# Master MIASHS
# Modèles Linéaires Mixtes

Christian Lavergne et Catherine Trottier

Année universitaire 2016-2017

## Le modèle linéaire gaussien à structure de covariance paramétrée et séparable ($\beta \perp \theta$)

- **Définition** $Y = X\beta + \epsilon$, avec $X$ de dimension $n \times p$

  $\epsilon \sim \mathcal{N}_{\mathbb{R}^N}(0, \Gamma_\theta)$ avec $\theta$ paramètre inconnu de $\mathbb{R}^K$.

- **La vraisemblance**

$$f(\beta, \theta; y) = \frac{1}{(2\pi)^{N/2}} |\Gamma_\theta|^{-1/2} \exp\{-\frac{1}{2}(y - X\beta)'\Gamma_\theta^{-1}(y - X\beta)\}$$

$$\text{et } l(\beta, \theta) = -2 \log(f(\beta, \theta; y))$$

- **Cas particulier : le modèle linéaire élémentaire.** Repose sur l'hypothèse d'erreur gaussienne et de structure de covariance $\sigma_\epsilon^2 \Gamma$ avec $\Gamma$ connue. Le cas le plus répandu étant $\Gamma = \mathrm{Id}_n$.

## Le maximum de vraisemblance dans un modèle linéaire général

- **Estimateur du maximum de vraisemblance : ML**

$$\hat{\beta}_{ML} = \arg\min_{\beta}((y - X\beta)'\Gamma_{\theta}^{-1}(y - X\beta))$$

$$\hat{\theta}_{ML} = \arg\min_{\theta}((y - X\beta)'\Gamma_{\theta}^{-1}(y - X\beta) + \log(|\Gamma_{\theta}|))$$

- **Fonction score**

$$\begin{aligned} U_{\beta}(\hat{\theta}_{ML}) &= \frac{\partial l(\beta,\theta)}{\partial\beta}\Big|_{\theta=\hat{\theta}_{ML}} \\ &= -2\,X'\Gamma_{\theta}^{-1}(y - X\beta)\Big|_{\theta=\hat{\theta}_{ML}} \qquad (1) \end{aligned}$$

donc $\hat{\beta}_{ML} = (X'\Gamma_{\hat{\theta}_{ML}}^{-1}X)^{-1}X'\Gamma_{\hat{\theta}_{ML}}^{-1}y.$

$$
\begin{aligned}
U_\theta^{ML}(\hat\beta_{ML}) &= \frac{\partial l(\beta,\theta)}{\partial\theta}\Big|_{\beta=\hat\beta_{ML}} \\
&= -(y-X\beta)'\Gamma_\theta^{-1}\frac{\partial\Gamma_\theta}{\partial\theta}\Gamma_\theta^{-1}(y-X\beta)+\underline{\mathrm{tr}}(\Gamma_\theta^{-1}\frac{\partial\Gamma_\theta}{\partial\theta})\Big|_{\beta=\hat\beta_{ML}} \\
&= -y'P_\theta\frac{\partial\Gamma_\theta}{\partial\theta}P_\theta\, y+\underline{\mathrm{tr}}(\Gamma_\theta^{-1}\frac{\partial\Gamma_\theta}{\partial\theta}) \qquad (2)
\end{aligned}
$$

où

- $P_\theta = \Gamma_\theta^{-1}(\mathrm{Id} - X(X'\Gamma_\theta^{-1}X)^{-1}X'\Gamma_\theta^{-1})$,
- $u'A_\theta\frac{\partial\Gamma_\theta}{\partial\theta}A_\theta\, u =$ est le vecteur de composantes $[u'A_\theta\frac{\partial\Gamma_\theta}{\partial\theta_j}A_\theta\, u]_{j=1,\dots,K}$
- resp. $\underline{\mathrm{tr}}(A_\theta\frac{\partial\Gamma_\theta}{\partial\theta}) = [\mathrm{tr}(A_\theta\frac{\partial\Gamma_\theta}{\partial\theta_j})]_{j=1,\dots,K}$

rappel : $\frac{\partial\Gamma_\theta^{-1}}{\partial\theta_j} = -\Gamma_\theta^{-1}\frac{\partial\Gamma_\theta}{\partial\theta_j}\Gamma_\theta^{-1}$ et $\frac{\partial\log(|\Gamma_\theta|)}{\partial\theta_j} = \mathrm{tr}(\Gamma_\theta^{-1}\frac{\partial\Gamma_\theta}{\partial\theta_j})$

**Estimation du paramètre de dispersion $\sigma^2$ : cas $\Gamma_\theta = \sigma^2 \mathrm{Id}$**

$\widehat{\sigma^2} = \dfrac{1}{n-p} \|Y - X\widehat{\beta}\|^2$ est un estimateur sans biais de $\sigma^2$ .

$\|Y - X\widehat{\beta}\|^2 = \|Y - X({}^tXX)^{-1}{}^tXY\|^2$
$= \|[Id - X({}^tXX)^{-1}{}^tX]Y\|^2 = \|MY\|^2 = {}^tY{}^tMMY$

Or $M$ projecteur orthogonal sur $\mathcal{X}^\perp$ donc : ${}^tM = M$; $M.M = M$ et $MX = 0$.
Donc

$$\|Y - X\widehat{\beta}\|^2 = {}^tYMY = {}^t(X\beta + \epsilon)M(X\beta + \epsilon) = {}^t\epsilon M\epsilon.$$

D'après le lemme :

$$E(\|Y - X\widehat{\beta}\|^2) = E({}^t\epsilon M\epsilon) = \sigma^2 \mathrm{trace} M = \sigma^2(n-p) \bullet$$

## Estimation dans la cas Var$(\epsilon) = \sigma^2\Gamma$

- $\Gamma$ est une matrice symétrique, définie positive donc diagonalisable et à valeurs propres $> 0$.

$$\Gamma = {}^tU\Lambda U \quad \text{avec} \quad {}^tUU = Id$$

Posons $\sqrt{\Lambda}$ la matrice diagonale des racines carrés des valeurs propres de $\Gamma$ et le changement de variable $Z = HY$ avec $H = \sqrt{\Lambda}^{-1}U$ alors : Var$(Z) = \sigma^2$Id et on applique au modèle $Z = HX\beta + \epsilon_z$ les résultats précédents.

## Interprétation géométrique ; propriété de W. Kruskal 1968

$$\widehat{\beta} = (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}Y = (X'X)^{-1}X'Y$$

l'espace $\mathcal{X}$ est invariant par l'opérateur $\Gamma$ ($\Gamma X = X$)

$$\widehat{\sigma}^2 = \frac{1}{n}||Y - X\widehat{\beta}_{ML}||^2_{\Gamma^{-1}} = \frac{1}{n}||Y - X\widehat{\beta}_{\mathrm{Id}}||^2$$

les espaces $\mathcal{X}$ et $\mathcal{X}^\perp$ sont invariants par l'opérateur $\Gamma$

**Théorème de Kruskal, démonstration**

Les deux estimateurs $\widehat{\beta}_{\mathrm{Id}}$ et $\widehat{\beta}_{\Gamma^{-1}}$ sont identiques si et seulement si $\mathcal{X}$ est invariant par la matrice $\Gamma$ .

$\widehat{\beta}_{\mathrm{Id}}$ est définie par : $\quad < y - X\widehat{\beta}_{\mathrm{Id}}, z >_{\mathrm{Id}} = 0, \ \forall z \in \mathcal{X};$

$\widehat{\beta}_{\Gamma^{-1}}$ est définie par : $\quad < y - X\widehat{\beta}_{\Gamma^{-1}}, z >_{\Gamma^{-1}} = < y - X\widehat{\beta}_{\Gamma^{-1}}, \Gamma^{-1} z >_{\mathrm{Id}} = 0.$

De plus $\mathcal{X}$ invariant par la matrice $\Gamma$ (ou $\Gamma$ inv.)

$$\Leftrightarrow \Gamma\mathcal{X} = \mathcal{X} \Leftrightarrow \Gamma^{-1}\mathcal{X} = \mathcal{X} \Leftrightarrow \Gamma^{-1}z \in \mathcal{X}, \ \forall z \in \mathcal{X}.$$

a) $\widehat{\beta}_{\mathrm{Id}} = \widehat{\beta}_{\Gamma^{-1}} \Rightarrow < y - X\widehat{\beta}_{\mathrm{Id}}, \Gamma^{-1} z >_{\mathrm{Id}} = 0$, donc $\Gamma^{-1} z \in \mathcal{X}, \ \forall z \in \mathcal{X}$.

b) $\mathcal{X}$ est invariant par $\Gamma$ (ou $\Gamma^{-1}$) alors la condition sur $\widehat{\beta}_{\Gamma^{-1}}$ :
$< y - X\widehat{\beta}_{\Gamma^{-1}}, \Gamma^{-1} z >_{\mathrm{Id}} = 0$ se transforme en : $< y - X\widehat{\beta}_{\Gamma^{-1}}, z >_{\mathrm{Id}} = 0$.

De par l'unicité $\widehat{\beta}_{\mathrm{Id}} = \widehat{\beta}_{\Gamma^{-1}}$.

**Matrice d'information :**
Posons $\gamma = (\beta, \theta)$ et

$$\mathcal{I}_{\beta,\theta} = E_{\beta,\theta}\left[-\frac{\partial^2 \log f(\beta,\theta;y)}{\partial\gamma\partial\gamma'}\right]$$

alors

$$\mathcal{I}_{\beta,\theta} = E_{\beta,\theta}\left[\frac{1}{2}\frac{\partial^2 l(\beta,\theta)}{\partial\gamma\partial\gamma'}\right] = E_{\beta,\theta}\left[\frac{1}{2}\begin{pmatrix} \dfrac{\partial U_\beta}{\partial\beta'} & \dfrac{\partial U_\beta}{\partial\theta'} \\ \\ \dfrac{\partial U_\theta^{ML}}{\partial\beta'} & \dfrac{\partial U_\theta^{ML}}{\partial\theta'} \end{pmatrix}\right].$$

On vérifie aisément que $\dfrac{\partial U_\beta}{\partial\beta'} = 2\,X'\Gamma_\theta^{-1}X$ et que $E_{\beta,\theta}\left[\dfrac{\partial U_\beta}{\partial\theta'}\right] = 0.$

D'autre part :

$$\left[\frac{\partial U_\theta^{ML}}{\partial \theta'}\right]_{ij} = \mathrm{tr}(\Gamma_\theta^{-1}(\frac{\partial^2 \Gamma_\theta}{\partial\theta_i \partial\theta_j} - \frac{\partial \Gamma_\theta}{\partial\theta_i}\Gamma_\theta^{-1}\frac{\partial \Gamma_\theta}{\partial\theta_j})) \\ -(y - X\beta)'\Gamma_\theta^{-1}(\frac{\partial^2 \Gamma_\theta}{\partial\theta_i \partial\theta_j} - 2\,\frac{\partial \Gamma_\theta}{\partial\theta_i}\Gamma_\theta^{-1}\frac{\partial \Gamma_\theta}{\partial\theta_j})\Gamma_\theta^{-1}(y - X\beta)$$

Donc

$$\mathcal{I}_{\beta,\theta} = \begin{pmatrix} \mathcal{I}_\beta = X'\Gamma_\theta^{-1}X & 0 \\ 0 & \mathcal{I}_\theta^{ML} = \left[\frac{1}{2}\mathrm{tr}(\Gamma_\theta^{-1}\frac{\partial \Gamma_\theta}{\partial\theta_i}\Gamma_\theta^{-1}\frac{\partial \Gamma_\theta}{\partial\theta_j})\right]_{i,j=1,\ldots,K} \end{pmatrix}$$

# Le maximum de vraisemblance restreint : REML

- **La vraisemblance restreinte :** c'est la vraisemblance marginale après intégration sur $\beta$

$$f_{RE}(\theta; y) = \frac{1}{(2\pi)^{(N-p)/2}} |\Gamma_\theta|^{-1/2} |X'\Gamma_\theta^{-1}X|^{-1/2} \exp\{-\frac{1}{2} y' P_\theta \ y\}$$

et

$$l_{RE}(\theta) = -2 \log(f_{RE}(\theta; y))$$

- **Estimateur maximun de vraisemblance restreinte**

$$\hat{\theta}_{REML} = \arg\min_\theta (y' P_\theta \ y + \log(|\Gamma_\theta|) + \log(|X'\Gamma_\theta^{-1}X|)$$

## Fonction score

$$
\begin{aligned}
U_\theta^{REML} &= \frac{\partial l_{RE}(\theta)}{\partial \theta} \\
&= -y' P_\theta \frac{\partial \Gamma_\theta}{\partial \theta} P_\theta\, y + \underline{\mathrm{tr}}(\Gamma_\theta^{-1} \frac{\partial \Gamma_\theta}{\partial \theta}) \\
&\quad -\underline{\mathrm{tr}}((X'\Gamma_\theta^{-1}X)^{-1} X'\Gamma_\theta^{-1} \frac{\partial \Gamma_\theta}{\partial \theta} \Gamma_\theta^{-1} X) \\
&= -y' P_\theta \frac{\partial \Gamma_\theta}{\partial \theta} P_\theta\, y + \underline{\mathrm{tr}}(P_\theta \frac{\partial \Gamma_\theta}{\partial \theta}) \qquad (3)
\end{aligned}
$$

et $\hat{\beta}_{REML}$ est solution de l'équation $U_\beta(\hat{\theta}_{REML}) = 0$.

# Matrice d'information associée au vecteur $y$

Posons

$$\mathcal{I}_\theta^{REML} = E_\theta\left[-\frac{\partial^2 \log f_{RE}(\theta; y)}{\partial\theta\partial\theta'}\right]$$

alors

$$\mathcal{I}_\theta^{REML} = E_\theta\left[\frac{1}{2}\frac{\partial^2 l_{RE}(\theta)}{\partial\theta\partial\theta'}\right] = E_\theta\left[\frac{1}{2}\frac{\partial U_\theta^{REML}}{\partial\theta'}\right].$$

Un calcul similaire au cas précédent

$$\mathcal{I}_\theta^{REML} = \left[\frac{1}{2}\mathrm{tr}(P_\theta\frac{\partial\Gamma_\theta}{\partial\theta_i}P_\theta\frac{\partial\Gamma_\theta}{\partial\theta_j})\right]_{i,j=1,\ldots,K}$$

# Remarque :

- En résumé :
  $\hat{\theta}_{ML}$ est donc solution de la minimisation du critère :
  $$\text{CritèreML} = y' P_\theta \; y + \log(|\Gamma_\theta|) \, ;$$
  $\hat{\theta}_{REML}$ est solution de la minimisation du critère :
  $$\text{CritèreREML} = \text{CritèreML} + \log(|X' \Gamma_\theta^{-1} X|)$$
  et $\hat{\beta}_{.} = (X' \Gamma_{\hat{\theta}_.}^{-1} X)^{-1} X' \Gamma_{\hat{\theta}_.}^{-1} y.$

- Remarquant que $E_\theta(P_\theta \; y) = 0$ et $P_\theta \Gamma_\theta P_\theta = P_\theta$ alors
  $$E_\theta(y' P_\theta \frac{\partial \Gamma_\theta}{\partial \theta} P_\theta \; y) = \underline{\text{tr}}(P_\theta \frac{\partial \Gamma_\theta}{\partial \theta}).$$
  La fonction score $U_\theta^{REML}$ est donc une statistique centrée contrairement à la fonction score $U_\theta^{ML}$.

# Critères de choix de modèles.

Les critères d'informations que l'on cherche à minimiser sont construits sur la log-vraisemblance en $\widehat{\beta}_{ML}$ et $\widehat{\theta}_{ML}$ notée $\mathcal{L}(\mathcal{M})$.
On désigne par $q(\mathcal{M})$ le nombre de paramètres estimés dans le modèle $(q(\mathcal{M}) = p + K)$.

AIC : Akaike Information Criterium

$$-2\mathcal{L}(\mathcal{M}) + 2 * q(\mathcal{M})$$

BIC : Bayesian Information Criterium

$$-2\mathcal{L}(\mathcal{M}) + \ln(n) * q(\mathcal{M})$$

## Le modèle linéaire mixte : L2M

$$Y = X\beta + \mathbb{Z}U + \epsilon \;=\; X\beta + \sum_{k=1}^{K} Z_k U_k + \epsilon$$

- $\epsilon \sim \mathcal{N}_{I\!R^N}(0, R = \theta_0 V_0)$, $V_0$ est une matrice connue.
- $U_k \sim \mathcal{N}_{I\!R^{q_k}}(0, \theta_k G_k)$ pour tout $k = 1, \ldots, K$ ; les $G_k$ sont des matrices connues.

$U_1, U_2, \ldots, U_k$ sont les effets aléatoires non observés ; ils sont indépendants entre eux et indépendants de $\epsilon$.

- $U \sim \mathcal{N}_{I\!R^q}(0, G_\theta)$ où $q = \sum_{k=1}^{K} q_k$

$$\mathrm{Var}(Y) = \Gamma_\theta = R + \mathbb{Z} G_\theta \mathbb{Z}' = \sum_{k=0}^{K} \theta_k V_k \text{ où } V_k = Z_k G_k Z_k'$$

les $\theta_k$ ($\in I\!R^+$) sont appelés les composantes de la variance.

# Simplification des fonctions scores $U_\theta^{ML}$ et $U_\theta^{REML}$ du modèle mixte

Dans ce cas particulier on a : $\frac{\partial \Gamma_\theta}{\partial \theta_k} = V_k$ :

$$\begin{array}{rcrcl} U_\theta^{ML} & : & y' P_\theta V_k P_\theta \, y & = & \mathrm{tr}(\Gamma_\theta^{-1} V_k) \text{ pour } k = 0, \ldots, K \\ U_\theta^{REML} & : & y' P_\theta V_k P_\theta \, y & = & \mathrm{tr}(P_\theta V_k) \text{ pour } k = 0, \ldots, K \end{array}$$

Comme pour $k = 0, \ldots, K$

$$\begin{array}{rrcl} & \mathrm{tr}(\Gamma_\theta^{-1} V_k) & = & \sum_{j=1}^{K} \mathrm{tr}(\Gamma_\theta^{-1} V_k \Gamma_\theta^{-1} V_j) \theta_j \\ \text{et} & \mathrm{tr}(P_\theta V_k) & = & \sum_{j=1}^{K} \mathrm{tr}(P_\theta V_k P_\theta V_j) \theta_j \end{array}$$

On obtient ainsi 2 systèmes linéaires identiques :

$2\mathcal{I}_\theta \, \theta = F_\theta$ où $F_\theta$ est le vecteur $\{y' P_\theta V_k P_\theta \, y\}$

# L'algorithme EM

Principe : étant donné un vecteur aléatoire observé $Y$ et un vecteur aléatoire non observé (ici les effets aléatoires) $U$, on cherche des statistiques exhaustives, fonction des données complètes $x = (y', u')'$ qui permettent de réaliser l'estimation des paramètres inconnus $\gamma$ : notées $t(x)$. On obtient un algorithme qui à chaque itération $[t]$ pour une valeur courante du paramètre $\gamma^{[t]}$ se décompose en 2 étapes :

• étape E : calculer l'espérance conditionnelle de $t(x)$ sachant les données observées $y$ et la valeur $\gamma^{[t]}$.

• étape M : maximiser la vraisemblance des données complètes en remplaçant les statistiques exhaustives par l'espérance conditionnelle et obtenir ainsi $\gamma^{[t+1]}$.

## Le modèle linéaire mixte : L2M

$$Y = X\beta + \sum_{k=1}^{K} Z_k U_k + \epsilon \ ; \epsilon \sim \mathcal{N}_{\mathbb{R}^N}(0, \theta_0 V_0), \ U_k \sim \mathcal{N}_{\mathbb{R}^{q_k}}(0, \theta_k G_k).$$

Donc si les $U_k$ étaient observées, les estimations des paramètres seraient obtenues par :

$$\widehat{\theta}_k = \frac{U_k G_k^{-1} U_k'}{q_k} \text{ et } \widehat{\beta} = (X' V_0^{-1} X)^{-1} X' V_0^{-1} (y - \sum_{k=1}^{K} Z_k U_k)$$

On vérifie alors aisément que :

$$E(U_k G_k^{-1} U_k' | y) = \theta_k^2 (y - X\beta)' \Gamma_\theta^{-1} V_j \Gamma_\theta^{-1} (y - X\beta) + q_k \theta - \theta^2 \mathrm{tr}(\Gamma_\theta^{-1} V_j)$$

$$E(y - \sum_{k=1}^{K} Z_k U_k | y) = X\beta + \theta_0 V_0 \Gamma_\theta^{-1} (y - X\beta)$$

D'où l'algorithme EM pour ML :

$$\begin{cases} q_k\theta_k^{[t+1]} & = & \theta_k^{2[t]}(y - X\beta^{[t]})'\Gamma_{\theta^{[t]}}^{-1}V_j\Gamma_{\theta^{[t]}}^{-1}(y - X\beta^{[t]}) \\ & & +q_k\theta^{[t]} - \theta^{2[t]}\mathrm{tr}(\Gamma_{\theta^{[t]}}^{-1}V_j) \\ X\beta^{[t+1]} & = & (X'V_0^{-1}X)^{-1}X'V_0^{-1}[X\beta^{[t]} + \theta_0^{[t]}V_0\Gamma_{\theta^{[t]}}^{-1}(y - X\beta^{[t]})] \end{cases}$$

La $1^{re}$ partie de l'algorithme peut aussi s'écrire :

$$\begin{cases} q_k\theta_k^{[t+1]} & = & \theta_k^{2[t]}y'P_{\theta^{[t]}}V_jP_{\theta^{[t]}}y + q_k\theta^{[t]} - \theta^{2[t]}\mathrm{tr}(\Gamma_{\theta^{[t]}}^{-1}V_j) \end{cases}$$

et en remplaçant $\Gamma_{\theta^{[t]}}^{-1}$ par $P_{\theta^{[t]}}$ on obtient l'algorithme EM pour REML.