MÉTHODE DE MOINDRE NORME (H3)
(07 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Une **méthode de moindre norme** est une méthode d'**estimation** utilisée lorsque la fonction objectif considérée est de nature vectorielle. Elle se base donc sur une **norme** et procède par **optimisation**.

Un exemple en est la méthode de **régression** basée sur une **norme** de $\mathbf{R}^N$ (cf aussi **valeur typique de FRÉCHET**, **estimateur à distance minimale**, **estimateur à distance minimum**).

(i) Ainsi, on considère un **modèle de régression linéaire** multiple, écrit $\eta = \xi' b + \varepsilon$ dans un **espace de variables**, et « observé » dans l'**espace d'observation** $\mathbf{R}^N$ selon :

(1) $y = X b + u$, avec $E u = 0$, $V u = \sigma^2 . I_N$.

(a) le **statisticien** est généralement conduit à estimer b en minimisant la **norme** $||.||_2$ définie par la somme des carrés :

(2) $||u||_2^2 = \Sigma_{n=1}^N u_n^2$ (ou $||u||_2^2 = u' u$),

et ceci fonde la **méthode des moindres carrés ordinaires**, associée à l'**hypothèse** d'**homoscédasticité** $V u = \sigma^2 . I_N$;

(b) dans certaines circonstances (eg V u inconnue, variables à valeurs dans L^p, étude de **robustesse**), il peut aussi estimer b en minimisant la norme $||.||_p$ définie par (norme dans l^p) :

(3) $||u||_p^p = \Sigma_{n=1}^N |u_n|^p = \Sigma_{n=1}^N |y_n - X_n b|^p$, avec $p \in [1, +\infty[$,

où (X_n , y_n) est la n-ième observation de (ξ, η), ie la n-ième ligne de la (N,K+1)-**matrice** « augmentée » [X, y]. L'entier p doit être tq la norme soit convexe (cf **convexité**, **fonction convexe**), ce qui est vérifié lorsque $p \in 2\, \mathbf{N}^*$.

Ce procédé conduit à résoudre le problème de **programmation mathématique** (sans contrainte sur b) suivant :

(4) $\min_{b \in \mathbf{R}K} ||u||_p^p$ ou $\min_{b \in \mathbf{R}K} ||u||_p$ (où **R**K désigne $\mathbf{R}^K$).

La solution $b_p^\#$ ainsi obtenue est appelée **estimateur de moindre norme** de b.

(ii) Les propriétés statistiques de $b_p^\#$ sont peu connues lorsque le modèle est à distance finie (ie lorsque $N << +\infty$), notamment du fait de la **complexité** de la résolution de (4) dans le cas général ($p \neq 2$). En effet, la solution ne s'explicite pas, en général, en fonction de (X, y).

Des **simulations** peuvent être effectuées dans chaque cas d'espèce.

Ainsi, lorsque p est pair ($p \in 2\,\mathbf{N}^*$), on doit calculer la **dérivée** première du polynôme (de degré p pr à b) suivant :

(5) $\quad P_p(b) = \Sigma_{n=1}^{N} (y_n - \Sigma_{k=1}^{K} x_{nk} b_k)^p$,

où $X = (x_{nk})_{(n,k)}$ et $y = (y_n)_n$, ce qui conduit à un système comportant K équations polynômiales de degré p - 1 en $b \in \mathbf{R}^K$:

(6) $\quad P_{k,p-1}(b) = 0, \qquad \forall\, k \in N_K^*$,

avec $P_{k,p-1}(b) = -\, p\, \Sigma_{n=1}^{N} (y_n - \Sigma_{l=1}^{K} x_{nl} b_l)^{p-1} x_{nk}$.

L'ensemble des points $b \in \mathbf{R}^K$ annulant une famille finie de fonctions polynômiales tq les précédentes est un ensemble algébrique, qui se réduit ici à un point $b_p^{\#}$ puisque la fonction $b \mapsto ||u||_p^p$ définie par (4) est strictement convexe.

(iii) L'application $s : \mathbf{N}^* \mapsto \mathbf{R}^K$ définit une **suite** d'**estimateurs** à valeurs dans $\mathbf{R}^K$ selon l'application (ou **statistique**) :

(7) $\quad p \in \mathbf{N}^* \mapsto b_p^{\#} = s(X, y) \in \mathbf{R}^K$.

A titre d'exemple :

(a) lorsque p = 2, on a $b_2^{\#} = b^{\wedge}$ (**estimateur des mco**) ;

(b) lorsque p = 1, $b_1^{\#}$ est l'estimateur du minimum de la somme des valeurs absolues (des perturbations u_n), ou estimateur de la **régression médiane**, utilisée dans la théorie de la **régression robuste** (cf **régression quantilaire**) ;

(c) lorsque $p \to +\infty$, on définit l'**estimateur du « minimax »**, ou **estimateur « minimax » de la régression**, $b_\infty^{\#}$ de b. En effet, la définition (3) et la méthode (4) s'étendent à ce cas en posant :

$$(8) \qquad ||u||_\infty = \begin{cases} \lim_{p \to +\infty} \{\Sigma_{n=1}^{N} |y_n - \Sigma_{k=1}^{K} x_{nk} b_k|^p\}^{1/p} \\ \max_n |y_n - \Sigma_{k=1}^{K} x_{nk} b_k| = \max_n |u_n|, \end{cases}$$

et en résolvant le programme mathématique $\min_b \max_n |u_n|$.

Chaque estimateur $b_p^{\#}$ (où $p \in [1, +\infty[$) est un **estimateur du mv** particulier. En effet, si la **perturbation** ε conduit à une **suite iid** selon une **lp** à densité exponentielle de la forme (cf aussi **loi exponentielle**) :

(9) $\quad t \in \mathbf{R} \mapsto f(t) = dP^{\varepsilon} / d\lambda_1(t) = c_p \,.\, \exp(-|t|^p)$,

où c_p est une **constante** de **normalisation**, alors $b_p^{\#}$ est l'**estimateur du maximum de vraisemblance** associé à f (même lorsque $p \to +\infty$ pour l'estimateur minimax).