

Quel BIC pour les modèles mixtes?

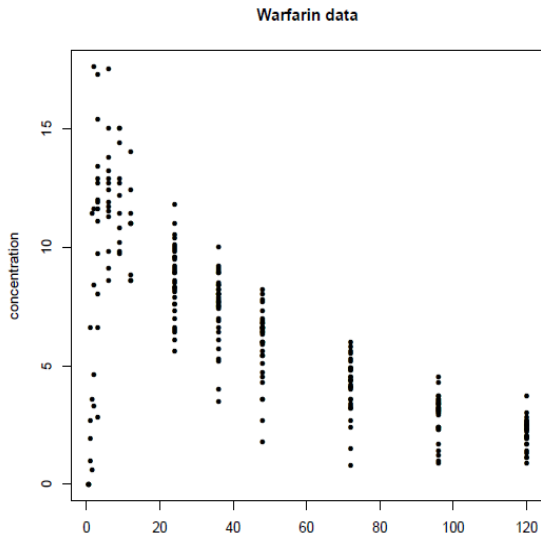
Maud Delattre¹, Marc Lavielle², Marie-Anne Poursat²

1. UMR 518 MIA, AgroParisTech
2. Laboratoire de Mathématiques, Université Paris-Sud

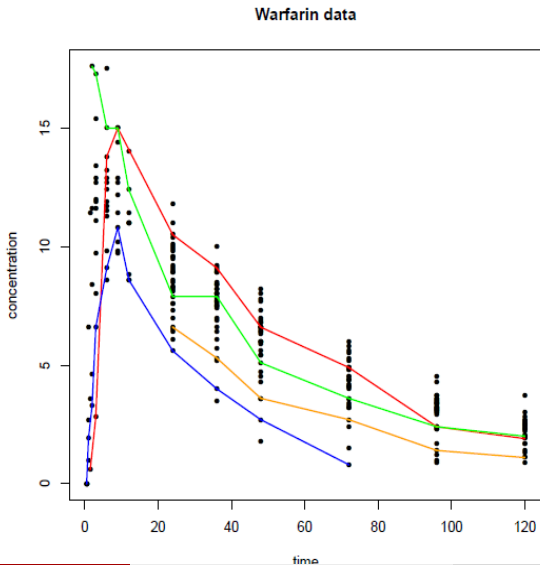


- ① Modèles à effets mixtes : formulation et exemples
- ② Sélection de variables
- ③ Résultat : pénalité du critère BIC
- ④ Comportement sur données d'études et simulées

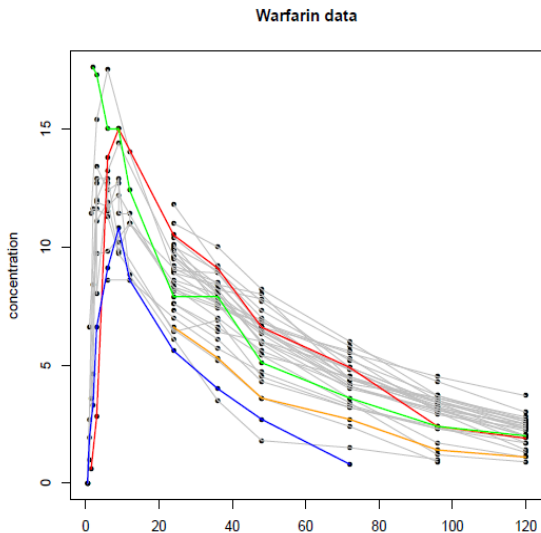
Mesures répétées ou longitudinales



individu i : $(t_{ij}, y_{ij}), j = 1, \dots, n_i$



N=32 individus



Modèle général : formulation hiérarchique

$Y_i = (y_{i1}, \dots, y_{in_i})'$: observations de l'individu i

y_{ij} : mesure au temps t_{ij}

Y_1, \dots, Y_N indépendants

1 Modèle des observations :

$$Y_i \sim p(\cdot | \psi_i; t_{ij})$$

\Leftrightarrow paramètres individuels ψ_i (vecteur $d \times 1$)

Modèle général : formulation hiérarchique

$Y_i = (y_{i1}, \dots, y_{in_i})'$: observations de l'individu i

y_{ij} : mesure au temps t_{ij}

Y_1, \dots, Y_N indépendants

1 Modèle des observations :

$$Y_i \sim p(\cdot | \psi_i; t_{ij})$$

\hookrightarrow paramètres individuels ψ_i (vecteur $d \times 1$)

2 Modèle des paramètres individuels

(variabilité inter-individus) :

$$g(\psi_i) = C_i \beta + \eta_i$$

β : vecteur $m \times 1$ des effets fixes,

η_i : vecteur $d \times 1$ des effets aléatoires, $\eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega)$

$C_i = C(c_i)$: matrice $d \times m$, dépend de covariables c_i ,
caractéristiques individuelles (poids, âge, ...)

Exemple linéaire

1. Observations :

$$y_{ij} = \psi_{0,i} + \psi_{1,i}t_{ij} + \psi_{2,i}t_{ij}^2 + \varepsilon_{ij}, \quad \varepsilon_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$$

2. Paramètres individuels :

$$\begin{cases} \psi_{0,i} = \mu_0 + \eta_{0,i} \\ \psi_{1,i} = \mu_1 + \alpha_1 c_i + \eta_{1,i} \\ \psi_{2,i} = \mu_2 + \alpha_2 c_i + \eta_{2,i} \end{cases}$$

ou encore

$$\underbrace{\begin{pmatrix} \psi_{0,i} \\ \psi_{1,i} \\ \psi_{2,i} \end{pmatrix}}_{\psi_i} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & c_i & 0 \\ 0 & 0 & 1 & 0 & c_i \end{pmatrix}}_{C_i} \underbrace{\begin{pmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \\ \alpha_1 \\ \alpha_2 \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} \eta_{0,i} \\ \eta_{1,i} \\ \eta_{2,i} \end{pmatrix}}_{\eta_i}$$

Paramètres de population : $\theta = (\beta, \sigma^2, \Omega)$

Modèle linéaire à effets mixtes

1. Observations :

$$y_{ij} = \psi_{0,i} + \psi_{1,i}t_{ij} + \psi_{2,i}t_{ij}^2 + \varepsilon_{ij}$$

2. Paramètres individuels :

$$\psi_i = C_i\beta + \eta_i$$

Modèle linéaire (Y_i) + modèle linéaire (ψ_i)

↔ Linear Mixed Effects model of Laird and Ware (1982) :

$$Y_i = X_i\beta + Z_i\eta_i + \varepsilon_i$$

où $X_i = X(c_i, t_i)$ et $Z_i = Z(c_i, t_i)$.

On a $Y_i \sim \mathcal{N}(X_i\beta, Z_i\Omega Z_i' + \sigma^2\text{Id})$.

Exemple Warfarin

1. Observations :

$$y_{ij} = f_{ij} + f_{ij}\varepsilon_{ij}$$

$$f_{ij} = f(t_{ij}, D_i, \psi_i)$$

2. Paramètres individuels :

$$\log \psi_i = C_i\beta + \eta_i$$

$$f(t, D, k_a, V, Cl) = \frac{D k_a}{V k_a - Cl} \left(e^{-(Cl/V)t} - e^{-k_a t} \right)$$

$$\begin{aligned} \log k_{a,i} &= \mu_1 \\ \log V_i &= \mu_V + \alpha_V c_i + \eta_{V,i} \\ \log Cl_i &= \mu_{Cl} + \eta_{Cl,i} \end{aligned}$$

Modèle non-linéaire à effets mixtes

1. Observations :

$$y_{ij} = f(t_{ij}, \psi_i) + h(t_{ij}, \psi_i)\varepsilon_{ij}, \quad \varepsilon_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$$

2. Paramètres individuels :

$$g(\psi_i) = C_i\beta + \eta_i$$

↔ Nonlinear Mixed Effects model of Davidian and Giltinan (1995), Pinheiro and Bates (2009), `nlme` dans R.

Modèles mixtes considérés

- Modèles linéaires à effets mixtes
- Modèles non-linéaires à effets mixtes
- Modèles linéaires et non-linéaires généralisés à effets mixtes (Schelldorfer *et al.* (2013) GLMMLasso)
- Modèles de Markov cachés à effets mixtes (M. Delattre (2012))

Sélection de variables

$$\begin{aligned} Y_i &= (y_{i1}, \dots, y_{in_i}) \sim p(\cdot | \psi_i) \\ \psi_i &\sim \mathcal{N}(C_i \beta, \Omega), \quad i = 1, \dots, N \\ \Omega &= \begin{pmatrix} 0 & 0 \\ 0 & \Omega_R \end{pmatrix} \end{aligned} \tag{1}$$

Sélectionner les covariables c_i pertinentes pour expliquer la variabilité inter-individus :

↔ sélectionner les composantes non nulles de β .

▶ Exemple Warfarin

Sélection de variables

Depuis 2006, nombreux articles :

- 1 Critères d'information mAIC, cAIC, BIC
 - 2 Méthodes LASSO
 - 3 méthodes bayésiennes, méthodes "fence"
- Ambiguïté de la sélection sur β et/ou sur les paramètres de Ω .
 - La plupart des méthodes de sélection supposent la **structure de covariance connue**.
 - Les méthodes de sélection proposées sont spécifiques des **modèles linéaires** à effets mixtes et des **modèles linéaires généralisés** à effets mixtes.

Sélection de variables

Bayesian information criterion :

$$BIC = -2\mathcal{L} + \text{pen}$$

Sélection de variables

Bayesian information criterion :

$$BIC = -2\mathcal{L}\mathcal{L} + \text{pen}$$

- Log-vraisemblance : $\mathcal{L}\mathcal{L} = \sum_{i=1}^N \log p(Y_i|\hat{\theta})$,

$$p(Y_i|\theta) = \int p(y_i|\psi_i)p(\psi_i|\theta)d\psi_i.$$

Sélection de variables

Bayesian information criterion :

$$BIC = -2\mathcal{L}\mathcal{L} + \text{pen}$$

- Log-vraisemblance : $\mathcal{L}\mathcal{L} = \sum_{i=1}^N \log p(Y_i|\hat{\theta})$,

$$p(Y_i|\theta) = \int p(y_i|\psi_i)p(\psi_i|\theta)d\psi_i.$$

- Pénalité :

$$\text{pen} = (\text{nb paramètres}) \times \log(\text{taille de l'échantillon})$$

Quelle taille d'échantillon ?

- Dans un modèle classique à **effets fixes** uniquement :

$n_{\text{tot}} = \sum_{i=1}^N n_i$ observations indépendantes :

y_{ij} , $i = 1, \dots, n_i$, $j = 1, \dots, N$.

$$y_{ij} = \mu + \varepsilon_{ij}$$

↪ *effective sample size* = n_{tot}

Quelle taille d'échantillon ?

- Dans un modèle classique à **effets fixes** uniquement :

$n_{\text{tot}} = \sum_{i=1}^N n_i$ observations indépendantes :

y_{ij} , $i = 1, \dots, n_i$, $j = 1, \dots, N$.

$$y_{ij} = \mu + \varepsilon_{ij}$$

↪ *effective sample size* = n_{tot}

- Dans un modèle à **effets aléatoires** uniquement :

N observations indépendantes : (Y_1, \dots, Y_N) .

$$y_{ij} = \psi_i + \varepsilon_{ij}$$

$$\psi_i = \mu + \eta_i$$

↪ *effective sample size* = N

Quelle taille d'échantillon ?

Lien avec les propriétés asymptotiques de l'EMV :

$$y_{ij} = \psi_i + \varepsilon_{ij},$$

$$\psi_i = \mu + \eta_i,$$

où $\eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \omega^2)$, $\varepsilon_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

$$\text{Var}(\hat{\mu}_{\text{MLE}}) = \frac{\omega^2}{N} + \frac{\sigma^2}{n_{\text{tot}}}$$

si $\omega > 0$, $\hat{\mu}_{\text{MLE}}$ est \sqrt{N} -consistant,

si $\omega = 0$, $\hat{\mu}_{\text{MLE}}$ est $\sqrt{n_{\text{tot}}}$ -consistant.

Quelle pénalité du BIC ?

En pratique :

- $\text{BIC} = -2\mathcal{L} + d \log(n_{\text{tot}})$ avec R (nlme), SPSS (MIXED)
- $\text{BIC} = -2\mathcal{L} + d \log N$ avec MONOLIX, SAS (NLMIXED)

où $d = \dim(\theta)$, avec des corrections *ad-hoc*.

Que dit la théorie ?

BIC classique

- Choix de modèle bayésien : loi *a posteriori* d'un modèle \mathbf{m} paramétré par θ , $p(\mathbf{m}|Y) \propto \pi(\mathbf{m})p(Y|\mathbf{m})$,

$$p(Y|\mathbf{m}) = \int p(Y|\theta, \mathbf{m})p(\theta|\mathbf{m})d\theta.$$

BIC classique

- Choix de modèle bayésien : loi *a posteriori* d'un modèle \mathbf{m} paramétré par θ , $p(\mathbf{m}|Y) \propto \pi(\mathbf{m})p(Y|\mathbf{m})$,

$$p(Y|\mathbf{m}) = \int p(Y|\theta, \mathbf{m})p(\theta|\mathbf{m})d\theta.$$

- Approximation de Laplace :

$$\log p(Y|\mathbf{m}) \approx \log p(Y|\hat{\theta}, \mathbf{m}) - \frac{1}{2} \log \det (H_{\hat{\theta}})$$

$$\text{où } H_{\hat{\theta}} = -\frac{\partial^2 \log p(Y|\mathbf{m})}{\partial \theta \partial \theta'} \Big|_{\theta=\hat{\theta}}.$$

BIC classique

- Choix de modèle bayésien : loi *a posteriori* d'un modèle \mathbf{m} paramétré par θ , $p(\mathbf{m}|Y) \propto \pi(\mathbf{m})p(Y|\mathbf{m})$,

$$p(Y|\mathbf{m}) = \int p(Y|\theta, \mathbf{m})p(\theta|\mathbf{m})d\theta.$$

- Approximation de Laplace :

$$\log p(Y|\mathbf{m}) \approx \log p(Y|\hat{\theta}, \mathbf{m}) - \frac{1}{2} \log \det (H_{\hat{\theta}})$$

où $H_{\hat{\theta}} = -\frac{\partial^2 \log p(Y|\mathbf{m})}{\partial \theta \partial \theta'} \Big|_{\theta=\hat{\theta}}$.

- Dans un modèle à effets fixes, n obs. indépendantes,
 $H_{\hat{\theta}} \approx n\mathcal{I}_{\hat{\theta}}$ et

$$\frac{1}{2} \log \det (H_{\hat{\theta}}) \approx \frac{\dim(\theta)}{2} \log n + \frac{1}{2} \log \det (\mathcal{I}_{\hat{\theta}}),$$

BIC dans les modèles mixtes

Cadre : $N \rightarrow \infty$, $\min_i n_i \rightarrow \infty$ et $\frac{N}{\min_i n_i} \rightarrow 0$.

Partition des paramètres individuels :

$$\psi_i = C_i \beta + \eta_i, \quad \eta_i \sim_{i.i.d.} \mathcal{N}(0, \Omega)$$

où

$$\Omega = \begin{pmatrix} 0 & 0 \\ 0 & \Omega_R \end{pmatrix}$$

d'où

$$\eta_i = \begin{pmatrix} 0 \\ \eta_{R,i} \end{pmatrix}, \quad \psi_i = \begin{pmatrix} \psi_{F,i} \\ \psi_{R,i} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_F \\ \beta_R \end{pmatrix}.$$

BIC dans les modèles mixtes

En écrivant $\psi_i = \begin{pmatrix} \psi_{F,i} \\ \psi_{R,i} \end{pmatrix}$,

- $H_{\hat{\theta}} \approx$ diagonale par blocs,
- $\log \det (H_{\hat{\theta}}) \approx \log \det(N \mathcal{I}_1) + \log \det(n_{\text{tot}} \mathcal{I}_2)$.

Nie (2006, 2007), *Convergence rate of the MLE in generalized linear and nonlinear mixed-effects models*.

- dimensions de \mathcal{I}_1 et \mathcal{I}_2 ?

Partition des paramètres de population :

$$\theta = \begin{pmatrix} \theta_F \\ \theta_R \end{pmatrix}, \theta_F = \beta_F, \theta_R = (\beta_R, \Omega_R).$$

BIC dans les modèles mixtes

Résultat

Dans les modèles à effets mixtes considérés ici, la procédure BIC consiste à sélectionner le modèle qui minimise

$$BIC_h = -2\mathcal{L} + \dim(\theta_R) \log N + \dim(\theta_F) \log n_{\text{tot}}.$$

BIC dans les modèles mixtes

Résultat

Dans les modèles à effets mixtes considérés ici, la procédure BIC consiste à sélectionner le modèle qui minimise

$$BIC_h = -2\mathcal{L} + \dim(\theta_R) \log N + \dim(\theta_F) \log n_{\text{tot}}.$$

Remarques

- La sélection de variables dépend de Ω via l'ordre de grandeur des termes de l'information de Fisher des observations.
- Approximation asymptotique de l'ordre de $\mathcal{O}(N/\min_i n_i)$.

BIC dans les modèles mixtes

$$BIC_h = -2\mathcal{L}\mathcal{L} + \dim(\theta_R) \log N + \dim(\theta_F) \log n_{\text{tot}}.$$

Remarque :

BIC_h est une version hybride des BICs existants :

- Si $\dim(\theta_R) = 0$,

$$BIC_{n_{\text{tot}}} = -2\mathcal{L}\mathcal{L} + \dim(\theta) \log n_{\text{tot}}.$$

- Si $\dim(\theta_F) = 0$,

$$BIC_N = -2\mathcal{L}\mathcal{L} + \dim(\theta) \log N.$$

Simulations

Modèle linéaire

1. Observations :

$$y_{ij} = \psi_{0,i} + \psi_{1,i}t_{ij} + \psi_{2,i}t_{ij}^2 + \varepsilon_{ij}, \quad \varepsilon_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$$

2. Paramètres individuels : $\psi_i = (\psi_{i0}, \psi_{i1}, \psi_{i2})'$ tels que

$$\begin{cases} \psi_{0,i} &= \mu_0 + \eta_{0,i} \\ \psi_{1,i} &= \mu_1 + \alpha_1 c_i + \eta_{1,i} \\ \psi_{2,i} &= \mu_2 + \alpha_2 c_i + \eta_{2,i} \end{cases},$$

$$\eta_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \Omega), \quad \Omega = \text{Diag}(\omega_0^2, \omega_1^2, \omega_2^2).$$

Simulations

Modèles testés

$$\left\{ \begin{array}{l} \psi_{0,i} = \mu_0 + \eta_{0,i} \\ \psi_{1,i} = \mu_1 + \alpha_1 c_i + \eta_{1,i} \\ \psi_{2,i} = \mu_2 + \alpha_2 c_i + \eta_{2,i} \end{array} \right. , \quad \beta = \begin{pmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

4 modèles de covariables :

$$M_1 : \alpha_1 = 0 , \alpha_2 = 0,$$

$$M_2 : \alpha_1 \neq 0 , \alpha_2 = 0,$$

$$M_3 : \alpha_1 = 0 , \alpha_2 \neq 0,$$

$$M_4 : \alpha_1 \neq 0 , \alpha_2 \neq 0.$$

Simulations

Modèles testés

$$\begin{cases} \psi_{0,i} &= \mu_0 + \eta_{0,i} \\ \psi_{1,i} &= \mu_1 + \alpha_1 c_i + \eta_{1,i} \\ \psi_{2,i} &= \mu_2 + \alpha_2 c_i + \eta_{2,i} \end{cases}, \quad \Omega = \begin{pmatrix} \omega_0^2 & 0 & 0 \\ 0 & \omega_1^2 & 0 \\ 0 & 0 & \omega_2^2 \end{pmatrix}$$

4 modèles de variance :

$$O_1 : \omega_1^2 = 0, \omega_2^2 = 0,$$

$$O_2 : \omega_1^2 \neq 0, \omega_2^2 = 0,$$

$$O_3 : \omega_1^2 = 0, \omega_2^2 \neq 0,$$

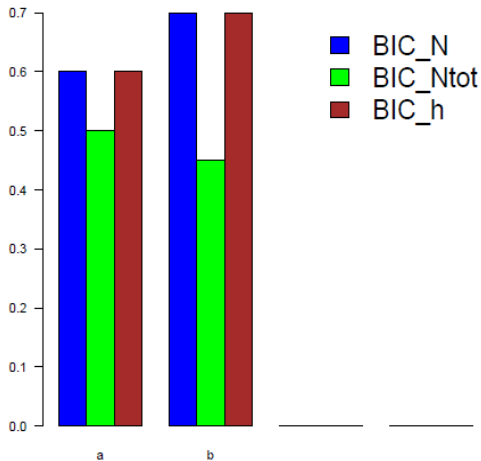
$$O_4 : \omega_1^2 \neq 0, \omega_2^2 \neq 0.$$

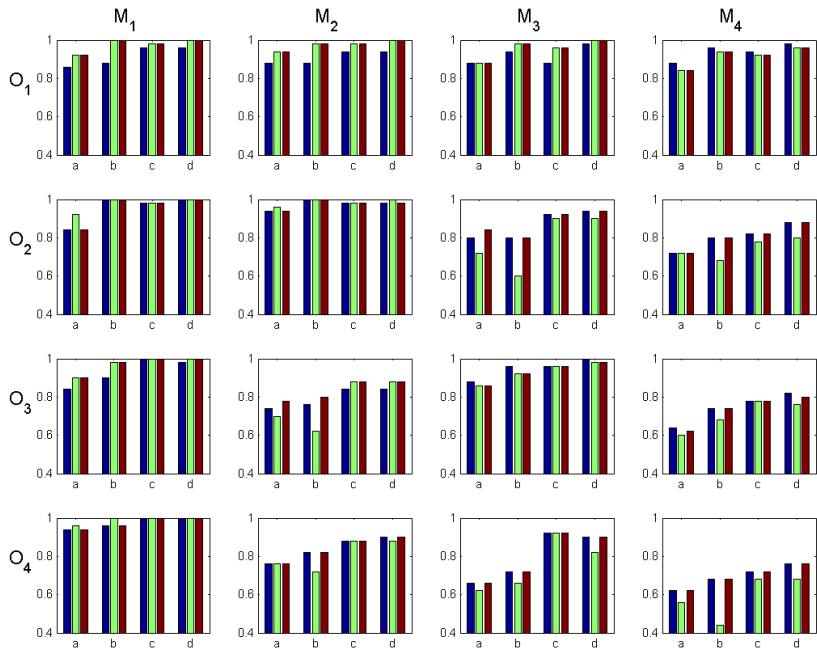
Simulations

- Pour chaque combinaison de modèles (M_k, O_k) , $1 \leq k \leq 4$, 4 design :
 - $(N = 20, n = 5)$,
 - $(N = 20, n = 100)$,
 - $(N = 100, n = 5)$,
 - $(N = 100, n = 100)$
- ↔ 64 situations possibles
- Pour chaque situation, simulation de 50 échantillons.
Pour chaque échantillon : calcul des 3 critères BIC_N (bleu), $BIC_{n_{tot}}$ (vert) et BIC_h (marron).

Exemple de lecture des résultats

Fréquence de choix de modèle correct pour chacun des critères dans 2 situations





Exemple Warfarin

$$y_{ij} = f(t_{ij}, D_i, \psi_i) + f(t_{ij}, D_i, \psi_i) \varepsilon_{ij}, \quad N = 32, \quad n_{\text{tot}} = 251$$

$$\psi_i = (k_{a,i}, V_i, Cl_i)'$$

Effet de la covariable $c_i = \text{poids}$?

$$\log V_i = \mu_V + \alpha_V c_i + \eta_{V,i}$$

$\mathcal{M}_1 : \log(ka_i) = \mu_{ka}$	$;$ $\log(Cl_i) = \mu_{Cl} + \eta_{Cl,i}$,
$\mathcal{M}_2 : \log(ka_i) = \mu_{ka}$	$;$ $\log(Cl_i) = \mu_{Cl} + \alpha_{Cl} c_i + \eta_{Cl,i}$,
$\mathcal{M}_3 : \log(ka_i) = \mu_{ka} + \alpha_{ka} c_i$	$;$ $\log(Cl_i) = \mu_{Cl} + \eta_{Cl,i}$,
$\mathcal{M}_4 : \log(ka_i) = \mu_{ka} + \alpha_{ka} c_i$	$;$ $\log(Cl_i) = \mu_{Cl} + \alpha_{Cl} c_i + \eta_{Cl,i}$,

Exemple Warfarin

$$y_{ij} = f(t_{ij}, D_i, \psi_i) + f(t_{ij}, D_i, \psi_i)\varepsilon_{ij}, \quad N = 32, \quad n_{\text{tot}} = 251$$

$$\psi_i = (k_{a,i}, V_i, Cl_i)'$$

Effet de la covariable $c_i = \text{poids}$?

Modèle	$\dim(\theta_R)$	$\dim(\theta_F)$	$-2 \mathcal{LL}$	BIC_N	$BIC_{n_{\text{tot}}}$	BIC_h
\mathcal{M}_1	5	2	928.1	952.4	966.8	956.5
\mathcal{M}_2	6	2	923.3	951.0	967.5	955.1
\mathcal{M}_3	5	3	923.0	950.7	967.2	956.9
\mathcal{M}_4	6	3	918.1	949.3	967.8	955.5

Conclusion

- Le critère proposé BIC_h lève l'ambiguïté sur le choix de la pénalité.
- BIC_h se comporte aussi bien que le meilleur des critères BIC usuels
- BIC_h s'adapte à la structure des effets aléatoires
- BIC_h implémenté dans MONOLIX
- BIC_h permet de faire de la sélection de covariables mais pas de la sélection d'effets aléatoires (éléments non nuls de Ω).

Références

- *A note on BIC in mixed-effects models*,
M. Delattre, M. Lavielle and M-A Poursat (2014), EJS 8(1), pp 456-475
- *Model Selection in Linear Mixed Models, a review*,
S. Müller, J.L. Scealy and A.H. Welsh (2013) Statistical Science
- 2 articles de *sélection conjointe* des effets fixes et aléatoires par une approche LASSO (*low-dimensional settings*) :
Bondell et al. *Biometrics* 2010
Ibrahim et al. *Biometrics* 2011
↔ choix des paramètres de régularisation : BIC_h !

Détails techniques : évaluation de $\mathcal{I}_{\hat{\theta}}$

Partition des paramètres de population

$$\mathcal{I}_{\hat{\theta}} = -\mathbb{E}_{\theta=\hat{\theta}} \begin{pmatrix} \frac{\partial^2 \log p(Y_1|\theta)}{\partial \theta_R \partial \theta'_R} & \frac{\partial^2 \log p(Y_1|\theta)}{\partial \theta_R \partial \theta'_F} \\ \frac{\partial^2 \log p(Y_1|\theta)}{\partial \theta_F \partial \theta'_R} & \frac{\partial^2 \log p(Y_1|\theta)}{\partial \theta_F \partial \theta'_F} \end{pmatrix}$$

Approximation de Laplace de $p(Y_1|\theta)$

- Rappel : $p(Y_1|\theta) = \int p(Y_1|\psi_1)p(\psi_1|\theta)d\psi_1$
- Approximation de chaque bloc de $\mathcal{I}_{\hat{\theta}}$

$$-\frac{\partial^2 \log p(Y_i|\theta)}{\partial \theta \partial \theta'} = -\frac{\partial^2 \log p(\psi_i|\theta)}{\partial \theta \partial \theta'} + \frac{\partial^2 \log p(\psi_i|\theta)}{\partial \theta \partial \psi'_i} \left(\frac{\partial^2 \log p(Y_i, \psi_i|\theta)}{\partial \psi_i \partial \psi'_i} \right)^{-1} \frac{\partial^2 \log p(\psi_i|\theta)}{\partial \theta' \partial \psi_i} \Big|_{\psi_i=\hat{\psi}_i} + \mathcal{O}(n^{-1}).$$

Hypothèses

(H1) Pour tout $i = 1, \dots, N$, et tout $\theta \in \vartheta$, $p(y_i|\theta)$ est trois fois dérivable en θ pour presque tout Y_i .

(H2) (i) Il existe $M > 0$ t.q. $\forall i = 1, \dots, N, \forall \theta \in \vartheta$ et $\forall k, l = 1, \dots, q$,

$$\mathbb{E}_{Y_i|\theta^*} \left[\frac{\partial \log p(Y_i|\theta)}{\partial \theta_k} \Big|_{\theta=\theta^*} \right]^2 < M \text{ et } \mathbb{E}_{Y_i|\theta^*} \left[\frac{\partial^2 \log p(Y_i|\theta)}{\partial \theta_k \partial \theta_l} \Big|_{\theta=\theta^*} \right]^2 < M.$$

(ii) Il existe $\{G_1(Y_1), \dots, G_N(Y_N)\}$ t.q. $\theta \in \vartheta, \forall i = 1, \dots, N$ et $\forall k, l, h = 1, \dots, q$,

$$\left| \frac{\partial^3 \log p(Y_i|\theta)}{\partial \theta_k \partial \theta_l \partial \theta_h} \right| \leq G_i(Y_i) \text{ et } \mathbb{E}_{Y_i|\theta^*} [G_i^2(Y_i)] \leq M.$$

(H3) Soit $V_\theta = (H_{\theta^*}^{-\frac{1}{2}})H_\theta(H_{\theta^*}^{-\frac{1}{2}})' \forall \theta \in \vartheta$. H_{θ^*} est définie positive et inversible, et

$$\max_{\theta \in \vartheta} \|V_\theta - I_q\| \xrightarrow{N \rightarrow +\infty} 0,$$

où I_q : matrice identité $q \times q$.

(H4) $\liminf_{N \rightarrow +\infty} \lambda_N = \lambda > 0$ où λ_N est la plus petite valeur propre de $\mathbb{E}(\frac{1}{N}H_{\theta^*})$.

Hypothèses

(H5) (i)

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E}_{Y_i|\theta^*} \left\{ \left[\psi_{R,i}' \left(-\frac{\partial^2 \log p(y_i|\psi_{R,i}, \theta_F)}{\partial \psi_{R,i} \partial \psi_{R,i}'} + \Omega_R^{-1} \right)^{-1} \psi_{R,i} \right]_{|\psi_i = \hat{\psi}_i} \right\} = o(n^{-1}),$$

(ii)

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E}_{Y_i|\theta^*} \left[\frac{\partial^2 \log p(Y_i|\psi_{R,i}, \theta_F)}{\partial \theta_F \partial \psi_{R,i}'} \left(-\frac{\partial^2 \log p(Y_i|\psi_{R,i}, \theta_F)}{\partial \psi_{R,i} \partial \psi_{R,i}'} + \Omega_R^{-1} \right)^{-1} \psi_{R,i} \right]_{|\psi_i = \hat{\psi}_i} = \mathcal{O}(1)$$

