

Méthode des moindres carrés généralisés

Jean Debord

30 avril 2003

1 Théorie

Soit à ajuster une fonction à une série de m points expérimentaux. On suppose que, pour chaque observation k , la valeur observée de la variable dépendante (y_k) est distribuée selon une loi normale de moyenne \hat{y}_k (\hat{y}_k étant la valeur calculée de la variable dépendante) et de variance v_k .

La valeur calculée \hat{y}_k est une fonction de la variable indépendante x_k et du vecteur des paramètres de la régression, $\mathbf{B} = [b_0, b_1, \dots]$:

$$\hat{y}_k = f(x_k, \mathbf{B}) \quad (1)$$

v_k est une fonction de \hat{y}_k et du vecteur des paramètres de la variance, $\Theta = [\theta_0, \theta_1, \dots]$, telle que :

$$v_k = \theta_0 \cdot g(\hat{y}_k, \theta_1, \dots) \quad (2)$$

où g est une fonction spécifiée par l'utilisateur, par exemple une fonction puissance :

$$g(\hat{y}_k, \theta_1, \dots) = (\hat{y}_k)^{\theta_1} \quad (3)$$

Si l'on dispose de N courbes, correspondant chacune à un vecteur de paramètres \mathbf{B}_i ($i = 1 \dots N$), et si l'on considère que la fonction de variance est la même pour toutes les courbes, le modèle général est alors le modèle de régression non linéaire hétéroscédastique :

$$y_{ik} = f(x_{ik}, \mathbf{B}_i) + \theta_0 \cdot g[f(x_{ik}, \mathbf{B}_i), \theta_1, \dots] \cdot \epsilon_{ik} \quad (4)$$

Les ϵ_{ik} sont des variables aléatoires supposées indépendantes et identiquement distribuées selon une loi normale de moyenne nulle et de variance unité. Dans ces conditions, la méthode qui s'impose de façon théorique est

la méthode du maximum de vraisemblance, qui consiste à minimiser simultanément en tous les \mathbf{B}_i et Θ la pseudo-vraisemblance (*pseudo-likelihood*, PL) :

$$PL = \sum_{i=1}^N PL_i = \sum_{i=1}^N \sum_{k=1}^{m_i} \left[\frac{(y_{ik} - \hat{y}_{ik})^2}{v_{ik}} + \ln v_{ik} \right] \quad (5)$$

soit :

$$PL_i = \sum_{k=1}^{m_i} \left[\frac{[y_{ik} - f(x_{ik}, \mathbf{B}_i)]^2}{\theta_0 \cdot g[f(x_{ik}, \mathbf{B}_i), \theta_1, \dots]} + \ln \theta_0 \cdot g[f(x_{ik}, \mathbf{B}_i), \theta_1, \dots] \right] \quad (6)$$

et :

$$PL = \sum_{i=1}^N \sum_{k=1}^{m_i} \left[\frac{[y_{ik} - f(x_{ik}, \mathbf{B}_i)]^2}{\theta_0 \cdot g[f(x_{ik}, \mathbf{B}_i), \theta_1, \dots]} + \ln \theta_0 \cdot g[f(x_{ik}, \mathbf{B}_i), \theta_1, \dots] \right] \quad (7)$$

où m_i désigne le nombre d'observations pour la courbe i

Cependant, la mise en oeuvre pratique de ce calcul pour une pseudo-vraisemblance un peu compliquée peut présenter des difficultés. Il vient alors naturellement à l'esprit de choisir la méthode des moindres carrés pondérés. Malheureusement il est possible parfois que l'estimateur ainsi construit ne soit pas consistant, c'est-à-dire n'approche jamais la vraie valeur du paramètre. Pour nous tirer de ce mauvais pas, nous allons faire une adaptation simple de ce procédé en utilisant une méthode à deux étapes, c'est-à-dire que les paramètres de la régression sont déterminés pour chaque courbe et les paramètres de la variance pour toutes les courbes en même temps [réf. 1,2].

Plus précisément, le processus itératif est tel qu'à l'itération courante :

1. Les paramètres de la variance sont estimés en minimisant la pseudo-vraisemblance globale PL par rapport à Θ , avec les valeurs courantes des paramètres de régression \mathbf{B}_i
2. Les paramètres de la régression sont déterminés, pour chaque courbe, en minimisant la pseudo-vraisemblance individuelle PL_i par rapport à \mathbf{B}_i , en utilisant les paramètres de variance obtenus à l'étape précédente.

Ces deux étapes sont répétées jusqu'à convergence des paramètres de la variance.

Les estimations initiales des paramètres de régression \mathbf{B}_i sont obtenues par régression non linéaire pondérée, les variances étant approchées par :

$$v_{ik} = g(y_{ik}, \theta_1, \dots) \quad (8)$$

où $\theta_1, ..$ ont des valeurs fixées par l'utilisateur (p. ex. 1). Ceci revient à calculer, pour chaque courbe i , le vecteur \mathbf{B}_i qui minimise la somme des carrés des écarts pondérée :

$$S_i = \sum_{k=1}^{m_i} \frac{[y_{ik} - f(x_{ik}, \mathbf{B}_i)]^2}{g[f(x_{ik}, \mathbf{B}_i), \theta_1, \dots]} \quad (9)$$

Le paramètre linéaire θ_0 peut alors être estimé à partir des résultats de cette régression pondérée, en minimisant la pseudo-vraisemblance totale PL par rapport à θ_0 seul, ce qui donne :

$$\theta_0 = \frac{1}{M_t} \sum_{i=1}^N S_i \quad (10)$$

où M_t désigne le nombre total d'observations, $M_t = \sum_{i=1}^N m_i$

2 Programmation

Les calculs précédents ont fait l'objet d'une programmation en Pascal [réf. 3] selon l'algorithme suivant :

1. Initialiser $\theta_j (j > 0)$
2. Pour chaque courbe i : Trouver \mathbf{B}_i qui minimise (9)
3. Estimer θ_0 par (10)
4. $Iter \leftarrow 0$
5. Répéter :
 - (a) $Iter \leftarrow Iter + 1$
 - (b) $\Theta^0 \leftarrow \Theta$
 - (c) Trouver Θ qui minimise (7)
 - (d) Pour chaque courbe i : Trouver \mathbf{B}_i qui minimise (6)

jusqu'à $(|\theta_j - \theta_j^0| < TOL \cdot |\theta_j^0| \quad \forall j)$ ou $(Iter > MAXITER)$

où $Iter$ désigne le numéro de l'itération, $MAXITER$ le nombre maximal d'itérations et TOL la précision requise sur les paramètres de la variance.

Le coeur du programme est constitué par la procédure de minimisation. En principe, n'importe lequel des nombreux algorithmes existants peut convenir. Nous avons pour notre part utilisé l'algorithme BFGS avec détermination numérique du gradient.

3 Références

1. M. Davidian, D. M. Giltinan, *Biometrics* (1993) **49**, 59-73
2. J. Debord, T. Dantoine, K. Suchaud, M. Harel, B. Verneuil, L. Merle. *Analusis* (1997) **25**, 293-297
3. J. Debord, K. Suchaud. Programme GLS.PAS (http://www.unilim.fr/pages_perso/jean.debord/tpmath/gls.zip)