# Modèles Linéaires Mixtes

**Christian Lavergne/ Catherine Trottier**

**Christian.Lavergne@math.univ-montp2.fr**
**Catherine.Trottier@math.univ-montp2.fr**

web :

**http://www.univ-montp3.fr/miap/ $\sim$ lavergne**

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

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

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

**La vraisemblance**

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

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

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

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

### Fonction score

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

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

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

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 $\Gamma_\theta = \sigma^2 \mathbf{Id}$**

$\widehat{\sigma^2} = \dfrac{1}{n-p} \|Y - X\widehat{\beta}\|^2$ 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 :

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

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

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

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

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 : $\mathrm{Var}(Z) = \sigma^2\mathrm{Id}$ et on applique au modèle $Z = HX\beta + \epsilon_z$ les résultats précédents.

## Interprétation géométrique ; propriété de W. Kruskal 1968

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

$$\Updownarrow$$

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

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

$$\Updownarrow$$

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

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

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 :

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

Donc

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

## 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**

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

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

## 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 :

$\widehat{\theta}_{ML}$ est donc solution de la minimisation du critère :

$$\text{CritèreML} = y' P_\theta\, y + \log(|\Gamma_\theta|)\ ;$$

$\widehat{\theta}_{REML}$ est solution de la minimisation du critère :

$$\text{CritèreREML} = \text{CritèreML} + \log(|X'\Gamma_\theta^{-1} X|)$$

et $\widehat{\beta}_{.} = (X'\Gamma_{\widehat{\theta}_{.}}^{-1} X)^{-1} X'\Gamma_{\widehat{\theta}_{.}}^{-1} y.$

- 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 $\widehat{\beta}$ et $\widehat{\theta}^2_{ML}$ 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

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