MÉTHODE DES MOINDRES CARRÉS GÉNÉRALISÉS (H3, J1)
(02 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **méthode des moindres carrés généralisés** est une méthode générale d'estimation des paramètres d'un **modèle de régression**, linéaire ou non, à **perturbation** additive. Elle consiste à minimiser une **distance** (ou une **norme**) euclidienne entre la **variable endogène** $y \in \mathbf{R}^N$ et le terme « certain » z (**espérance**) de la **régression** : z est astreint à parcourir une **partie** spécifiée de l'**espace d'observation**.

Cette méthode généralise la **méthode des moindres carrés ordinaires**. Elle s'analyse aussi comme une application du théorème de la projection dans un **espace de HILBERT** (cf **théorème de la projection orthogonale**).

(i) Ainsi, $\mathbf{R}^N$ étant l'**espace d'observation** ambiant, la **méthode des moindres carrés généralisés** (**mcg**) (de A.C. AITKEN) appliquée au modèle de régression :

(1) $y = z + u, \quad \text{avec } E\, u = 0,$

consiste, en supposant que $\Sigma \in S_N(\mathbf{R})$ est une **matrice définie positive** ($\Sigma \geq 0$), à minimiser la distance suivante :

(2) $\varphi_\Sigma(z) = ||y - z||_\Sigma^2 = (y - z)' \, \Sigma^{-1} (y - z)$

pr à $z \in \mathcal{V}$, où $\mathcal{V} \subset \mathbf{R}^N$ est une sous-variété donnée (ie spécifiée par le **statisticien**).

On note $b_g^\wedge$ ou $b^{\wedge\wedge}$, ou parfois β_g , la **solution des moindres carrés généralisés** ainsi définie.

Si la régression est une **régression linéaire**, $z = X\, b$ représente l'équation paramétrée par b de la variété linéaire, de dimension K, $\mathcal{V} \triangleleft \mathbf{R}^N$ (sous-**espace vectoriel**). La méthode revient à minimiser pr à $b \in \mathbf{R}^K$ la **forme quadratique** (non homogène) :

(3) $q_\Sigma(b) = ||y - X\, b||_\Sigma^2 = (y - X\, b)' \, \Sigma^{-1} (y - X\, b).$

Le fondement de la méthode est ainsi associé à l'hypothèse (stochastique) du second ordre selon laquelle Σ n'est autre que la **matrice de dispersion** de la **perturbation aléatoire** u :

(4) $V\, u = V\, y = \Sigma > 0.$

La méthode équivaut donc à minimiser la forme quadratique $q_\Sigma(b) = (y - X\, b)' \, (V\, u)^{-1} (y - X\, b)$ dans la **métrique** définie par la matrice Σ.

La solution $b_g^\wedge$ ou $b^{\wedge\wedge}$ s'appelle **estimateur des moindres carrés généralisés** (mcg), ou **estimateur de A.C. AITKEN - C.F. GAUSS - A.A. MARKOV**, du paramètre b.

Il existe toujours un nombre réel $\sigma > 0$ tq $\Sigma = \sigma^2 . \Omega$, où Ω est une matrice définie positive ($\Omega > 0$). La méthode des moindres carrés ordinaires (mco) correspond au cas où $\Omega = I_N$. On suppose alors que (cf **trace**) :

(5) $\quad \mathrm{tr}\, \Omega = \mathrm{rg}\, \Sigma = N.$

(ii) Dans le cas d'un modèle de régression linéaire, on montre que :

(a) si Ω est connue et régulière et si rg X = K, la solution de (2), appelée **estimateur des moindres carrés généralisés** (mcg) de b, est donnée par :

(6) $\quad b_g^{\wedge} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y.$

Cette solution vérifie les propriétés suivantes :

$(a)_1$ $E\, b_g^{\wedge} = b$ (**estimateur sans biais**) (cf aussi **biais**) ;

$(a)_2$ $V\, b_g^{\wedge} = (X' \Sigma^{-1} X)^{-1} = \sigma^2 (X' \Omega^{-1} X)^{-1}$ (**matrice de dispersion**) ;

$(a)_3$ ($b_g^{\wedge}$ vérifie) le **théorème de AITKEN-GAUSS-MARKOV** ;

$(a)_4$ si l'on pose $y_g^{\wedge} = X\, b_g^{\wedge} = z_g^{\wedge}$ et $u_g^{\wedge} = y - y_g^{\wedge}$, la **statistique** :

(7) $\quad (\sigma^2)_g^{\wedge} = (u_g^{\wedge}{}' \, \Omega^{-1} u_g^{\wedge}) / (N - K)$

constitue un **estimateur sans biais** de σ^2 (ie $E\, (\sigma^2)_g^{\wedge} = \sigma^2$) ;

$(a)_5$ l'estimateur des mcg estimé sur le modèle :

(8) $\quad y = X\, b + u, \qquad$ avec $E\, u = 0$ et $V\, u = \Sigma = \sigma^2 . \Omega,$

est identique à l'estimateur des mco du **modèle « sphéricisé »** suivant :

(9) $\quad y° = X°\, b + u°, \qquad$ avec $E\, u° = 0$ et $V\, u° = \sigma^2 I_N$,

où $y° = \Omega^{-1/2} y$, $X° = \Omega^{-1/2} X$ et $u° = \Omega^{-1/2} u$;

$(a)_6$ si $u \sim \mathcal{N}_N (0, \sigma^2 \Omega)$ (**loi normale multidimensionnelle** centrée), l'**estimateur du maximum de vraisemblance** de (b, σ^2) n'est autre que $(b_g^{\wedge}, (\sigma^2)_{gc}^{\wedge})$, avec $(\sigma^2)_{gc}^{\wedge} = \{N / (N - K)\} (\sigma^2)_g^{\wedge}$ (**variance « corrigée »**) ;

$(a)_7$ on a :

(10) $\quad e_N' u_g^{\wedge} = 0$ (P-p.s.) $\Leftrightarrow \Omega\, e_N \in \mathrm{Im}\, X.$

(b) si Ω est inconnue et si l'on estime b à l'aide de l'**estimateur des mco** $b^{\wedge}$ (au lieu de l'**estimateur des mcg** $b_g^{\wedge}$), alors :

(11) $\quad b^{\wedge} = b_g^{\wedge} \Leftrightarrow \mathrm{Im}\, (\Omega X) \subset \mathrm{Im}\, X.$

Par suite, l'**efficacité relative** (scalaire) de $b^\wedge$ pr à $b_g^\wedge$ vaut :

(12) $e_N (b^\wedge / b_g^\wedge) = \text{Dét} (V\, b_g^\wedge) / \text{Dét} (V\, b^\wedge) = |X'\, X|^2 / (|X'\, \Omega^{-1}\, X|.|X'\, \Omega\, X|)$;

(iii) Si Ω est inconnue (cas général), plusieurs procédés d'estimation de b sont possibles :

a) la **méthode des mco** (cf infra). Celle-ci semble d'autant plus « efficace » que la « distance » entre Ω et I_N (ou celle de Σ et $\sigma^2\, I_N$) est moindre ;

b) la **méthode du maximum de vraisemblance**. Celle-ci permet d'estimer le couple (b, Σ) dans lequel Σ joue le rôle de **paramètre importun**. Cette matrice dépend seulement de N (N+1) / 2 paramètres scalaires distincts au plus ;

c) la **méthode des moindres carrés quasi-généralisés** ;

d) une **méthode à distance minimale** (cf aussi **estimateur à distance minimum**, **méthode de moindre norme**).

(iv) Des considérations analogues aux précédentes peuvent être développées pour un modèle de **régression non linéaire** (pr à b). On minimise alors (pr à b) l'expression suivante :

(13) $q_\Sigma (b) = ||y - F (b)||_\Sigma^2 = (y - F (b))'\, \Sigma^{-1}\, (y - F (b))$.