0,1,2,3.. Construire la matrice S et résoudre le système S.A=T. Définir les éléments a_i, pour i=1..3, à partir du vecteur A.
4. Application : on pose x=[3.8, 4.11, 5.18, 4.25, 5.49, 1.11, 5.35, 2.09, 0.6, 6.53, 1.40] et y=[56.1, 56.2, 40.0, 45.9, 38.5, 32.4, 31.0, 50.0, 18.2, 30.1, 45.8]. Calculez les coefficients a_i , puis représentez dans un même graphe le nuage des points (x_i, y_i) et la fonction $f(a_0, a_1, a_2, a_3,x)=a_0+a_1.x+a_2.x^2+a_3.x^3$.
5. Ecrire une procédure Moin_poly d'argument les tableaux Xvals et Yvals qui réalise la régression polynomiale.

Notions utiles

With, plots, plot, point plot, display, color, proc, stats, fit, zip, leastsquare, rhs, sum, nops, RETURN, evalf, simplify, solve, value, for do od, min, max, linalg, matrix, vecteur, linsolve,

Scanned by CamScanner