

REML

Jean-Louis Foulley
INRA/Génétique quantitative/78350 Jouy-en-Josas
(foulley@jouy.inra.fr)

Plan

- Introduction
- Exemple simple
 - ML
 - Correction par l'espérance conditionnelle
 - Factorisation de la vraisemblance
- Cas général
 - Concept de vraisemblance marginale
 - Application au modèle linéaire mixte gaussien
 - Interprétation bayésienne
 - Traitement par EM
 - Vraisemblance profilée ajustée
 - Approximations de Wolfinger et de Kachman-Gianola
- Exemples
 - Données de croissance
 - Données d'hormone de gestation
- Conclusion
- Bibliographie, logiciels

Introduction

- Méthodes quadratiques
 - Estimateurs quadratiques, sans biais dérivés de l'ANOVA et méthode de Yates: Henderson I, II, III (1953)
- Minque
 - Estimateurs quadratiques, sans biais et localement de norme (variance) minimum: CR Rao (1971), LaMotte (1970,1973)
- Maximum de vraisemblance
 - Estimation ML des effets fixes et des paramètres de dispersion dans les modèles linéaires mixtes: Hartley & JNK Rao (1967)
 - Estimation ML basée sur des résidus libres des effets fixes (contrastes d'erreur): Patterson & Thompson (1971); Harville (1977); Searle (1979)

Pourquoi REML?/Ex simple

$$y_i \sim_{\text{iid}} \mathcal{N}(\mu, \sigma^2); i = 1, 2, \dots, N; \mathbf{y} = \{y_i\}$$

$$-2L(\mu, \sigma^2; \mathbf{y}) = N(\ln 2\pi + \ln \sigma^2) + \sum_{i=1}^N (y_i - \mu)^2 / \sigma^2$$

$$\text{Posons: } s^2 = \sum_{i=1}^N (y_i - \bar{y})^2 / N$$

$$\partial(-2L/N) / \partial \mu = 0 \Rightarrow \hat{\mu} = \bar{y}$$

$$\partial(-2L/N) / \partial \sigma^2 = 0 \Rightarrow \hat{\sigma}_{ML}^2 = s^2 + (\bar{y} - \hat{\mu})^2$$

$$\boxed{\hat{\sigma}_{ML}^2 = s^2}, \forall N \geq 2 \text{ et } \boxed{E(\hat{\sigma}_{ML}^2) = (N-1)\sigma^2 / N}$$

Exemple simple/espérance conditionnelle

Si μ connu, $\hat{\sigma}^2 = s^2 + (\bar{y} - \mu)^2 \geq s^2$

μ étant inconnu, que faire?

Prédire la contribution moyenne de $(\bar{y} - \mu)^2$

via $E[(\bar{y} - \mu)^2 | \mathbf{y}, \sigma^2]$

$$\frac{\bar{y} - \mu}{\sqrt{\sigma^2 / N}} \sim \mathcal{N}(0,1) \Leftrightarrow \begin{cases} \bar{y} | \mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2 / N) \\ \mu | \bar{y}, \sigma^2 \sim \mathcal{N}(\bar{y}, \sigma^2 / N) \end{cases}$$

$$E[(\bar{y} - \mu)^2 | \mathbf{y}, \sigma^2] = \text{Var}(\mu | \bar{y}, \sigma^2) = \sigma^2 / N$$

$$\hat{\sigma}^2 = s^2 + \hat{\sigma}^2 / N \Rightarrow \hat{\sigma}^2 (1 - 1/N) = s^2 \Rightarrow \hat{\sigma}^2 = Ns^2 / (N - 1)$$

Ex simple/factorisation de la vraisemblance

$$\mathbf{y}_{(N \times 1)} = \{y_i\} \Leftrightarrow \mathbf{y}^*_{(N \times 1)} = (\mathbf{z}'_{N-1}, \bar{y})'$$

$$\mathbf{z}_{N-1} = \{z_i = y_i - \bar{y}; i = 1, 2, \dots, N - 1\}$$

$$p_Y(\mathbf{y} | \mu, \sigma^2) = p_{Y^*}(\mathbf{y}^* | \mu, \sigma^2) |J|$$

$$p_{Y^*}(\mathbf{y}^* | \mu, \sigma^2) = p_Z(\mathbf{z}_{N-1} | \sigma^2) p_{\bar{Y}}(\bar{y} | \mu, \sigma^2)$$

$$L(\mu, \sigma^2; \mathbf{y}) = L_1(\sigma^2; \mathbf{z}_{N-1}) + L_2(\mu, \sigma^2; \bar{y}) + cste$$

Ex simple/factorisation de la vraisemblance

$$\begin{aligned} \mathbf{z}_{N-1} &\sim \mathcal{N}(\mathbf{0}, \mathbf{V}_Z); \mathbf{V}_Z = \sigma^2 (\mathbf{I}_{N-1} - \mathbf{J}_{N-1} / N) \\ -2L_1(\sigma^2; \mathbf{z}_{N-1}) &= (N-1)(\ln 2\pi + \ln \sigma^2) - \ln N \\ &\quad + \left[\sum_{i=1}^N (y_i - \bar{y})^2 \right] / \sigma^2 \\ \frac{\partial(-2L_1)}{\partial \sigma^2} &= \frac{N-1}{\sigma^2} - \frac{Ns^2}{\sigma^4} \Rightarrow \boxed{\hat{\sigma}^2 = Ns^2 / (N-1)}; N \geq 2 \end{aligned}$$

Cas général/concept de vraisemblance marginale

Ce concept développé par Kalbfleisch et Sprott (1970).

Transformation biunivoque de \mathbf{y} en $(\mathbf{u}', \mathbf{v}')$ telle que les deux conditions suivantes

portant sur $f(\mathbf{u}, \mathbf{v} | \boldsymbol{\beta}, \gamma) = f(\mathbf{v} | \boldsymbol{\beta}, \gamma) g(\mathbf{u} | \mathbf{v}, \boldsymbol{\beta}, \gamma)$ soient réalisées:

a) $f(\mathbf{v} | \boldsymbol{\beta}, \gamma) = f(\mathbf{v} | \gamma)$

b) $g(\mathbf{u} | \mathbf{v}, \boldsymbol{\beta}, \gamma)$ «contains no available information concerning γ in the

absence of knowledge of $\boldsymbol{\beta}$ »

Barnard mit en avant le caractère indissociable des informations imputables à $\boldsymbol{\beta}$ et

γ dans $g(\mathbf{u} | \mathbf{v}, \boldsymbol{\beta}, \gamma)$

(«This information is inextricably mixed up with the nuisance parameters»).

Patterson & Thompson (1971) considèrent la vraisemblance de $\mathbf{v} = (\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{OLS}$ d'où l'appellation de "residual" likelihood

$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ projecteur des moindres carrés

$\mathbf{v} = \mathbf{S}\mathbf{y} = (\mathbf{I}_N - \mathbf{H})\mathbf{y} = \hat{\boldsymbol{\varepsilon}}$: résidu de moindres carrés

$\mathbf{v} = (\mathbf{I}_N - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = (\mathbf{I}_N - \mathbf{H})\boldsymbol{\varepsilon}$

Expression de la logvraisemblance résiduelle

Harville (1977) propose de considérer $N - r(\mathbf{X})$ éléments linéairement indépendants de \mathbf{v} appelés contrastes d'erreur (notés $\mathbf{K}'\mathbf{y}$)

$$-2L(\boldsymbol{\gamma}; \mathbf{K}'\mathbf{y}) = (N - p) \ln 2\pi + \ln |\mathbf{K}'\mathbf{V}\mathbf{K}| + \mathbf{y}'\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}'\mathbf{y}$$

Searle (1979) montre que

$$-2L(\boldsymbol{\gamma}; \mathbf{K}'\mathbf{y}) = (N - p) \ln 2\pi + \ln |\mathbf{V}| + \ln |\underline{\mathbf{X}}'\mathbf{V}^{-1}\underline{\mathbf{X}}| + \mathbf{y}'\underline{\mathbf{P}}\mathbf{y}$$

$$\underline{\mathbf{P}}\mathbf{y} = \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}) = \mathbf{V}^{-1}(\mathbf{I} - \mathbf{Q})$$

Dans les équations: $\underline{\mathbf{P}}$ (REML) remplace \mathbf{V}^{-1} (ML)
comparaison avec la vraisemblance profilée

$$-2L_p(\boldsymbol{\gamma}; \mathbf{y}) = N \ln 2\pi + \ln |\mathbf{V}| + \mathbf{y}'\underline{\mathbf{P}}\mathbf{y}$$

Vraisemblance profilée

$$\text{Max}_{\beta, \gamma} L(\beta, \gamma; \mathbf{y}) = \text{Max}_{\gamma} \text{Max}_{\beta} L(\beta, \gamma; \mathbf{y})$$

$$\text{Max}_{\beta} L(\beta, \gamma; \mathbf{y}) = L[\hat{\beta}(\gamma), \gamma; \mathbf{y}] = L_p(\gamma; \mathbf{y})$$

$L_p(\gamma; \mathbf{y})$: "profile likelihood" Cox & Reid (1987)

Ici $\hat{\beta}(\gamma)$ solution des eq GLS $\hat{\beta}(\gamma) = (\mathbf{X}' \mathbf{V}_{\gamma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}_{\gamma}^{-1} \mathbf{y}$

$$-2 L_p(\gamma; \mathbf{y}) = N \ln(2\pi) + \ln |\mathbf{V}_{\gamma}| + (\mathbf{y} - \mathbf{X} \hat{\beta}_{\gamma})' \mathbf{V}_{\gamma}^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta}_{\gamma})$$

$$-2 L_p(\gamma; \mathbf{y}) = N \ln(2\pi) + \ln |\mathbf{V}| + \mathbf{y}' \mathbf{P} \mathbf{y}$$

ML-REML/remarques

Argument de Cressie et Lahiri (1993)

$$1) \frac{\partial [L(\gamma; \mathbf{K}' \mathbf{y})]}{\partial \gamma_k} = \frac{1}{2} [\mathbf{y}' \mathbf{P} \mathbf{V}_k \mathbf{P} \mathbf{y} - \text{tr}(\mathbf{P} \mathbf{V}_k)]$$

$$2) \frac{\partial [L_p(\gamma; \mathbf{y})]}{\partial \gamma_k} = \frac{1}{2} [\mathbf{y}' \mathbf{P} \mathbf{V}_k \mathbf{P} \mathbf{y} - \text{tr}(\mathbf{V}^{-1} \mathbf{V}_k)]$$

$$\text{Par définition } E \left(\frac{\partial [L(\gamma; \mathbf{K}' \mathbf{y})]}{\partial \gamma_k} \right) = 0 \text{ vs } E \left(\frac{\partial [L_p(\gamma; \mathbf{y})]}{\partial \gamma_k} \right) \neq 0$$

Cf Théorie des fonctions estimantes: Godambe (1960, 1991)

Liang & Zeger (1995)

1) FE sans biais; 2) FE biaisée

Interprétation bayésienne

concept de vraisemblance marginale, cette fois au sens bayésien du terme (Dawid, 1980) par intégration des paramètres parasites.

$$p(\mathbf{y}|\boldsymbol{\gamma}) = \int p(\mathbf{y}, \boldsymbol{\beta}|\boldsymbol{\gamma}) d\boldsymbol{\beta}$$

$$p(\mathbf{y}|\boldsymbol{\gamma}) = \int p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma})p(\boldsymbol{\beta}|\boldsymbol{\gamma}) d\boldsymbol{\beta}$$

Ici prendre $p(\boldsymbol{\beta}|\boldsymbol{\gamma}) = cste$

Interprétation bayésienne/suite

Soit à expliciter $p(\mathbf{y}|\boldsymbol{\gamma}) = \int p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) d\boldsymbol{\beta}$

$$p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\theta}) = (2\pi)^{-N/2} |\mathbf{V}|^{-1/2} \exp\left[-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) / 2\right]$$

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{V}^{-1} \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

$$p(\mathbf{y}|\boldsymbol{\gamma}) = (2\pi)^{-(N-p)/2} |\mathbf{V}|^{-1/2} |\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}|^{-1/2} \exp\left[-(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) / 2\right]$$

$$-2 \ln p(\mathbf{y}|\boldsymbol{\gamma}) = -2 L(\boldsymbol{\gamma}; \mathbf{y}) = (N-p) \ln 2\pi + \ln |\mathbf{V}| + \ln |\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}| + \mathbf{y}' \mathbf{P} \mathbf{y}$$

Résultat du à Harville (1974)

REML et équivalences

Dispositifs équilibrés: REML=ANOVA

REML=MINQUE itéré

$$\gamma_{REML} = \text{Arg max ln } \underbrace{p(\mathbf{y} | \gamma)}_{\text{vrais.marginale}} \Leftrightarrow \text{REML} = \text{MaxMarginalLikelihood}$$

en prenant un a priori non informatif sur β

$$\gamma_{REML} = \text{Arg max ln } \underbrace{p(\gamma | \mathbf{y})}_{\text{a posteriori marginal}} \Leftrightarrow \text{REML} = \text{MAP de la distr. marginale}$$

en prenant un a priori non informatif sur β et idem sur γ

EM-REML dans un modèle simple

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$\mathbf{u}_{(q \times 1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{G} = \sigma_1^2 \mathbf{I}_q)$$

$$\mathbf{e}_{(N \times 1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{R} = \sigma_0^2 \mathbf{I}_N)$$

$$\mathbf{V} = \text{Var}(\mathbf{y}) = \sigma_1^2 \mathbf{Z}\mathbf{Z}' + \sigma_0^2 \mathbf{I}_N$$

Données manquantes: $\mathbf{z} = (\beta', \mathbf{u}')$

$$\mathbf{x} = (\mathbf{y}', \beta', \mathbf{u}')'; \phi = (\sigma_1^2, \sigma_0^2)'$$

$$L(\phi; \mathbf{x}) = L_0(\sigma_0^2; \mathbf{e}) + L_1(\sigma_1^2; \mathbf{u}) + cste$$

EM-REML modèle simple/ phase E

$$\text{Phase E: } Q(\boldsymbol{\phi}; \boldsymbol{\phi}^{[t]}) = E_c^{[t]} [L(\boldsymbol{\phi}; \mathbf{x})]$$

$$Q(\boldsymbol{\phi}; \boldsymbol{\phi}^{[t]}) = Q_0(\sigma_0^2; \boldsymbol{\phi}^{[t]}) + Q_1(\sigma_1^2; \boldsymbol{\phi}^{[t]})$$

$$Q_1(\sigma_1^2; \boldsymbol{\phi}^{[t]}) = E_c^{[t]} [L_1(\sigma_1^2; \mathbf{u})]$$

$$= -\frac{1}{2} \left[q \ln 2\pi + q \ln \sigma_1^2 + E_c^{[t]}(\mathbf{u}'\mathbf{u}) / \sigma_1^2 \right]$$

$$E_c^{[t]}(\cdot) \text{ par rapport à } \mathbf{z} \mid \mathbf{y}, \boldsymbol{\phi} = \boldsymbol{\phi}^{[t]}$$

$$\frac{\partial}{\partial \sigma_1^2} Q_1(\sigma_1^2; \boldsymbol{\phi}^{[t]}) = 0 \Rightarrow \sigma_1^{2[t+1]} = E_c^{[t]}(\mathbf{u}'\mathbf{u}) / q$$

$$\text{idem pour } \sigma_0^{2[t+1]} = E_c^{[t]}(\mathbf{e}'\mathbf{e}) / N$$

EM-REML vs ML

REML

$$E \left(\underbrace{\mathbf{e}'\mathbf{e}}_{N\sigma_0^{2[t+1]} \mid \sigma_0^{2[t+1]}} \mid \sigma_0^{2[t+1]} \right) = E_c^{[t]} \left(\mathbf{e}'\mathbf{e} \mid \mathbf{y}, \sigma_0^2 = \sigma_0^{2[t]}, \sigma_1^2 = \sigma_1^{2[t]} \right)$$

$$E \left(\underbrace{\mathbf{u}'\mathbf{u}}_{q\sigma_1^{2[t+1]} \mid \sigma_1^{2[t+1]}} \mid \sigma_1^{2[t+1]} \right) = E_c^{[t]} \left(\mathbf{u}'\mathbf{u} \mid \mathbf{y}, \sigma_0^2 = \sigma_0^{2[t]}, \sigma_1^2 = \sigma_1^{2[t]} \right)$$

ML: idem mais distribution conditionnelle sachant

$$\mathbf{y}, \boldsymbol{\beta} = \boldsymbol{\beta}^{[t]}, \sigma_0^2 = \sigma_0^{2[t]}, \sigma_1^2 = \sigma_1^{2[t]}$$

EM-REML modèle simple/ formules

$$E_c^{[l]}(\mathbf{e}'\mathbf{e}) = E(\mathbf{e} | \mathbf{y}, \boldsymbol{\phi} = \boldsymbol{\phi}^{[l]})' E(\mathbf{e} | \mathbf{y}, \boldsymbol{\phi} = \boldsymbol{\phi}^{[l]}) \\ + \text{tr} \left[\text{var}(\mathbf{e} | \mathbf{y}, \boldsymbol{\phi} = \boldsymbol{\phi}^{[l]}) \right]$$

$$p_D^{[l]} = \text{rang}(\mathbf{X}) + q - \lambda^{[l]} \text{tr}(\mathbf{C}_{uu}^{[l]})$$

Facteur de complexité du modèle mixte développé
par Spiegelhalter et al, (2002)

$$\hat{\sigma}_0^{2[l+1]} = (\hat{\mathbf{e}}^{[l]'} \hat{\mathbf{e}}^{[l]} + p_D^{[l]} \hat{\sigma}_0^{2[l]}) / N \quad \text{ou} \quad (N - \hat{p}_D) \hat{\sigma}_0^2 = \hat{\mathbf{e}}' \hat{\mathbf{e}}$$

$$\text{Modèle fixe : } (N - \text{rang}(\mathbf{X})) \hat{\sigma}_0^2 = \hat{\mathbf{e}}' \hat{\mathbf{e}}$$

Vraisemblance profilée ajustée/ approche de Liao-Lipsitz pour GLMM

Liao, Lipsitz, Biometrika, 2002, 89, 401-409

1) $\mathbf{y} | \mathbf{u}, \mathbf{b}$ = modèle GLM exp. classique

$$E(y_i | \mathbf{u}) = \mu_{u,i} = h(\eta_{u,i}); \quad \eta_{u,i} = \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{u}$$

$$\text{Var}(y_i | \mathbf{u}) = \phi v(\mu_{u,i}), \phi \text{ connu}$$

2) $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\theta)$

Vraisemblance profilée ajustée/
approche de Liao-Lipsitz pour GLMM

Ajustement de la vraisemblance profilée
(McCullagh&Tibshirani,1990)

$\hat{\boldsymbol{\theta}}_b^y = \text{MLE de } \boldsymbol{\theta} \text{ basé sur } \mathbf{y} \text{ à } \mathbf{b} \text{ connu}$

Si \mathbf{b} connu, $L(\boldsymbol{\theta}; \mathbf{y}, \mathbf{u}) = \ln p(\mathbf{u}) + C$

car $\mathbf{y} | \mathbf{u}$ ne dépend que de \mathbf{b}

$$E : Q(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = -\frac{1}{2} \left(\ln |\mathbf{V}_\theta| + \text{tr} \left[\mathbf{V}^{-1} E(\mathbf{u}\mathbf{u}' | \mathbf{y}, \boldsymbol{\theta}^*, \mathbf{b}) \right] \right)$$

$$M : \frac{\partial}{\partial \boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)_{\boldsymbol{\theta}^* = \boldsymbol{\theta}} = \mathbf{h} \left(\boldsymbol{\theta}, \underbrace{E(\mathbf{u}\mathbf{u}' | \mathbf{y}, \boldsymbol{\theta}, \mathbf{b})}_S \right) = 0$$

Vraisemblance profilée ajustée/
approche de Liao-Lipsitz pour GLMM/suite

$$\mathbf{h}(\boldsymbol{\theta}, \mathbf{S})_k = \text{tr} \left[\mathbf{V}_\theta^{-1} (\mathbf{S} - \mathbf{V}_\theta) \mathbf{V}_\theta^{-1} \frac{\partial \mathbf{V}_\theta}{\partial \theta_k} \right]$$

Si $E(\mathbf{S}) = \mathbf{V}_\theta$ alors $E[\mathbf{h}(\boldsymbol{\theta}, \mathbf{S})] = 0$

(fonction estimante sans biais)

Vrai pour $\mathbf{h}[\boldsymbol{\theta}, E(\mathbf{u}\mathbf{u}' | \mathbf{y}, \boldsymbol{\theta}, \mathbf{b})] = 0$

Faux pour $\mathbf{h}[\boldsymbol{\theta}, E(\mathbf{u}\mathbf{u}' | \mathbf{y}, \boldsymbol{\theta}, \hat{\mathbf{b}}_\theta^y)] = 0$

$\hat{\mathbf{b}}_\theta^y = \text{MLE de } \mathbf{b} \text{ à } \boldsymbol{\theta} \text{ connu}$

Eq MLE basée sur la vraisemblance profilée.

Vraisemblance profilée ajustée/ approche de Liao-Lipsitz pour GLMM/suite

Evaluer le biais $B(\boldsymbol{\theta}, \mathbf{b}) = E_{\mathbf{y}|\boldsymbol{\theta}, \mathbf{b}} \left[E(\mathbf{u}\mathbf{u}' | \mathbf{y}, \boldsymbol{\theta}, \hat{\mathbf{b}}_{\boldsymbol{\theta}}^y) \right] - E_{\mathbf{y}|\boldsymbol{\theta}, \mathbf{b}} \left[E(\mathbf{u}\mathbf{u}' | \mathbf{y}, \boldsymbol{\theta}, \mathbf{b}) \right]$ par des méthodes de Monte Carlo

En fait, on résout $h \left[\boldsymbol{\theta}, E(\mathbf{u}\mathbf{u}' | \mathbf{y}, \boldsymbol{\theta}, \hat{\mathbf{b}}_{\boldsymbol{\theta}}^y) + B(\boldsymbol{\theta}, \hat{\mathbf{b}}_{\boldsymbol{\theta}}^y) \right] = 0$

Modèles NL/appr. de Wolfinger

Wolfinger, 1993, Biometrika;

Wolfinger & Lin, 1997, CSDA

$$1) y_i = f(\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{u}) + e_i$$

$$\mathbf{u} \sim \mathcal{N}(0, \mathbf{G}); \mathbf{e} \sim \mathcal{N}(0, \mathbf{R})$$

$$2) \tilde{\mathbf{X}} = \left\{ \tilde{\mathbf{x}}_i' = f'(\mathbf{x}_i' \hat{\boldsymbol{\beta}} + \mathbf{z}_i' \hat{\mathbf{u}}) \tilde{\mathbf{x}}_i' \right\}; \text{idem pour } \tilde{\mathbf{Z}}$$

$$\begin{bmatrix} \tilde{\mathbf{X}}' \mathbf{R}^{-1} \tilde{\mathbf{X}} & \tilde{\mathbf{X}}' \mathbf{R}^{-1} \tilde{\mathbf{Z}} \\ \tilde{\mathbf{Z}}' \mathbf{R}^{-1} \tilde{\mathbf{X}} & \tilde{\mathbf{Z}}' \mathbf{R}^{-1} \tilde{\mathbf{Z}} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{X}}' \mathbf{R}^{-1} \mathbf{w} \\ \tilde{\mathbf{Z}}' \mathbf{R}^{-1} \mathbf{w} \end{bmatrix}$$

$$w_i = y_i - f(\mathbf{x}_i' \hat{\boldsymbol{\beta}} + \mathbf{z}_i' \hat{\mathbf{u}}) + \tilde{\mathbf{x}}_i' \hat{\boldsymbol{\beta}} + \tilde{\mathbf{z}}_i' \hat{\mathbf{u}}$$

$$3) -2RL = (N - p) \ln 2\pi + \ln |\mathbf{V}| + \ln |\tilde{\mathbf{X}}' \mathbf{V}^{-1} \tilde{\mathbf{X}}| + (\mathbf{w} - \tilde{\mathbf{X}} \hat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{w} - \tilde{\mathbf{X}} \hat{\boldsymbol{\beta}})$$

itérer sur 2) et max 3)

version REML de Lindstrom & Bates (1990)

Modèles NL/appr. de Kachman-Gianola

Kachman & Gianola, 1984; Rodriguez et al, 2000

$$2) \begin{bmatrix} \tilde{\mathbf{X}}' \mathbf{R}^{-1} \tilde{\mathbf{X}} & \tilde{\mathbf{X}}' \mathbf{R}^{-1} \tilde{\mathbf{Z}} \\ \tilde{\mathbf{Z}}' \mathbf{R}^{-1} \tilde{\mathbf{X}} & \tilde{\mathbf{Z}}' \mathbf{R}^{-1} \tilde{\mathbf{Z}} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{X}}' \mathbf{R}^{-1} \mathbf{w} \\ \tilde{\mathbf{Z}}' \mathbf{R}^{-1} \mathbf{w} \end{bmatrix}$$

$$w_i = y_i - f(\mathbf{x}_i' \hat{\boldsymbol{\beta}} + \mathbf{z}_i' \hat{\mathbf{u}}) + \tilde{\mathbf{x}}_i' \hat{\boldsymbol{\beta}} + \tilde{\mathbf{z}}_i' \hat{\mathbf{u}}$$

3) raisonnement EM + espérance et variance de la loi conditionnelle de $(\mathbf{u} | \mathbf{y}, \text{par})$ approchée sous l'hypothèse de normalité de cette loi.

Ex: si $\mathbf{u} \sim \mathcal{N}(0, \sigma_u^2 \mathbf{I}_q)$

$$\sigma_u^{2[t+1]} = \left[\hat{\mathbf{u}}^{[t]'} \hat{\mathbf{u}}^{[t]} + \text{tr}(\mathbf{C}_{uu}^{[t]}) \right] / q$$

$\mathbf{C}_{uu}^{[t]}$ = bloc (u,u) dans l'inv. de la matrice des coefts

Exemples

- Données de croissance faciale à 4 âges (8, 10, 12, 14 ans) chez 11 filles et 16 garçons. Pothoff & Roy (1964); Verbeke & Molenberghs (1997, 2000)
- Dosages de la protéine sérique de gestation (PSP60) effectués de 2.5 à 9 mois chez 46 vaches de différents âges vêlant à la même saison (Foulley, Maugrion, 2001)

Ex de croissance/données

Growth measurements in 11 girls and 16 boys: Pothoff and Roy, 1964; Little and Rubin, 1987

Girl	Age (years)				Boy	Age (years)			
	8	10	12	14		8	10	12	14
1	210	200	215	230	1	260	250	290	310
2	210	215	240	255	2	215		230	265
3	205		245	260	3	230	225	240	275
4	235	245	250	265	4	255	275	265	270
5	215	230	225	235	5	200		225	260
6	200		210	225	6	245	255	270	285
7	215	225	230	250	7	220	220	245	265
8	230	230	235	240	8	240	215	245	255
9	200		220	215	9	230	205	310	260
10	165		190	195	10	275	280	310	315
11	245	250	280	280	11	230	230	235	250
					12	215		240	280
					13	170		260	295
					14	225	255	255	260
					15	230	245	260	300
					16	220		235	250

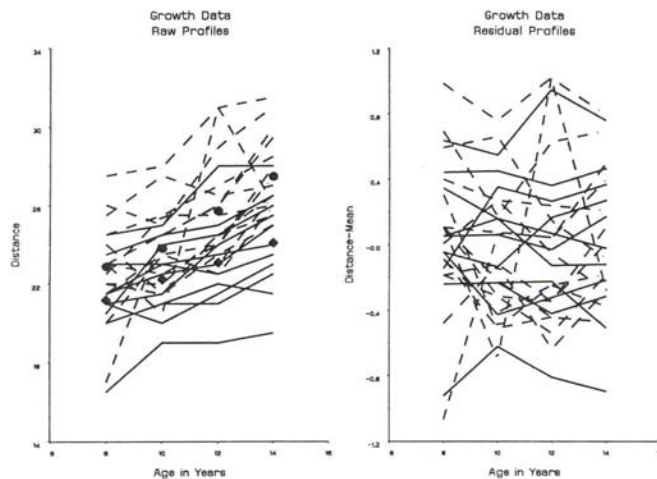
distance from the centre of the pituitary to the pterygomaxillary fissure (unit: 10^{-6} m)

02/03/2004

Monolix, Orsay-JL Foulley

27

Exemple "croissance faciale": profils



02/03/2004

Monolix, Orsay-JL Foulley

28

Modèle [0] : Modèle à effets fixes

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i \text{ avec } \mathbf{V}_i = \mathbf{I}_{n_i} \sigma_e^2$$

Modèle [1] : Modèle mixte à intercept aléatoire

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i + \mathbf{e}_i \text{ avec } \mathbf{V}_i = \mathbf{Z}_i \mathbf{Z}_i' g_{00} + \sigma_e^2 \mathbf{I}_{n_i}$$

Modèle [2] : Modèle avec processus temporel

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{W}_i \text{ avec } \mathbf{V}_i = \sigma^2 \mathbf{H}_i \quad \mathbf{H}_i = \{h_{i,t,t'} = \rho^{|t-t'|}\}$$

Modèle [3] : Modèle avec processus temporel et erreurs de mesure

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{W}_i + \mathbf{e}_i \text{ avec } \mathbf{V}_i = \sigma^2 \mathbf{H}_i + \sigma_e^2 \mathbf{I}_{n_i}$$

Modèle [4] : Modèle mixte à intercept aléatoire et processus temporel

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_{0i} \mathbf{u}_{0i} + \mathbf{W}_i \text{ avec } \mathbf{V}_i = \mathbf{Z}_i \mathbf{Z}_i' g_{00} + \sigma^2 \mathbf{H}_i$$

Modèle [5] : Modèle de régression linéaire à coefficients aléatoires

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_{0i} \mathbf{u}_{0i} + \mathbf{Z}_{1i} \mathbf{u}_{1i} + \mathbf{e}_i \text{ avec } \mathbf{V}_i = \mathbf{Z}_i \mathbf{G}_i \mathbf{Z}_i' + \sigma_e^2 \mathbf{I}_{n_i} \quad \mathbf{G}_i = \begin{pmatrix} g_{00} & g_{01} \\ g_{01} & g_{11} \end{pmatrix}$$

ML vs REML/exemple « croissance »

Variance components

Method	Model	Random effects ^a			Time process ^b		Error ^c
		g_{00}	g_{01}	g_{11}	σ^2	ρ	σ_e^2
REML	[0]						527.93
	[1]	337.24					207.49
	[2]				545.41	0.802	
	[3]				380.95	0.966	164.99
	[4]	331.42			213.60	0.239	
	[5]	835.50	-46.53	4.41			176.66
ML	[0]						506.60
	[1]	309.49					201.74
	[2]				510.95	0.792	
	[3]				342.74	0.971	168.70
	[4]	307.36			203.94	0.152	
	[5]	678.62	-34.99	3.37			177.00
	[6] ^f						

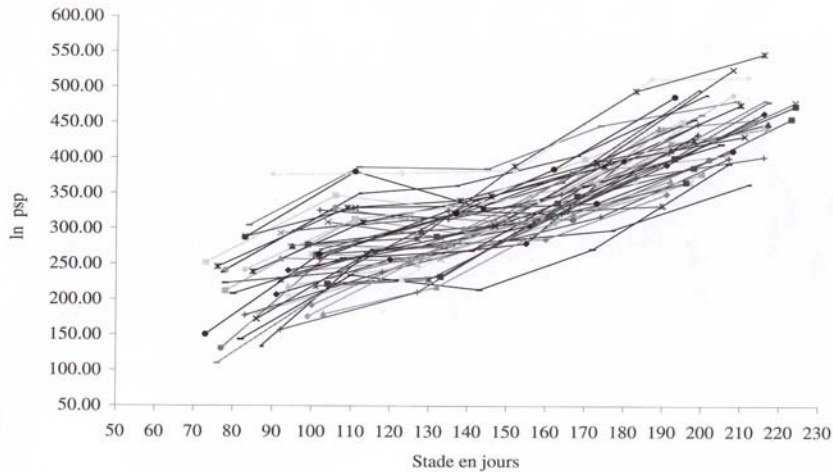
^aRandom effects model intercept (0) and/or slope (1) : fixed part with sex, linear regression on age varying according to sex

^bPower process: $Cov_{\rho} = \rho^s \sigma^s$

^cResidual in random regression models and measurement error σ_e^2 («nugget»)
for models involving $\mathbf{R}_i = \sigma^2 \mathbf{H}_i + \sigma_e^2 \mathbf{I}_i$

^fModel with an unspecified covariance structures (10 parameters)

PSP au cours de la gestation



02/03/2004

Monolix, Orsay-JL Foulley

31

ML vs REML/exemple « PSP »

Variance-covariance components of several models for the analysis of PSP60 concentration records

Method	Model	Random effects ^a			Time Process ^b		Error ^c σ_e^2	-2L
		g_{00}	g_{01}	g_{11}	σ^2	ρ		
REML	[0]						2386.33	2305.340
	[1]	1432.15					977.73	2205.338
	[2]				2490.51	0.989		2195.537
	[3]				2096.60	0.994	398.87	2193.071
	[4]	1182.43			1287.54	0.973		2188.937
	[5]	1365.54	-2.937	0.0944			776.36	2193.534
ML	[0]						2342.54	2310.117
	[1]	1363.09					966.38	2210.738
	[2]				2396.83	0.988		2202.404
	[3]				1976.45	0.994	423.92	2199.613
	[4]	1133.83			1244.61	0.971		2195.213
	[5]	1297.63	-2.888	0.0857			777.01	2199.739

^aRandom effects model, intercept (0) and/or slope (1) : fixed part with parity, linear regression on dayof blood sample collection varying according to parity

^bPower process : $Cov_{\mu} = \rho^{|i-j|} \sigma^2$

^cResidual in random regression models and measurement error σ_e^2 («nugget») for models involving

$$\mathbf{R}_i = \sigma^2 \mathbf{H}_i + \sigma_e^2 \mathbf{I}_{n_i}$$

02/03/2004

Monolix, Orsay-JL Foulley

32

Critères d'information/exemple PSP/résultats

Table. Likelihood statistics (REML, -2L, 2AIC, -2BIC) of several models for the analysis of PSP60 concentration records (10x log PSP60 in ng/ml)

Mod	Random Effects ^a			Time process ^b		Error ^c		Likelihood					
	g_{00}	g_{01}	g_{11}	σ^2	ρ	σ_e^2	#par	-2L	-2AIC ^e (-2BIC)	comp	$A[-2L]$	Distr ^d	P-value
[0]						2386.33	1	2305.340	2307.34 (2310.71)				
[1]	1432.12					977.73	2	2205.338	2209.34 (2216.07)	[1]-[0]	100.02	0:1	0
[2]				2490.51	0.989		2	2195.537	2199.54 (2206.26)				
[3]				2096.17	0.994	399.05	3	2193.071	2199.08 (2209.17)	[3]-[2]	1.827	0:1	0.0882
[4]	1182.43			1287.54	0.973		3	2188.937	2195.94 (2205.03)	[3]-[1]	12.267	0:1	0.0002
										[4]-[2]	6.600	0:1	0.0051
[5]	1365.54	-2.937	0.0944			776.37	4	2193.534	2201.54 (2214.99)	[4]-[1]	16.401	0:1	2E-5
										[5]-[1]	11.804	1:2	0.0016

^aRandom effects model with intercept (0) and/or slope (1): fixed part with parity, linear regression on day of blood sample collection

^bPOW process: $Cov_{\rho} = \rho^2 \sigma^2$

^cResidual in random regression models and measurement error σ_e^2 for models involving $\mathbf{R}_i = \sigma^2 \mathbf{H}_i + \sigma_e^2 \mathbf{I}_i$

^dAsymptotic distribution of the likelihood ratio under the null hypothesis: Chi-square or mixture of chi-squares

^e-2AIC (-2BIC) minus twice Akaike's (Schwarz's) criterion: #observations=218 on 46 subjects: #location parameters=4 (a regression line in each parity)

Conclusion

- ML-REML: problème du traitement des paramètres parasites
 - « Profile, Marginal, Conditional, Integrated, Modified likelihoods »
 - Réduction du biais (données manquantes MAR) mais MSE?
 - Équivalence: ANOVA eq, MINQUE itéré
- Obtention plus difficile de REML
 - Cas des modèles GLMM et NL
 - Vraisemblance profilée ajustée (FE)
 - Approche bayésienne

Bibliographie/modèles mixtes

References sur les modèles mixtes

- Davidian M, Giltinan DM (2003) Nonlinear models for repeated measurement data: an overview and update. *JABES* 8, 387-419
- Dempster A, Laird N, Rubin R (1977) Maximum likelihood estimation from incomplete data via the EM algorithm. *J R Stat Soc B* 39, 1-10
- Fouley JL (2001) Le modèle linéaire mixte. Cours ENSAI, polycopié.
- Fouley JL (2002) Algorithme EM et modèle mixte. *Journal de la Société Française de Statistique*, 143, 57-109
- Fouley JL, Delmas C, Robert-Granié C (2002) Méthodes du maximum de vraisemblance en modèle linéaire mixte. *Journal de la Société Française de Statistique*, 143, 5-51
- Hartley HO, Rao JNK (1967) Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 54, 93-108
- Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J Amer Stat Assoc* 72, 320-340
- Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Guelph, Ontario.
- Liao JG, Lipsitz SR (2002) A type of restricted maximum likelihood estimator of variance components in generalized linear mixed models. *Biometrika* 89, 401-409
- Lindley DV, Smith AFM (1972) Bayes Estimates for the Linear Model. *J R Statist Soc*, 34, 1-41
- Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545-554
- Rao CR, Kleffe J (1988) Estimation of variance components and applications. *Noth Holland series in statistics and probability*, Elsevier
- Searle SR (1979) Notes on variance component estimation. A detailed account of maximum likelihood and kindred methodology, Paper BU-673-M, Cornell University, Ithaca
- Searle SR, Casella G, McCulloch CE (1992) Variance components. *J Wiley & sons, New-York*
- Sorensen D., Gianola D. (2002), Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics, Springer Verlag, Berlin, 756 pages
- Verbeke G, Molenberghs G. Linear mixed models in practice, Springer Verlag, New York, 1997
- Wolfinger R (1993) Laplace's approximation for nonlinear mixed models. *Biometrika* 80, 791-795

02/03/2004

Monolix, Orsay-JL Fouley

35

Logiciels: sites Web

- ASREML: <http://www.vsn-intl.com/asreml/>
- DFREML: <http://aqbu.une.edu.au/~kmeyer/dfreml.html>
- BLUPF90: <http://nce.ads.uga.edu/~ignacy/newprograms.html>
- VCE5: <http://www.tzv.fal.de/~eg/>
- NLME: <http://cm.bell-labs.com/cm/ms/departments/sia/project/nlme/index.html>
- WINBUGS: <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>

02/03/2004

Monolix, Orsay-JL Fouley

36