Méthodes d'estimation

dans les

modèles linéaires généralisés mixtes

(GLMM)

Présentation du modèle

Modèle linéaire
LM

généralisé
GLM

mixte
LMM

GLMM

Présentation du modèle

$$\begin{tikzcd}
 & \begin{array}{l}\text{généralisé}\\ \text{GLM}\end{array} \arrow[dr] & \\
\begin{array}{l}\text{Modèle linéaire}\\ \text{LM}\end{array} \arrow[ur] \arrow[dr] & & \text{GLMM} \\
 & \begin{array}{l}\text{mixte}\\ \text{LMM}\end{array} \arrow[ur] &
\end{tikzcd}$$

Fléxibilité :

- données catégorielles, discrètes ou continues “non gaussiennes”

  ↪ lois classiques dans la famille exponentielle

- introduction d’effets aléatoires

  ↪ modélisation : dépendance entre données

  données répétées - corrélation

  ↪ surdispersion

Description formelle du modèle et notations

▷ Deux sources d'aléa :

- $u_j \qquad j = 1, .., q$ (q clusters) $\quad j^{\text{ème}}$ réalisation de l'effet aléatoire
- $y_{ij} \qquad i = 1, .., n_j \quad i^{\text{ème}}$ observation dans le $j^{\text{ème}}$ cluster

▷ Modèle hiérarchique à 2 niveaux :

- distribution de Y conditionnellement aux effets aléatoires :

$$Y|U \sim FamExp(\mu, \phi)$$

avec $\forall i, j$:

$$\begin{aligned} E(Y_{ij}|U) &= \mu_{ij} = g^{-1}(\eta_{ij}) = g^{-1}(x'_{ij}\beta + u_j) \\ \text{var}(Y_{ij}|U) &= \phi v(\mu_{ij}) \end{aligned}$$

- distribution des effets aléatoires U :

$$U \sim \mathcal{N}_q(0, \sigma^2 A)$$

Estimation dans les GLMM

Approche par maximisation de vraisemblance

$$\eta = X\beta + Zu$$

- paramètres à estimer : β, σ^2
- données observées : y

$$\begin{aligned} l(\beta, \sigma^2; y) &= \int_{\mathbb{R}^q} f_{Y|U}(y)\ f_U(u)\ du \\ &= \int_{\mathbb{R}^q} \prod_{i,j} f_{Y_{ij}|U_j}(y_{ij})\ f_U(u)\ du \end{aligned}$$

Problème : intégrale non calculable formellement

exceptions : loi normale, modèle béta-binomial, modèle Poisson-gamma.

$\hookrightarrow$ forme analytique de la distribution marginale de Y ?

$\hookrightarrow$ approximation nécessaire

PLAN

1 - Approximation numérique de l'intégrale

$$
\begin{aligned}
l(\beta,\sigma^2;y) &= \int_{\mathbb{R}^q} \prod_{i,j} f_{Y_{ij}|U_j}(y_{ij}) \ \prod_j f_{U_j}(u_j)\, du_j \\
&= \prod_j \int_{\mathbb{R}} \prod_i f_{Y_{ij}|U_j}(y_{ij}) \ f_{U_j}(u_j)\, du_j \\
&= \prod_j \int_{\mathbb{R}} \prod_i f_{Y_{ij}|U_j^*}(y_{ij}) \ f_{U_j^*}(u_j^*)\, du_j^*
\end{aligned}
$$

▷ Quadrature gaussienne :

intégrale simple approchée par une somme discrète pondérée :

- choix des points de quadrature : $u_{j,l}^{*c}$

- choix des poids : $w_{j,l}$

cf. Abramowitz & Stegun (1964)

$$
L(\beta,\sigma^2;y) \simeq \sum_j \ln \left[\sum_l \left(\prod_i f_{Y_{ij}|U_j^*}(y_{ij}, u_{j,l}^{*c}) \right) \ w_{j,l} \right]
$$

↪ les paramètres sont alors des paramètres fixes à estimer

▷ Inconvénients (limites de cette approche) :

- en dimension q : numériquement lourd
- plusieurs effets aléatoires   - si emboîtés : ok
  - si croisés : ?
- choix des points de quadratures

▷ Logiciel :

SAS - macro nlmixed

quadrature gaussienne adaptative basée sur une méthode pour les modèles non linéaires mixtes.

▷ Méthodes d'intégration stochastique

• Échantillonage préférentiel

cf. Geyer & Thompson (1992)

- principe : utilisation d'une fonction d'importance $h_U(u)$ pour simuler :

$$l(\beta, \sigma^2; y) = \int_{\mathbb{R}^q} \frac{f_{Y|U}(y)\; f_U(u)}{h_U(u)}\; h_U(u)\; du$$

$$\text{d'où} \quad l(\beta, \sigma^2; y) \simeq \frac{1}{N} \sum_{k=1}^{N} \frac{f_{Y|U}(y)\; f_U(u^{(k)})}{h_U(u^{(k)})}$$

- inconvénients : - choix fonction d'importance

  - estimations peuvent être très erronées

  - efficacité ?

- Algorithme MC-EM

cf. Mc Culloch (1994, 1997)

- principe : étape E : approximation stochastique de l'intégrale par algorithme de Metropolis
- inconvénients :
  - calculs informatiques intensifs
  - étape de maximisation : pas si facile ?
  - problème de EM au bord de l'espace des paramètres

- Paradigme bayésien

cf. Zeger & Karim (1991), Tanner (1993)

- principe : a priori sur β et σ^2 et on travaille avec les lois a posteriori
- inconvénients :
  - a posteriori peut ne pas exister
  - échantillonnage de Gibbs peut ne pas le détecter et produire des estimations fausses.

2 - Approche conditionnelle

cf . Schall (1991)

Présentation de la méthode (en 2 étapes) :

▷ conditionnellement à U : GLM

- estimation de β dans GLM : moindres carrés pondérés

$$X'W_\beta^{-1}X\beta = X'W_\beta^{-1}z_\beta$$

 - sur la variable de travail : $z_\beta = \eta + (y - \mu_\beta)g'(\mu_\beta)$
 - avec la matrice de variance : $W_\beta = diag\{g'(\mu_\beta)^2 \ \mathrm{var}(y)\}$

- modèle linéarisé $\mathcal{M}^{[t]}$ (procédure itérative) :

$$Z^{[t]} = X\beta + \varepsilon \quad \text{où} \quad \mathrm{var}(\varepsilon) = W_{\beta^{[t]}}$$

- ici : modèle linéarisé $\mathcal{M}_u^{[t]}$:

$$Z^{[t]} = X\beta + ZU + \varepsilon \quad \text{où} \quad \text{var}(\varepsilon) = W_{\beta^{[t]},u^{[t]}} = W^{[t]}$$

▷ déconditionnement : estimation des paramètres dans le LMM :

- étape 1 : estimation β, u : équations de Henderson

$$\begin{pmatrix} X'W^{[t]-1}X & X'W^{[t]-1}Z \\ Z'W^{[t]-1}X & Z'W^{[t]-1}Z + D^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ u \end{pmatrix} = \begin{pmatrix} X'W^{[t]-1}z^{[t]} \\ Z'W^{[t]-1}z^{[t]} \end{pmatrix}$$

- étape 2 : estimation de σ^2 à l'aide de $\hat{u}$

↪ procédure itérative et alternante en β, u et σ^2

ou système direct / démarche EM

Autres éclairages

▷ Quasi-vraisemblance pénalisée (PQL) :

cf. Breslow & Clayton (1993)

$$\begin{aligned} l(\beta,\sigma^2;y) &= |D|^{-1/2} \int exp\{-\textstyle\sum_{i,j} d(y_{ij},\mu_{ij},\phi) - \frac{1}{2}u'D^{-1}u\}\, du \\ &= |D|^{-1/2} \int exp\{-\kappa(u)\}\, du \end{aligned}$$

Approximation de Laplace en $\tilde{u}$ solution de $\kappa'(u) = 0$:

$$L(\beta,\sigma^2;y) \simeq -\frac{1}{2} ln|D| - \frac{1}{2} ln|\kappa''(\tilde{u})| - \kappa(\tilde{u})$$

et on aboutit à : $L(\beta,\sigma^2;y) \simeq -\sum_{i,j} d(y_{ij},\mu_{ij},\phi) - \frac{1}{2}\tilde{u}'D^{-1}\tilde{u}$

$\hookrightarrow$ $\hat{\beta}$ et $\hat{u}$ maximisent conjointement la log-quasi-vraisemblance pénalisée : $-\sum_{i,j} d(y_{ij},\mu_{ij},\phi) - \frac{1}{2}u'D^{-1}u$

$\hookrightarrow$ $\hat{u}$ utilisé pour estimer σ^2

$\hookrightarrow$ mêmes solutions que celles du système de Henderson précédent.

▷ A posteriori bayésien :

Stiratelli et al. (1984)

• a priori : diffus sur β et normal sur U

• maximisation en β et u de la vraisemblance a posteriori → étape 1

• approximation a posteriori sur U par une loi normale multivariée et estimation de σ^2 → étape 2

▷ h-vraisemblance :

Lee & Nelder (1996)

Avantages :

- facilité de mise en œuvre
- rapidité numérique
- pas de limite sur le nombre ni le type d'effets aléatoires

Inconvénients :

- léger biais (correction proposée par *Lin & Breslow (1996)*)
- hypothèse de normalité des effets
- faible composante de la variance

Logiciels :

- *cf. Wolfinger (1993)* *SAS - macro glimmix*
- approche développée dans les modèles non linéaires mixtes

  cf. Lindström & Bates (1990) (données normales - lien non linéaire)

  Splus - fonction nlme

3 - Approche marginale

cf. Gilmour, Anderson & Rae (1985, 1988)

▷ Principe général :

• calcul des 2 premiers moments marginaux

• construction de la fonction de quasi-vraisemblance

• approximation de la matrice de variance marginale sous forme de la structure de covariance d'un LMM : $R + ZDZ'$

• estimation à l'aide des équations de Henderson

▷ Description de la méthode :

• les moments marginaux $(E \ / \ V)$:

- calcul direct :
  . cas binomial - lien probit
  . cas poisson - lien log
  . cas exponentiel - lien log
- par approximation :
  . cas binomial - lien logit
  . cas poisson - lien identité
  . cas exponentiel - lien inverse

moments conditionnels :

$$\begin{aligned} E(Y_i|U) &= \mu_{u,i} = h(x_i'\beta + z_i'u) \\ V(Y_i|U) &= v(\mu_{u,i}) \end{aligned}$$

moments marginaux (approximation) :

$$\begin{aligned} E(Y_i) &= h(\eta_i^*) = \mu_i \quad \text{avec} \quad \eta^* = MX\beta + C \\ V(Y_i) &\approx V_0 + KMZDZ'MK = \tilde{V} \quad \text{avec} \quad K = diag\{h'(\eta_i^*)\} \end{aligned}$$

- maximisation de la fonction de quasi-vraisemblance

$$P'\tilde{V}^{-1}(y-\mu)=0$$

avec $P=\dfrac{\partial\mu}{\partial\beta'}=KMX=LX$

Algorithme itératif :

$$(X'W^{[t]-1}X)\beta^{[t+1]}=X'W^{[t]-1}z^{[t]}$$

où
$$\begin{aligned} z^{[t]} &= X\beta^{[t]}+L^{[t]-1}(y-\mu^{[t]}) \\ W^{[t]} &= L^{[t]-1}\tilde{V}^{[t]}L^{[t]-1} \end{aligned}$$

donc moindres carrés généralisés avec la matrice de variance :

$$\tilde{W}=L^{-1}V_0L^{-1}+ZDZ'$$

$\hookrightarrow$ structure de variance d'un LMM

▷ Inconvénients :

• approximations valables pour les faibles valeurs des composantes de la variance

• à définir au cas par cas

• trouver l'échelle η^*

▷ Logiciels :

• méthode peu pratiquée

• pas présente sur les logiciels standards

Extension des GLMM ... avec hétérogénéité

$\hookrightarrow$ introduction d'une hétérogénéité sur la variance des effets aléatoires

• idée : un même effet aléatoire peut ne pas avoir la même variance selon l'environnement dans lequel il se réalise
cf. Foulley & Quaas (1995)

• modélisation : au niveau du prédicteur linéaire
dans l'environnement i :

$$\eta_i = X_i \beta + \sigma_i Z_i u$$

• estimation : approche conditionnelle
$\hookrightarrow$ modèle linéarisé $\mathcal{M}_u^{[t]}$: LMM pour la variable de travail
$\hookrightarrow$ LMM à variances hétérogènes : algorithme EM permet l'estimation