

An impressionistic landscape painting featuring a bright yellow field in the foreground, a blue river or path winding through the middle ground, and a large, dark green, spiky plant on the right. The background is a mix of light blue and white, suggesting a sky or distant hills. The overall style is expressive and textured, with visible brushstrokes.

MODELE LINEAIRE A EFFETS MIXTES THEORIE & APPLICATION

Amin EL GAREH et Bezeid
CHEICK-MOHAMED-LMAMI

PROJET JANVIER-MARS 2015, UNIVERSITÉ DE BOURGOGNE

Ce projet a été encadré par Hervé CARDOT.

Publié le 30 Mars 2015



Table des Matières

1	Introduction	5
1.1	Motivation	5
1.2	Modèle linéaire mixte gaussien à K facteurs aléatoires	5
2	Méthode ML	8
2.1	Construction des estimateurs	8
2.1.1	Estimation du coefficient β	9
2.1.2	Estimation des variances α_k	10
2.1.3	Fisher scoring pour l'estimation de θ	15
2.2	Etude de la qualité des estimateurs	18
2.2.1	Etude de la qualité de $\hat{\beta}$	18
2.2.2	Etude de la qualité de \hat{V}	22
3	Méthode REML	23
4	Tests d'hypothèses et critère de selection	25
4.1	Test de Wald	25
4.2	Test du ratio Log-Vraisemblance	27
4.3	Critère d'information AIC	29
4.4	Critère d'information BIC	29

5	Application	30
5.1	Introduction du modèle linéaire mixte pour l'étude de la dépendance alimentaire des oisillons	30
5.1.1	Présentation des données	30
5.1.2	Présentation du modèle	31
5.1.3	Description des données	33
5.2	Illustration de la régression linéaire mixte avec la méthode ML	35
5.2.1	Sélection des variables	35
5.2.2	Interprétation des résultats	39
5.3	Illustration de la régression linéaire mixte avec la méthode REML	43



1. Introduction

1.1 Motivation

Le modèle linéaire mixte a été mis en oeuvre dès les années 1950, essentiellement dans le domaine de la génétique animale (réf. Henderson^[5]). Il n'a toutefois connu une utilisation plus générale qu'au cours des années 1990, en relation avec le développement de nouvelles procédures de calcul dans le cadre des logiciels statistiques. L'utilisation du modèle linéaire mixte soulève, par rapport aux modèles classiques d'analyse de la variance, un certain nombre de difficultés particulières, tant en ce qui concerne l'estimation des différents paramètres que la réalisation des tests d'hypothèses. Des informations peuvent être trouvées à ce sujet dans les articles de Littell [2002], McLean et al. [1991], et Piepho et al. [2003], et dans les livres de Demidenko [2004], McCulloch et Searle [2001],

1.2 Modèle linéaire mixte gaussien à K facteurs aléatoires

On appelle modèle mixte un modèle statistique dans lequel on considère à la fois des facteurs à effets fixes (qui vont intervenir au niveau de la moyenne du modèle) et des facteurs à effets aléatoires (qui vont intervenir au niveau de la variance du modèle). Un modèle est dit mixte lorsqu'il y a au moins un facteur de chaque nature. Dans le cadre de ce rapport, nous ne considérons que des modèles linéaires gaussiens mixtes unidimensionnel à K facteurs aléatoires indépendants plus une résiduelle, mais la notion de modèle mixte se rencontre également dans d'autres contextes, notamment dans le modèle linéaire généralisé.

Un modèle linéaire à effets mixtes est un modèle (réf. Laird et Ware^[2]) qui satisfait:

$$\boxed{Y = X\beta + \sum_{k=1}^K Z_k \gamma_k + \varepsilon} \quad (1.1)$$

avec $(\gamma_k)_{k=1,\dots,K}$ le k ème vecteur aléatoire et ε le vecteur des résidus.

• $\gamma_k = \begin{bmatrix} \gamma_{k1} \\ \vdots \\ \gamma_{kq_k} \end{bmatrix}_{q_k,1}$ où $\gamma_{k1}, \dots, \gamma_{kq_k} \hookrightarrow \mathcal{N}(0, \sigma_k^2)$ et indépendantes les unes des autres.

Le vecteur aléatoire γ_k est gaussien puisque pour tout $c = (c_1, \dots, c_{q_k})' \in \mathbb{R}^{q_k}$ la variable $\sum_{i=1}^{q_k} c_i \gamma_{ki} = c' \gamma_k$ est une variable réelle gaussienne. γ_k est donc normalement distribué, $\gamma_k \hookrightarrow \mathcal{N}(0, \Sigma_k)$ de matrice de covariance $\Sigma_k = \sigma_k^2 I_{q_k}$

• $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n,1}$ où $\varepsilon_1, \dots, \varepsilon_n \hookrightarrow \mathcal{N}(0, \sigma_\varepsilon^2)$ et indépendantes les unes des autres.

• $X = \begin{bmatrix} | & | & & | \\ 1 & x_1 & \dots & x_p \\ | & | & & | \end{bmatrix}_{n,p+1}$ $\beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{p+1,1}$

$X \in \mathcal{M}_{n,p+1}(\mathbb{R})$ est la matrice formée d'une colonne 1_n et de variables explicatives fixées x_j .

β est le vecteur de \mathbb{R}^{p+1} qui réunit la constante de régression α et les coefficients β_j des variables à effets fixes. Tandis que les matrices Z_k dans $\mathcal{M}_{n,q_k}(\mathbb{R})$ et les vecteurs γ_k de \mathbb{R}^{q_k} jouent un rôle pour les composantes aléatoires du modèle.

Proposition 1.2.1 — Distribution du vecteur à expliquer. Y est normalement distribué de moyenne $\mu = X\beta$ et de matrice de covariance $V = \sum_{k=1}^K Z_k \sigma_k^2 Z_k' + \sigma_\varepsilon^2 I_n$ supposée symétrique, définie positive.

Démonstration.

Le vecteur $Y = (y_1, \dots, y_n)'$ est un vecteur gaussien de \mathbb{R}^n puisque pour tout $c = (c_1, \dots, c_n)' \in \mathbb{R}^n$ la variable $\sum_{i=1}^n c_i y_i = c'Y = c'(X\beta + \sum_{k=1}^K Z_k \gamma_k + \varepsilon) = c'X\beta + c'\sum_{k=1}^K Z_k \gamma_k + c'\varepsilon$ est une variable réelle gaussienne, car $c'X\beta$ est un réel et $c'\sum_{k=1}^K Z_k \gamma_k + c'\varepsilon$ est une somme de variables réelles gaussiennes.

Y est de moyenne μ et de matrice de covariance V tel que:

$$\mu = \mathbb{E}[Y] = \mathbb{E}[X\beta] + \sum_{k=1}^K \mathbb{E}[Z_k \gamma_k] + \mathbb{E}[\varepsilon] = X\beta \quad (\text{car } X \text{ est déterministe})$$

$$\begin{aligned} V &= \text{Var}(Y) \\ &= \text{Var}(X\beta) + \sum_{k=1}^K \text{Var}(Z_k \gamma_k) + \text{Var}(\varepsilon) \\ &= \sum_{k=1}^K Z_k \text{Var}(\gamma_k) Z_k' + \sigma_\varepsilon^2 I_n \\ &= \sum_{k=1}^K Z_k \sigma_k^2 Z_k' + \sigma_\varepsilon^2 I_n \quad \blacksquare \end{aligned}$$



2. Méthode ML

L'idée fondamentale de l'estimation par maximum de vraisemblance dite ML ("Maximum Likelihood") est, comme le nom l'implique, de trouver un ensemble d'estimations de paramètres, appelé $\hat{\theta}$, telles que la vraisemblance d'avoir obtenu l'échantillon que nous utilisons soit maximisée. Nous signifions par là que la densité de probabilité jointe pour le modèle que l'on estime est évaluée aux valeurs observées de la (des) variable(s) dépendante(s) et traitée comme une fonction de paramètres du modèle. Le vecteur $\hat{\theta}$ des estimations ML donne alors le maximum de cette fonction. Ce principe d'estimation est très largement applicable: si nous pouvons écrire la densité jointe de l'échantillon, nous pouvons en principe utiliser le maximum de vraisemblance, soumis bien sûr à certaines conditions de régularité.

2.1 Construction des estimateurs

On rappelle que le cadre général d'étude est:

$$Y = X\beta + \sum_{k=1}^K Z_k \gamma_k + \varepsilon$$

avec $Y \hookrightarrow \mathcal{N}(X\beta, V)$ de matrice de covariance $V = \sum_{k=1}^K \sigma_k^2 Z_k Z_k' + \sigma_\varepsilon^2 I_n$

Si Z_k est de plein rang alors V est inversible, de ce fait Y admet une densité par rapport à la mesure de Lebesgue λ_n , la fonction $f_Y : \mathbb{R}^n \rightarrow \mathbb{R}_+^*$ est définie par:

$$\forall Y \in \mathbb{R}^n, f_Y(Y) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(V)}} \exp\left(-\frac{1}{2}(Y - X\beta)' V^{-1} (Y - X\beta)\right)$$

On note $\alpha = (\sigma_1^2, \dots, \sigma_K^2, \sigma_\varepsilon^2)$.

La vraisemblance L est considéré ici comme une fonction dépendante du paramètre $\theta = (\beta', \alpha)$.

$$L(Y, \theta) = f_Y(Y) = (2\pi)^{-\frac{n}{2}} \det(V)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(Y - X\beta)' V^{-1} (Y - X\beta)\right)$$

Le logarithme de la vraisemblance L , dite logvraisemblance est donné par:

$$\begin{aligned} \ln(L(Y, \theta)) &= \ln(f_Y(Y)) \\ &= \ln\left((2\pi)^{-\frac{n}{2}} \det(V)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(Y - X\beta)' V^{-1} (Y - X\beta)\right)\right) \\ &= \ln((2\pi)^{-\frac{n}{2}}) + \ln(\det(V)^{-\frac{1}{2}}) - \frac{1}{2}(Y - X\beta)' V^{-1} (Y - X\beta) \end{aligned}$$

$$\ln(L(Y, \theta)) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln(\det(V)) - \frac{1}{2}(Y - X\beta)' V^{-1} (Y - X\beta) \quad (2.1)$$

2.1.1 Estimation du coefficient β

On s'intéresse au terme de (2.1) ne faisant intervenir que le coefficient β , et on développe l'expression ainsi retenue

$$-\frac{1}{2} [Y' V^{-1} Y - Y' V X\beta - (X\beta)' V^{-1} Y + (X\beta)' V X\beta]$$

En retenant toujours que les termes faisant intervenir β et en remarquant que $Y' V^{-1} X\beta = (X\beta)' V^{-1} Y$ on obtient l'expression suivante:

$$-\frac{1}{2} [-2Y' V^{-1} X\beta + (X\beta)' V^{-1} X\beta] \quad (2.2)$$

Puisque l'on souhaite calculer $\frac{\delta \ln(L(Y, \theta))}{\delta \beta}$, alors on dérive (2.2) suivant β

$$\begin{aligned} \frac{\delta \ln(L(Y, \theta))}{\delta \beta} &= \frac{\delta \left[-\frac{1}{2} \left[-2Y' V^{-1} X \beta + (X \beta)' V^{-1} X \beta \right] \right]}{\delta \beta} \\ &= \frac{\delta \left[Y' V^{-1} X \beta \right]}{\delta \beta} - \frac{1}{2} \frac{\delta \left[(X \beta)' V^{-1} X \beta \right]}{\delta \beta} \\ &= Y' V^{-1} X - X' V^{-1} X \beta \end{aligned}$$

Et en remarquant que $Y' V^{-1} X = X' V^{-1} Y$, on a:

$$\boxed{\frac{\delta \ln(L(Y, \theta))}{\delta \beta} = X' V^{-1} Y - X' V^{-1} X \beta} \quad (2.3)$$

$$\begin{aligned} \frac{\delta \ln(L(Y, \theta))}{\delta \beta} = 0 &\Leftrightarrow X' V^{-1} Y - X' V^{-1} X \beta = 0 \\ &\Leftrightarrow X' V^{-1} X \beta = X' V^{-1} Y \\ &\Leftrightarrow \beta = (X' V^{-1} X)^{-1} X' V^{-1} Y \end{aligned}$$

$\beta = \hat{\beta}$ est un maximum local si la condition suffisante: $\frac{\delta^2 \ln(L(Y, \theta))}{\delta \beta^2} < 0$ est vérifiée.

$$\boxed{\hat{\beta} = (X' V^{-1} X)^{-1} X' V^{-1} Y} \quad (2.4)$$

$$\boxed{\frac{\delta^2 \ln(L(Y, \theta))}{\delta \beta^2} = -X' V^{-1} X = -X' \left(\sum_{k=1}^K \sigma_k^2 Z_k Z_k' + \sigma_\varepsilon^2 I_n \right)^{-1} X} \quad (2.5)$$

2.1.2 Estimation des variances α_k

Cette fois on s'intéresse qu'aux termes de (2.1) faisant intervenir α_k c'est à dire:

$$-\frac{1}{2} \ln(\det(V)) - \frac{1}{2} (Y - X \beta)' V^{-1} (Y - X \beta) \quad (2.6)$$

On dérive les termes de (2.6) suivant α_k .

$$\frac{\delta \ln(L(Y, \theta))}{\delta \alpha_k} = -\frac{1}{2} \frac{\delta [\ln(\det(V))]}{\delta \alpha_k} - \frac{1}{2} \frac{\delta [(Y - X\beta)' V^{-1} (Y - X\beta)]}{\delta \alpha_k} \quad (2.7)$$

Quelques généralités sur les différentielles de matrice.

Theorem 2.1.1 Soit S une partie ouverte de $\mathbb{R}^{n \times q}$. Si la fonction matricielle $F : S \rightarrow T_+$ est k fois différentiable dans S , alors la fonction $\log(\det(F)) : S \rightarrow \mathbb{R}$ est donnée par $(\log(\det(F)))(X) = \log(\det(F(X)))$. De plus,

$$d \log(\det(F)) = \text{tr}(F^{-1} dF)$$

Theorem 2.1.2 Soit $T = (Y : Y \in \mathbb{R}^{m \times m}, \det(Y) \neq 0)$. Soit S une partie ouverte de $\mathbb{R}^{n \times q}$. Si la fonction matricielle $F : S \rightarrow T$ est k fois différentiable dans S , alors la fonction $F^{-1} : S \rightarrow T$ est donnée par $(F^{-1})(X) = (F(X))^{-1}$, et

$$dF^{-1} = -F^{-1}(dF)F^{-1}$$

D'après les résultats fondamentaux des théorèmes 2.1.1 et 2.1.2

$$\bullet \frac{\delta \ln(\det(V))}{\delta \alpha_k} = \text{tr} \left(V^{-1} \frac{\delta V}{\delta \alpha_k} \right) \quad (2.8)$$

$$\bullet \frac{\delta V^{-1}}{\delta \alpha_k} = -V^{-1} \frac{\delta V}{\delta \alpha_k} V^{-1} \quad (2.9)$$

En remplaçant les résultats (2.8) et (2.9) dans l'expression (2.7) on obtient:

$$\boxed{\frac{\delta \ln(L(Y, \theta))}{\delta \alpha_k} = -\frac{1}{2} \text{tr} \left(V^{-1} \frac{\delta V}{\delta \alpha_k} \right) + \frac{1}{2} (Y - X\beta)' V^{-1} \frac{\delta V}{\delta \alpha_k} V^{-1} (Y - X\beta)} \quad (2.10)$$

$$\frac{\delta \ln(L(Y, \theta))}{\delta \alpha_k} = 0 \Leftrightarrow -\text{tr} \left(V^{-1} \frac{\delta V}{\delta \alpha_k} \right) + (Y - X\beta)' V^{-1} \frac{\delta V}{\delta \alpha_k} V^{-1} (Y - X\beta) = 0 \quad (2.11)$$

On évalue l'équation (2.11) en $\beta = \hat{\beta}$ et $V = \hat{V}$

$$\text{tr} \left(\hat{V}^{-1} \frac{\delta \hat{V}}{\delta \alpha_k} \right) = (Y - X\hat{\beta})' \hat{V}^{-1} \frac{\delta \hat{V}}{\delta \alpha_k} \hat{V}^{-1} (Y - X\hat{\beta}) \quad (2.12)$$

On s'intéresse à $\hat{V}^{-1} (Y - X\hat{\beta})$ qui en remplaçant $\hat{\beta}$ par sa valeur calculée en (2.5) s'écrit comme $\hat{V}^{-1} (Y - X (X' V^{-1} X)^{-1} X' V^{-1} Y)$. On remarque que $Q = X (X' V^{-1} X)^{-1} X' V^{-1}$

est le projecteur des moindres carrés généralisés. Ce qui nous permet d'écrire que $\hat{V}^{-1} (Y - X\hat{\beta}) = \hat{V}^{-1} (I - Q) Y$, ou encore $\hat{V}^{-1} (Y - X\hat{\beta}) = \hat{P}Y$ avec $\hat{P} = \hat{V}^{-1} (I - Q)$.

La nouvelle écriture de (2.12) est:

$$tr \left(\hat{V}^{-1} \frac{\delta \hat{V}}{\delta \alpha_k} \right) = Y' \hat{P} \frac{\delta \hat{V}}{\delta \alpha_k} \hat{P} Y \quad (2.13)$$

On s'intéresse au terme de gauche de (2.13) qui dû à la linéarité de V s'explique comme:

$$tr \left(\hat{V}^{-1} \frac{\delta \hat{V}}{\delta \alpha_k} \right) = \sum_{k=1}^{K+1} tr \left(\hat{V}^{-1} \frac{\delta \hat{V}}{\delta \alpha_k} \hat{V}^{-1} \frac{\delta \hat{V}}{\delta \alpha_l} \right) \hat{\alpha}_l$$

Et donc

$$\sum_{l=1}^{K+1} tr \left(\hat{V}^{-1} \frac{\delta \hat{V}}{\delta \alpha_k} \hat{V}^{-1} \frac{\delta \hat{V}}{\delta \alpha_l} \right) \hat{\alpha}_l = Y' \hat{P} \frac{\delta \hat{V}}{\delta \alpha_k} \hat{P} Y$$

Soit encore sous forme matricielle (réf. Foulley^[6])

$$\boxed{\hat{F} \hat{\alpha} = \hat{g}} \quad (2.14)$$

où F est une matrice $(K+1) \times (K+1)$ symétrique et g un vecteur $(K+1)$ définis par:
 $F = [f_{kl}] = \left[tr \left(V^{-1} \frac{\delta V}{\delta \alpha_k} V^{-1} \frac{\delta V}{\delta \alpha_l} \right) \right]$ et $g = [g_k] = \left[Y' P \frac{\delta V}{\delta \alpha_k} P Y \right]$.
 \hat{F} et \hat{g} correspondent à F et g évalués en $\alpha = \hat{\alpha}$.

Le système en (2.14) est non linéaire qui, en général, n'a pas de solution analytique. On le résout numériquement par un algorithme itératif ayant la forme d'un système linéaire en α :

$$\boxed{F(\alpha^{(n)}) \alpha^{(n+1)} = g(\alpha^{(n)})} \quad (2.15)$$

On souhaite à présent calculer $\frac{\delta^2 \ln(L(Y, \theta))}{\delta \alpha_l \delta \alpha_k}$, pour cela on va d'abord dériver le premier terme de (2.10) suivant α_l .

$$\begin{aligned} -\frac{1}{2} \frac{\delta \left[\text{tr} \left(V^{-1} \frac{\delta V}{\delta \alpha_k} \right) \right]}{\delta \alpha_l} &= -\frac{1}{2} \text{tr} \left(\frac{\delta V^{-1}}{\delta \alpha_l} \frac{\delta V}{\delta \alpha_k} \right) - \frac{1}{2} \text{tr} \left(V^{-1} \frac{\delta^2 V}{\delta \alpha_l \delta \alpha_k} \right) \\ &= \frac{1}{2} \text{tr} \left(V^{-1} \frac{\delta V}{\delta \alpha_l} V^{-1} \frac{\delta V}{\delta \alpha_k} \right) - \frac{1}{2} \text{tr} \left(V^{-1} \frac{\delta^2 V}{\delta \alpha_l \delta \alpha_k} \right) \end{aligned} \quad (2.16)$$

Ensuite, on dérive le second terme de (2.10) suivant γ_l .

$$\begin{aligned} \frac{1}{2} \frac{\delta \left[(Y - X\beta)' V^{-1} \frac{\delta V}{\delta \alpha_k} V^{-1} (Y - X\beta) \right]}{\delta \alpha_l} &= \frac{1}{2} (Y - X\beta)' \frac{\delta V^{-1}}{\delta \alpha_l} \frac{\delta V}{\delta \alpha_k} V^{-1} (Y - X\beta) \\ &\quad + \frac{1}{2} (Y - X\beta)' V^{-1} \frac{\delta}{\delta \alpha_l} \left[\frac{\delta V}{\delta \alpha_k} V^{-1} \right] (Y - X\beta) \end{aligned}$$

On remarque d'abord que $\frac{\delta V^{-1}}{\delta \alpha_l} \frac{\delta V}{\delta \alpha_k} V^{-1} = V^{-1} \left(-\frac{\delta V}{\delta \alpha_l} V^{-1} \frac{\delta V}{\delta \alpha_k} \right) V^{-1}$

$$\begin{aligned} \text{La seconde remarque porte sur } V^{-1} \frac{\delta}{\delta \alpha_l} \left[\frac{\delta V}{\delta \alpha_k} V^{-1} \right] &= V^{-1} \left(\frac{\delta^2 V}{\delta \alpha_l \delta \alpha_k} V^{-1} + \frac{\delta V}{\delta \alpha_k} \frac{\delta V^{-1}}{\delta \alpha_l} \right) \\ &= V^{-1} \left(\frac{\delta^2 V}{\delta \alpha_l \delta \alpha_k} V^{-1} - \frac{\delta V}{\delta \alpha_k} V^{-1} \frac{\delta V}{\delta \alpha_l} V^{-1} \right) \\ &= V^{-1} \left(\frac{\delta^2 V}{\delta \alpha_l \delta \alpha_k} - \frac{\delta V}{\delta \alpha_k} V^{-1} \frac{\delta V}{\delta \alpha_l} \right) V^{-1} \end{aligned}$$

Ce qui permet de réécrire

$$\frac{1}{2} \frac{\delta \left[(Y - X\beta)' V^{-1} \frac{\delta V}{\delta \alpha_k} V^{-1} (Y - X\beta) \right]}{\delta \alpha_l} = \frac{1}{2} (Y - X\beta)' V^{-1} \left(-2 \frac{\delta V}{\delta \alpha_l} V^{-1} \frac{\delta V}{\delta \alpha_k} + \frac{\delta^2 V}{\delta \alpha_l \delta \alpha_k} \right) V^{-1} (Y - X\beta) \quad (2.17)$$

Finalement on obtient $\frac{\delta^2 \ln(L(Y, \theta))}{\delta \alpha_l \delta \alpha_k}$ en utilisant les résultats en (2.16) et (2.17).

$$\begin{aligned} \frac{\delta^2 \ln(L(Y, \theta))}{\delta \alpha_l \delta \alpha_k} &= \frac{1}{2} (Y - X\beta)' V^{-1} \left(-2 \frac{\delta V}{\delta \alpha_l} V^{-1} \frac{\delta V}{\delta \alpha_k} + \frac{\delta^2 V}{\delta \alpha_l \delta \alpha_k} \right) V^{-1} (Y - X\beta) \\ &\quad + \frac{1}{2} \text{tr} \left(V^{-1} \frac{\delta V}{\delta \alpha_l} V^{-1} \frac{\delta V}{\delta \alpha_k} \right) - \frac{1}{2} \text{tr} \left(V^{-1} \frac{\delta^2 V}{\delta \alpha_l \delta \alpha_k} \right) \end{aligned}$$

En particulier à partir de (2.18) on peut déterminer $\frac{\delta^2 \ln(L(Y, \theta))}{\delta^2 \sigma_k^2}$, $\frac{\delta^2 \ln(L(Y, \theta))}{\delta \sigma_k^2 \delta \sigma_\varepsilon^2}$ et $\frac{\delta^2 \ln(L(Y, \theta))}{\delta^2 \sigma_\varepsilon^2}$,

- $\frac{\delta^2 \ln(L(Y, \theta))}{\delta^2 \sigma_k^2} = - (Y - X\beta)' V^{-1} (Z_k Z_k' V^{-1} Z_k Z_k') V^{-1} (Y - X\beta) + \frac{1}{2} \text{tr} (V^{-1} Z_k Z_k' V^{-1} Z_k Z_k')$
- $\frac{\delta^2 \ln(L(Y, \theta))}{\delta \sigma_k^2 \delta \sigma_\varepsilon^2} = \frac{1}{2} \text{tr}(V^{-2} Z_k Z_k') - (Y - X\beta)' V^{-2} Z_k Z_k' V^{-1} (Y - X\beta)$
- $\frac{\delta^2 \ln(L(Y, \theta))}{\delta^2 \sigma_\varepsilon^2} = \frac{1}{2} (Y - X\beta)' V^{-3} (Y - X\beta) + \frac{1}{2} \text{tr}(V^{-2})$

2.1.3 Fisher scoring pour l'estimation de θ

Pour déterminer les estimateurs $\hat{\beta}$, $\hat{\sigma}_k^2$ et $\hat{\sigma}_\varepsilon^2$, on peut aussi utiliser l'algorithme des scores de Fisher (réf. Laurent DAVEZIES^[5]), qui se base sur la méthode de Newton-Raphson pour résoudre les équations du maximum de vraisemblance.

$$\theta^{(n+1)} = \theta^{(n)} + I^{-1}(\theta^{(n)}) H(\theta^{(n)})$$

où I est la matrice d'information de Fisher, et $H(\theta^{(n)}) \approx H(\theta^{(n-1)}) - I(\theta^{(n-1)})(\theta^{(n)} - \theta^{(n-1)})$.

Quelques rappels sur le score et l'information de Fisher.

On suppose que $\Theta \subseteq \mathbb{R}^d$, Θ ouvert et $d \geq 1$, et on considère Y une variable aléatoire vectorielle de dimension d à valeurs y dans χ .

Hypothèses: H1. $\{y \in \chi : f(y, \theta) > 0\}$ ne dépend pas de $\theta \in \Theta$

H2. La fonction $\theta \rightarrow f(y, \theta)$ est $C^2(\Theta)$

H3. $\forall A \subseteq \chi$

$$\frac{\delta}{\delta \theta_i} \int_{y \in A} f(y, \theta) dy = \int_{y \in A} \frac{\delta f(y, \theta)}{\delta \theta_i} dy \quad i = 1, \dots, d$$

$$\frac{\delta^2}{\delta \theta_i \delta \theta_j} \int_{y \in A} f(y, \theta) dy = \int_{y \in A} \frac{\delta^2 f(y, \theta)}{\delta \theta_i \delta \theta_j} dy \quad i, j = 1, \dots, d$$

Definition 2.1.1 — Score et information de Fisher. On appelle score la fonction $S(y, \theta) = (S_1(y, \theta), \dots, S_d(y, \theta))$ à valeurs dans \mathbb{R}^d où

$$S_i(y, \theta) = \frac{\delta \ln(f(y, \theta))}{\delta \theta_i}$$

On appelle l'information de Fisher la matrice symétrique $I(\theta)$ de dimension $d \times d$ donnée par:

$$I(\theta) = \mathbb{E}_\theta [S(y, \theta)S(y, \theta)']$$

et donc pour tout $1 \leq i, j \leq d$

$$I(\theta)_{ij} = \mathbb{E}_\theta [S_i(y, \theta)S_j(y, \theta)] = \mathbb{E}_\theta \left[\frac{\delta \ln(f(y, \theta))}{\delta \theta_i} \cdot \frac{\delta \ln(f(y, \theta))}{\delta \theta_j} \right]$$

Theorem 2.1.3 Si **H1**, **H2**, **H3** sont vérifiées, alors

$$\mathbb{E}_\theta [S(y, \theta)] = 0 \text{ et } I(\theta)_{ij} = -\mathbb{E}_\theta \left[\frac{\delta^2 \ln(f(y, \theta))}{\delta \theta_i \delta \theta_j} \right] = -\mathbb{E}_\theta \left[\frac{\delta S_j(y, \theta)}{\delta \theta_i} \right]$$

En utilisant les résultats (2.5), (2.19) et les résultats (réf. Laurent DAVEZIES^[5]) concernant les valeurs de $\frac{\delta^2 \ln(L(Y, \theta))}{\delta \beta \delta \sigma_k^2}$ et $\frac{\delta^2 \ln(L(Y, \theta))}{\delta \beta \delta \sigma_\varepsilon^2}$, on peut déterminer la matrice d'information de Fisher I de paramètre $\theta = (\beta', \alpha)$ associées à notre modèle.

$$I(\theta) = \mathbb{E} \begin{bmatrix} X'V^{-1}X & 0 & 0 \\ 0 & \frac{1}{2}tr(V^{-2}) & \frac{1}{2}tr(V^{-2}Z_k Z_k') \\ 0 & \frac{1}{2}tr(V^{-2}Z_k Z_k') & \frac{1}{2}tr(V^{-1}Z_k Z_k' V^{-1}Z_k Z_k') \end{bmatrix}$$

2.2 Etude de la qualité des estimateurs

2.2.1 Etude de la qualité de $\hat{\beta}$

Hypothèses: H1. Linéaire en les paramètres : $Y = X\beta + \sum_{k=1}^K Z_k \gamma_k + \varepsilon$

H2. Moyenne conditionnelle nulle : $\mathbb{E}[\varepsilon | X] = 0$ et $\mathbb{E}[\gamma_k | X] = 0 \quad \forall k = 1, \dots, K$

H3. Homoscédasticité : $\text{Var}(\varepsilon | X) = \sigma_\varepsilon^2 I_n$ et $\text{Var}(\gamma_k | X) = \sigma_k^2 I_{q_k} \quad \forall k = 1, \dots, K$

H4. Absence de corrélation : $\mathbb{E}[\varepsilon_i \varepsilon_j | X] = 0$ et $\mathbb{E}[\gamma_{ki} \gamma_{kj} | X] = 0 \quad \forall i, j = 1, \dots, n \text{ et } i$

R L'hypothèse **H3** assure que les variables aléatoires sont identiquement distribuées, et **H4** assure leurs indépendances.

Definition 2.2.1 Un estimateur $\hat{\theta}_0$ est optimal dans une classe d'estimateurs $\hat{\theta}$ si toute estimation d'une combinaison linéaire du paramètre est estimée plus précisément par $\hat{\theta}_0$ que par tout autre estimateur de la classe:

$$\text{Var}(\lambda' \hat{\theta}_0) \leq \text{Var}(\lambda' \hat{\theta}) \quad \forall \lambda$$

Theorem 2.2.1 — Théorème de Gauss-Markov. Sous les hypothèses **H1, H2, H3, H4**, l'estimateur $\hat{\beta}$ construit par ML à partir du modèle $Y = X\beta + \sum_{k=1}^K Z_k \gamma_k + \varepsilon$ est optimale dans la classe des estimateurs linéaires sans biais, et il est communément appelé BLUE ("Best Unbiased Linear Estimator").

Démonstration.

L'idée de la preuve consiste à prouver que $\hat{\beta}$ est **i.** linéaire, **ii.** sans biais et de **iii.** variance minimale par rapport aux autres estimateurs.

i. La linéarité de $\hat{\beta}$ est mise en évidence par une simple réécriture, on rappelle que

$$\hat{\beta} = (X' V^{-1} X)^{-1} X' V^{-1} Y$$

On pose

$$W = (X' V^{-1} X)^{-1} X' V^{-1}$$

La nouvelle écriture de $\hat{\beta}$ est:

$$\begin{aligned}
 \hat{\beta} &= WY \\
 &= W \left(X\beta + \sum_{k=1}^K Z_k \gamma_k + \varepsilon \right) \\
 &= WX\beta + W \sum_{k=1}^K Z_k \gamma_k + W\varepsilon \\
 &= (X'V^{-1}X)^{-1} X'V^{-1}X\beta + W \sum_{k=1}^K Z_k \gamma_k + W\varepsilon \\
 &= \beta + W \sum_{k=1}^K Z_k \gamma_k + W\varepsilon
 \end{aligned}$$

ii. Le caractère non biaisé de $\hat{\beta}$ est explicité par $Biais(\hat{\beta}) = \mathbb{E}[\hat{\beta}] - \beta = 0$

$$\begin{aligned}
 \mathbb{E}[\hat{\beta} | X] &= \mathbb{E}[\beta + W \sum_{k=1}^K Z_k \gamma_k + W\varepsilon | X] \\
 &= \beta + \mathbb{E}[W \sum_{k=1}^K Z_k \gamma_k | X] + \mathbb{E}[W\varepsilon | X] \\
 &= \beta + W \sum_{k=1}^K Z_k \mathbb{E}[\gamma_k | X] + W \mathbb{E}[\varepsilon | X]
 \end{aligned}$$

Sous l'hypothèse **H2** : $\mathbb{E}[\gamma_k | X] = 0$ et $\mathbb{E}[\varepsilon | X] = 0$, ce qui implique que $\mathbb{E}[\hat{\beta} | X] = \beta$

Comme $\mathbb{E}[\hat{\beta}] = \mathbb{E}[\mathbb{E}[\hat{\beta} | X]] = \beta$, alors $Biais(\hat{\beta}) = 0$

iii. (Reste à prouver.)

■

Proposition 2.2.2 — Distribution du vecteur des coefficients $\hat{\beta}$. $\hat{\beta}$ est normalement distribué de moyenne $\mu_{\hat{\beta}} = \beta$ et de matrice de covariance $V_{\hat{\beta}} = \sum_{k=1}^K W Z_k \Sigma_k (W Z_k)' + W \sigma_{\varepsilon}^2 I_n W'$.

Démonstration.

Le vecteur $\hat{\beta}$ est un vecteur gaussien de \mathbb{R}^{p+1} , en effet pour tout $c = (c_1, \dots, c_{p+1})' \in \mathbb{R}^{p+1}$ la variable $\sum_{i=1}^{p+1} c_i \hat{\beta}_i = c' \hat{\beta} = c' (\beta + W \sum_{k=1}^K Z_k \gamma_k + W \varepsilon) = c' \beta + c' W \sum_{k=1}^K Z_k \gamma_k + c' W \varepsilon$ est une variable réelle gaussienne, puisque $c' \beta$ est un réel et que $c' W \sum_{k=1}^K Z_k \gamma_k + c' W \varepsilon$ est une somme de variables réelles gaussiennes.

$\hat{\beta}$ est de moyenne $\mu_{\hat{\beta}}$ et de matrice de covariance $V_{\hat{\beta}}$ tel que:

$$\mu_{\hat{\beta}} = \mathbb{E}[\hat{\beta}] = \beta \quad (\text{d'après dem. ii. thm. 2.2.1})$$

$$\begin{aligned} V_{\hat{\beta}} &= \text{Var}(\hat{\beta}) \\ &= \text{Var}(\beta) + \text{Var}\left(W \sum_{k=1}^K Z_k \gamma_k\right) + \text{Var}(W \varepsilon) \\ &= \sum_{k=1}^K W Z_k \text{Var}(\gamma_k) (W Z_k)' + W \text{Var}(\varepsilon) W' \\ &= \sum_{k=1}^K W Z_k \Sigma_k (W Z_k)' + W \sigma_{\varepsilon}^2 I_n W' \quad \blacksquare \end{aligned}$$

Theorem 2.2.3 — Information de Fisher et loi asymptotique d'un estimateur par ML. —

Soit $P = \{P_{\theta} : \theta \in \Theta\}$ un modèle paramétrique vérifiant les hypothèses **H1**, **H2**, **H3** tel que $0 < I(\theta) < +\infty$. On suppose que l'estimateur par ML $\hat{\theta}_n$ existe et est unique. Soit θ_0 la vraie valeur du paramètre θ inconnu. S'il existe $\delta > 0$ et $k(y) \geq 0$ tels que:

- i. $\left| \frac{\delta^2 \ln(f(y, \theta))}{\delta \theta^2} \right| \leq k(y) \quad \forall \theta \in [\theta_0 - \delta, \theta_0 + \delta]$
- ii. $\mathbb{E}_{\theta_0}[k(Y)] < +\infty$

Alors

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow \mathcal{N}(0, I(\theta_0)^{-1})$$

D'après le théorème 2.2.3 à condition que les points **i.** et **ii.** soient vérifiés, la loi asymptotique de $\hat{\beta}$ est donnée par:

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta) &\rightarrow \mathcal{N}(0, I(\theta)_{11}^{-1}) &\Leftrightarrow &\sqrt{n}(\hat{\beta} - \beta) \rightarrow \mathcal{N}(0, (X' V^{-1} X)^{-1}) \\ &&\Leftrightarrow &\hat{\beta} \rightarrow \mathcal{N}\left(\beta, \frac{1}{n} (X' V^{-1} X)^{-1}\right)\end{aligned}$$

2.2.2 Etude de la qualité de \hat{V}

$$\hat{V} = \sum_{k=1}^K \hat{\sigma}_k^2 Z_k Z_k' + \hat{\sigma}_\varepsilon^2 I_n$$

Pour $\sigma_k^2 > 0$ et $\sigma_\varepsilon^2 > 0$, cet estimateur est asymptotiquement sans biais, et asymptotiquement optimal parmi les estimateurs asymptotiquement sans biais (réf. Laurent DAVEZIES^[3]). De plus pour $\sigma_k^2 > 0$ et $\sigma_\varepsilon^2 > 0$:

$$\sqrt{n} \left((\hat{\sigma}_k^2 | \hat{\sigma}_\varepsilon^2)' - (\sigma_k^2 | \sigma_\varepsilon^2)' \right) \rightarrow \mathcal{N}(0, V_a)$$

où la variance asymptotique de l'estimateur du maximum de vraisemblance est:

$$V_a = \lim_{n \rightarrow +\infty} \frac{1}{2n} \begin{bmatrix} \text{tr}(V^{-1} \sum_{k=1}^K Z_k Z_k' V^{-1} Z_k Z_k') & \text{tr}(V^{-1} Z_k Z_k' V^{-1}) \\ \text{tr}(V^{-1} Z_k Z_k' V^{-1}) & \text{tr}(V^{-2}) \end{bmatrix}^{-1}$$

R L'estimateur du maximum de vraisemblance doit ses propriétés à des développements de certaines fonctionnelles autour des paramètres (ici $\beta, \sigma_k^2, \sigma_\varepsilon^2$). Pour pouvoir faire ces développements de manière standard, il faut que le voisinage des paramètres soit suffisamment 'régulier', c'est le cas lorsque $\sigma_k^2 > 0, \sigma_\varepsilon^2 > 0$. Dans le cas contraire, ($\sigma_k^2 = 0$) σ_k^2 ne peut pas être approché par valeurs inférieures et on sort du cadre habituel des théorèmes de convergence de l'estimateur du maximum de vraisemblance. C'est pour cette raison que les propriétés sont énoncées sous la condition que $\sigma_k^2 > 0, \sigma_\varepsilon^2 > 0$.



3. Méthode REML

La méthode du maximum de vraisemblance, qui entraîne un biais systématique dans l'estimation de la variance, n'est pas la plus appropriée dans ce cas: on lui préfère, en général, la méthode du maximum de vraisemblance restreint dite REML ("REstricted Maximum Likelihood").

On rappelle que le cadre général d'étude est:

$$Y = X\beta + \sum_{k=1}^K Z_k \gamma_k + \varepsilon$$

avec $Y \hookrightarrow \mathcal{N}(X\beta, V)$ de matrice de covariance $V = \sum_{k=1}^K Z_k \sigma_k^2 Z_k' + \sigma_\varepsilon^2 I_n$

L'estimateur du maximum de vraisemblance restreint de σ_k^2 et σ_ε^2 est construit sur la vraisemblance de $U = A'Y$ où A est une matrice $n \times (n-p-1)$ de plein rang dont les colonnes sont orthogonales aux colonnes de X . Le vecteur U est normalement distribué de moyenne 0 et de matrice de covariance $A'VA$ qui ne dépend pas de β . Further, Harville[1974] ont montrés que la fonction de vraisemblance restreinte peut s'écrire comme (réf. Geert Verbeke, Geert Molenberghs^[4]):

$$L(U, \theta) = (2\pi)^{-\frac{(n-p-1)}{2}} \det(X'X)^{\frac{1}{2}} \det(X'V^{-1}X)^{-\frac{1}{2}} \det(V)^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (Y - X\beta)' V^{-1} (Y - X\beta) \right) \quad (3.1)$$

Finalement, en remarquant que $(2\pi)^{-\frac{(p-1)}{2}} \det(X'X)^{\frac{1}{2}} = cst$, et en négligeant ce terme on peut réécrire (3.1) comme:

$$L_{REML}(U, \theta) = \det(X'V^{-1}X)^{-\frac{1}{2}} L_{ML}(Y, \theta) \quad (3.2)$$



4. Tests d'hypothèses et critère de sélection

4.1 Test de Wald

Le test de Wald cherche à déterminer s'il y a une différence significative entre l'estimateur de β lorsque nous estimons le modèle restreint et le modèle non restreint.

$$\begin{cases} (\mathbf{H0}) : k' \beta = m \\ (\mathbf{H1}) : k' \beta \neq m \end{cases}$$

où k est une matrice $(p+1) \times r$ avec $r < p+1$, dont les r colonnes sont linéairements indépendantes et m est un vecteur $r \times 1$ de constantes.

On suppose que l'hypothèse $(\mathbf{H0}) : k' \beta = m$ est vraie.

Soit $\hat{\beta}$ l'estimateur par ML du coefficient β , d'après la remarque faite précédemment $\hat{\beta}$ est asymptotiquement normalement distribué. Dans ce cas on peut déduire la loi asymptotique de $k' \hat{\beta}$ comme suit:

$$\begin{aligned} \sqrt{n}(k' \hat{\beta} - m) &\rightarrow \mathcal{N}(0, k' I(\theta)^{-1} k) \Leftrightarrow \sqrt{n}(k' \hat{\beta} - m) \rightarrow \mathcal{N}(0, k'(X' V^{-1} X)^{-1} k) \\ &\Leftrightarrow \sqrt{n}(k'(X' V^{-1} X)^{-1} k)^{-\frac{1}{2}} (k' \hat{\beta} - m) \rightarrow \mathcal{N}(0, I_r) \end{aligned}$$

On note

$$J = \text{Var}(k' \hat{\beta}) = k'(X' V^{-1} X)^{-1} k$$

La statistique de test de Wald est:

$$w_n = n (k' \hat{\beta} - m)' J^{-1} (k' \hat{\beta} - m)$$

La statistique de Wald w_n suit asymptotiquement une loi de χ^2 à r degrés de liberté (réf. Z. Griliches and M.D. Intriligator^[8]).

On rappelle que le risque de première espèce α est la probabilité de rejeter (H_0) alors qu'elle est vraie.

Si l'hypothèse **(H1)** : $k' \hat{\beta} < m$ est vraie,
alors on rejette (H_0) si $w_n < q_{\frac{\alpha}{2}}$ où $q_{\frac{\alpha}{2}}$ est le quantile d'ordre $\frac{\alpha}{2}$ de χ_r^2 .

Si l'hypothèse **(H1)** : $k' \hat{\beta} > m$ est vraie,
alors on rejette (H_0) si $w_n > q_{1-\frac{\alpha}{2}}$ où $q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de χ_r^2 .

4.2 Test du ratio Log-Vraisemblance

Le test du ratio Log-Vraisemblance est un test pour comparer deux modèles imbriqués. Souvent noté "Likelihood Ratio", il correspond au log du ratio de vraisemblance entre le modèle complet et le modèle contraint.

Soit $\Theta \subseteq \mathbb{R}^d$, Θ ouvert et $d \geq 1$.

Pour $\theta \in \Theta$ on considère L le maximum de la Vraisemblance tel que $L = \sup\{L(Y, \theta)\}$.

$$\begin{cases} \text{(H0)} : \theta \in \Theta_0 \\ \text{(H1)} : \theta \notin \Theta_0 \end{cases} \quad \text{où } \Theta_0 \subseteq \Theta$$

On suppose que l'hypothèse **(H0)** : $\theta \in \Theta_0$ est vraie.

Sous **(H0)** : $\theta = \theta_0$ et le maximum de la Vraisemblance s'écrit:

$$L_0 = \sup\{L(Y, \theta_0)\}$$

La statistique de test du ratio Log-Vraisemblance est:

$$\lambda = 2 \ln\left(\frac{L}{L_0}\right) = 2 \ln(L) - 2 \ln(L_0)$$

λ suit asymptotiquement une loi de χ^2 à d_0 degrés de liberté, où $d_0 = \dim(\Theta_0)$.

Idée de la preuve.^[7]

Soit f la fonction de $\mathbb{R}_+^* \rightarrow \mathbb{R}$ définie par $f : x \rightarrow \ln(\sup(x))$
et la fonction de vraisemblance L de \mathbb{R}^d dans \mathbb{R}_+^* définie par $L : \theta \rightarrow L(Y, \theta)$.
Soit de plus la fonction composée h de $\mathbb{R}_+^* \rightarrow \mathbb{R}$ tel que $h = f \circ L$.

D'après le théorème de Taylor-Young,
le développement limité de la fonction composée h en $\theta = \hat{\theta}$ vaut:

$$h(\theta) \approx h(\hat{\theta}) + (\theta - \hat{\theta})' h'(\hat{\theta}) + \frac{1}{2} (\theta - \hat{\theta})' h''(\hat{\theta}) (\theta - \hat{\theta})$$

Puisqu'au point $\theta = \hat{\theta}$ le maximum est atteint, alors $h'(\hat{\theta}) = \ln(L(Y, \hat{\theta}))' = 0$
et $h''(\hat{\theta}) = \ln(L(Y, \hat{\theta}))''$, ce qui donne:

$$h(\theta) - h(\hat{\theta}) \approx \frac{1}{2} (\theta - \hat{\theta})' \ln(L(Y, \hat{\theta}))'' (\theta - \hat{\theta})$$

Sous $(H_0) : \theta = \theta_0$, et on a $h(\theta_0) - h(\hat{\theta}) \approx \frac{1}{2} (\theta_0 - \hat{\theta})' \ln(L(Y, \hat{\theta}))'' (\theta_0 - \hat{\theta})$

Comme $\lambda = 2 (h(\hat{\theta}) - h(\theta_0))$ alors:

$$\lambda \approx -(\hat{\theta} - \theta_0)' \ln(L(Y, \hat{\theta}))'' (\hat{\theta} - \theta_0) \quad (4.1)$$

On réécrit (2.16) afin de mettre en évidence la loi asymptotique de $\hat{\theta}$

$$\begin{aligned} \lambda &\approx -(\hat{\theta} - \theta_0)' \frac{n}{I(\theta_0)^{-1}} \frac{I(\theta_0)^{-1}}{n} \ln(L(Y, \hat{\theta}))'' (\hat{\theta} - \theta_0) \\ \Leftrightarrow \lambda &\approx -(\hat{\theta} - \theta_0)' \frac{\sqrt{n}}{(I(\theta_0)^{-1})^{\frac{1}{2}}} \frac{I(\theta_0)^{-1}}{n} \ln(L(Y, \hat{\theta}))'' \frac{\sqrt{n}}{(I(\theta_0)^{-1})^{\frac{1}{2}}} (\hat{\theta} - \theta_0) \end{aligned}$$

D'après le théorème 2.2.2 à condition que les point **i.** et **ii.** soient vérifiés, la loi asymptotique de $\hat{\theta}$ est donnée par:

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow \mathcal{N}(0, I(\theta_0)^{-1}) \Leftrightarrow \sqrt{n} \frac{(\hat{\theta} - \theta_0)}{(I(\theta_0)^{-1})^{\frac{1}{2}}} \rightarrow \mathcal{N}(0, I_{d_0})$$

On note $Z = -(\hat{\theta} - \theta_0)' \frac{n}{I(\theta_0)^{-1}} (\hat{\theta} - \theta_0)$

A fortiori Z suit une loi du χ^2 à d_0 degrés de liberté.

Soit maintenant θ^* qui est entre θ_0 et $\hat{\theta}$, on a donc sous (H_0) :

$$\frac{1}{n} \ln(L(Y, \theta^*))'' \rightarrow -\mathbb{E}_{\theta_0}[\ln(L(Y, \theta_0))''] = I(\theta_0) \quad (\text{convergence en probabilité})$$

On note $U = I(\theta_0)^{-1} \frac{1}{n} \ln(L(Y, \hat{\theta}))''$, et on remarque que U converge en probabilité vers 1.

On résume qu'à présent on a: $\lambda \approx ZU$ avec $Z \rightarrow \chi^2(d_0)$ et $U \rightarrow 1$.

D'après le théorème de Slutsky $\lambda \approx ZU \rightarrow \chi^2(d_0)$. ■

Règle de décision.

Si l'hypothèse **(H1)** : $\theta \notin \Theta_0$ est vraie, et en particulier si $\theta_0 < \theta_1$ alors on rejette (H_0) si $\lambda < q_{\frac{\alpha}{2}}$ où $q_{\frac{\alpha}{2}}$ est le quantile d'ordre $\frac{\alpha}{2}$ de $\chi_{q_0}^2$.

Si l'hypothèse **(H1)** : $\theta \notin \Theta_0$ est vraie, et en particulier si $\theta_0 > \theta_1$ alors on rejette (H_0) si $\lambda > q_{1-\frac{\alpha}{2}}$ où $q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de $\chi_{q_0}^2$.

4.3 Critère d'information AIC

Le critère d'information d'Akaike s'écrit comme suit:

$$AIC = 2k - 2\ln(L)$$

où k est le nombre de paramètres à estimer du modèle et L est le maximum de la fonction de vraisemblance du modèle.

Si l'on considère un ensemble de modèles candidats, le modèle choisi est celui qui aura la plus faible valeur d'AIC. Ce critère repose donc sur un compromis entre la qualité de l'ajustement et la complexité du modèle, en pénalisant les modèles ayant un grand nombre de paramètres, ce qui limite les effets de sur-ajustement (augmenter le nombre de paramètre améliore nécessairement la qualité de l'ajustement).

L'AIC est basé sur la théorie de l'information: il propose une estimation de la perte d'information lorsqu'on utilise le modèle considéré pour représenter le processus qui génère les données. L'AIC ne fournit pas un test de modèle dans le sens d'une hypothèse nulle, c'est-à-dire que ce test ne dit rien de la qualité absolue du modèle. Il ne rendrait ainsi pas compte du fait que tous les modèles candidats ne produisent pas de bons ajustements.

4.4 Critère d'information BIC

À la différence du critère d'information d'Akaike, la pénalité dépend de la taille de l'échantillon et pas seulement du nombre de paramètres.

Le critère d'information bayésien est:

$$BIC = -2\ln(L) + k\ln(n)$$

où L est le maximum de la fonction de vraisemblance du modèle, n est le nombre d'observations dans l'échantillon et k le nombre de paramètres du modèle.



5. Application

5.1 Introduction du modèle linéaire mixte pour l'étude de la dépendance alimentaire des oisillons

5.1.1 Présentation des données

L'exemple que nous introduisons est issu de la section 5.10 du livre "Mixed Effects Models and Extensions in Ecology with R" (réf. Alain F.Zuur, Elena N.Ieno, Neil J.Walker, Anatoly A.Saveliev, Graham M.Smith^[1]).

Roulin et Bersier (2007) ont étudiés le comportement de dépendance alimentaire des oisillons.

Les mesures ont été prises entre 21h30 et 05h30 dans 27 nids. Le nombre d'oisillons par nids varie entre 2 et 7. Le comportement de dépendance alimentaire a été évalué en calculant le nombre d'appel *SiblingNegotiation* que l'oisillon faisait sur une période de 30s avant l'arrivée d'un de ces parents divisé par le nombre d'oisillons *BroodSize* présent à ce moment dans le nid. On souhaite expliquer ce phénomène que l'on notera $NegPerChick_{ij} = \left(\frac{SiblingNegotiation}{BroodSize} \right)_{ij}$ pour j observations faites au nids i. Pour cela différentes variables explicatives ont été retenues telles que le sexe du parent, le traitement alimentaire, le temps d'arrivée du parent. Nous apportons une précision sur les deux dernières variables, le traitement alimentaire *Foodtreatment* comprend deux modalités *Deprived* et *Satiated* qui nous renseignent sur l'état nutritionnel des oisillons dans le nid. Le temps d'arrivée du parent *ArrivalTime* reflète l'heure à laquelle le parent arrive avec une proie dans le nids.

5.1.2 Présentation du modèle

Les raisons qui font que nous allons utiliser un modèle linéaire avec effet mixte sont les suivantes, d'abord nous sommes en présence de plusieurs observations pour les mêmes nids, alors ces observations devraient être corrélées. Si on fait par exemple le choix de considérer les 27 nids comme effets fixes, cela risque d'être coûteux en matière de degrés de liberté. De plus, on souhaite expliquer de manière générale un phénomène qui se produit dans un nid et non pas uniquement que pour les 27 nids d'étude. L'effet aléatoire sera donc le tirage du nid. Nous proposons pour résoudre ce problème le modèle suivant:

$$\begin{aligned} NegPerChick_{ij} = & \alpha + \beta_1 \times SexParent_{ij} + \beta_2 \times Foodtreatment_{ij} + \beta_3 \times ArrivalTime_{ij} \\ & + \beta_4 \times Foodtreatment_{ij} \times SexParent_{ij} + \beta_5 \times ArrivalTime_{ij} \times SexParent_{ij} + a_i + \varepsilon_{ij} \end{aligned}$$

Pout tout couple (i, j) dans $[1, \dots, p] \times [1, \dots, n_i]$, où p est le nombre de nids et n_i est le nombre d'observations faites au i -ème nid, on définit:

1. *SexParent* et *FoodTreatment* les variables nominatives comprenant deux modalités, et *ArrivalTime* la variable continue.
2. a_i l'identificateur du nid (son nom), de plus la variable a_i suit une loi normale centrée de variance d^2 .
3. Le résidu ε_{ij} suit une loi normale centrée de variance σ^2 .

De plus, les variables a_i et ε_{ij} sont mutuellement indépendantes.

On peut écrire (4.0) sous forme matricielle

$$\begin{pmatrix} NegPerChick_{11} \\ \vdots \\ NegPerChick_{1n_1} \\ \vdots \\ NegPerChick_{i1} \\ \vdots \\ NegPerChick_{in_i} \\ \vdots \\ NegPerChick_{p1} \\ \vdots \\ NegPerChick_{pn_p} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{n_1} & SP_{11} & Ft_{11} & AT_{11} & AT.SP_{11} & Ft.SP_{11} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{1}_{n_i} & SP_{i1} & Ft_{i1} & AT_{i1} & AT.SP_{i1} & Ft.SP_{i1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{1}_{n_p} & SP_{p1} & Ft_{p1} & AT_{p1} & AT.SP_{p1} & Ft.SP_{p1} \end{pmatrix} \cdot \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} + \begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_p \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{in_i} \\ \vdots \\ \varepsilon_{pn_p} \end{pmatrix} \quad (5.1)$$

On peut réécrire (3.1) comme

$$\begin{pmatrix} N^{(1)} \\ \vdots \\ N^{(i)} \\ \vdots \\ N^{(p)} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{n_1} & SP^{(1)} & Ft^{(1)} & AT^{(1)} & Ft.SP^{(1)} & AT.SP^{(1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{1}_{n_i} & SP^{(i)} & Ft^{(i)} & AT^{(i)} & Ft.SP^{(i)} & AT.SP^{(i)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{1}_{n_p} & SP^{(p)} & Ft^{(p)} & AT^{(p)} & Ft.SP^{(p)} & AT.SP^{(p)} \end{pmatrix} \cdot \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} + \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(i)} \\ \vdots \\ A^{(p)} \end{pmatrix} + \begin{pmatrix} E^{(1)} \\ \vdots \\ E^{(i)} \\ \vdots \\ E^{(p)} \end{pmatrix} \quad (5.2)$$

ou encore

$$\boxed{N = X\beta + A + E} \quad (5.3)$$

On pose $n = \sum_{i=1}^p n_i$,

Le vecteur aléatoire $A = \left(A^{(1)} | \dots | A^{(p)} \right)' \in \mathbb{R}^n$ est formé de vecteurs aléatoires $A^{(i)} \hookrightarrow \mathcal{N}(0, d^2 I_{n_i})$ qui sont mutuellement indépendants. Ce qui revient à dire que A est distribué normalement de moyenne 0 et de matrice de covariance $D = d^2 I_n$.

Le vecteur aléatoire des résidus $E = \left(E^{(1)} | \dots | E^{(p)} \right)' \in \mathbb{R}^n$ est formé de vecteurs aléatoires $E^{(i)} \hookrightarrow \mathcal{N}(0, \sigma^2 I_{n_i})$ qui sont mutuellement indépendants. Ce qui revient à dire que E est distribué normalement de moyenne 0 et de matrice de covariance $\Sigma = \sigma^2 I_n$.

Comme le modèle (1.5) n'est rien d'autre qu'un modèle linéaire mixte gaussien unidimensionnel à un facteur aléatoire plus une résiduelle, alors la proposition 1.2.1 assure que N est normalement distribué de moyenne $\mu = X\beta$ et de matrice de covariance $V = Z_1 d^2 Z_1' + Z_2 \sigma^2 Z_2'$, et puisque $Z_1 = Z_2 = I_n$ alors $V = (d^2 + \sigma^2) I_n$.

5.1.3 Description des données

La distribution de la variable à expliquer $NegPerChick_{ij}$ étant asymétrique (de nombreuses faibles valeurs et quelques grandes valeurs) alors il est utile d'effectuer une transformation des données pour donner moins de poids aux variations des grandes valeurs. Soit donc la nouvelle variable à expliquer $LogNeg_{ij} = \log(NegPerChick_{ij} + 1)$ pour la j -ième observation au nid i .

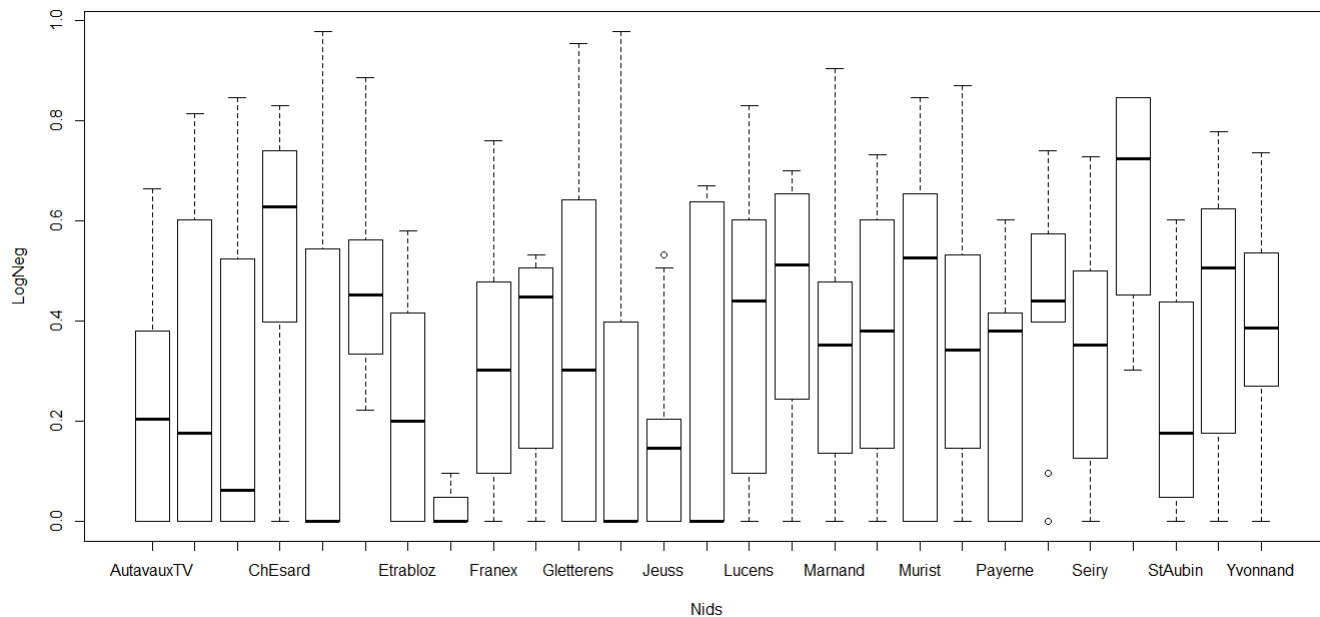


Figure 1- Répartition de LogNeg en fonctions du nid

Dans la plupart des cas, on observe une dispersion intra-nid relativement élevée (les box-plot par nid sont assez larges), par contre la variation inter-nid est faible (exception faite pour certains nids).

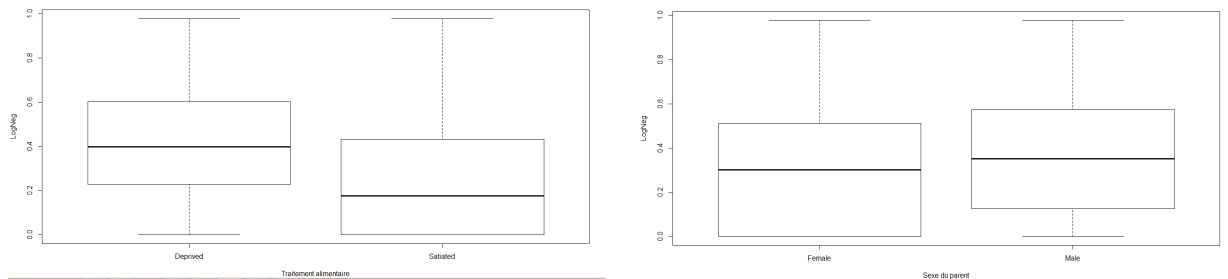


Figure 2- Répartition de LogNeg en fonction du traitement alimentaire (Deprived | Satiated), puis en fonction du sexe des parents (Female | Male)

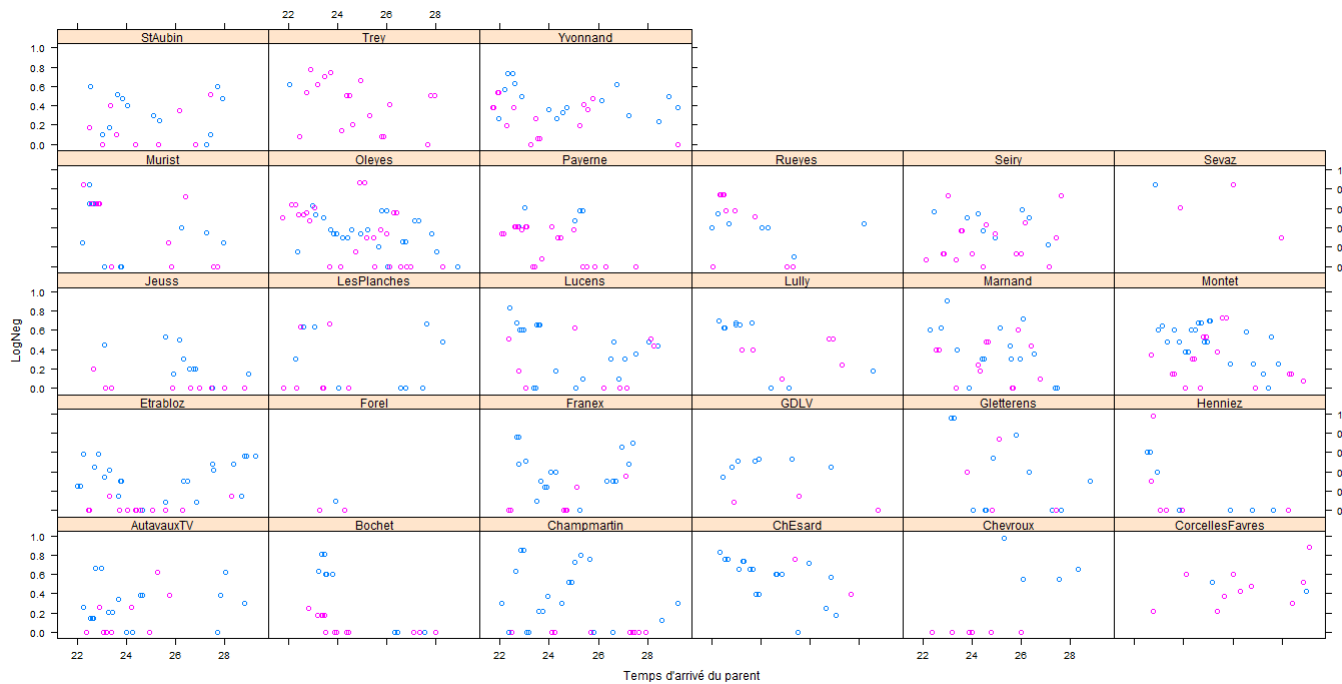


Figure 3- Répartition de LogNeg en fonction du temps d'arrivée du parent, par traitement alimentaire (Satiated en mangenta et Deprived en bleu) et par nid

On constate que dans la majorité des nids le traitement alimentaire des oisillons a été varié. Cependant dans certains nids les oisillons ont recus un traitement alimentaire exclusif, c'est à dire soit ils ont été totalement privés de nourriture, soit ils ont été absolument rassasiés.

5.2 Illustration de la régression linéaire mixte avec la méthode ML

5.2.1 Selection des variables

Etape 1 : Modèle complet

```
> form = formula( LogNeg ~ FoodTreatment + SexParent + ArrivalTime +
FoodTreatment*SexParent + ArrivalTime*SexParent )
> res.ml = lme( form, random= ~ 1 | Nest, method="ML", data=Owls )
> summary( res.ml )
```

```
Fixed effects: list(form)

              Value Std.Error DF   t-value p-value
(Intercept)    1.1234836 0.19533093 567   5.751693  0.0000
FoodTreatmentSatiated -0.1815306 0.03064658 567  -5.923355  0.0000
SexParentMale     0.1071656 0.25477617 567   0.420627  0.6742
ArrivalTime      -0.0290209 0.00782461 567  -3.708922  0.0002
FoodTreatmentSatiated:SexParentMale 0.0139114 0.03973709 567   0.350086  0.7264
SexParentMale:ArrivalTime -0.0037772 0.01020614 567  -0.370094  0.7115
```

On remarque que les variables SexParentMale, Foodtreatment:Sex et ArrivalTime:Sex ne sont pas significatives. On va comparer le modèle complet avec d'abord le sous-modèle **a)** sans la variable Foodtreatment:Sex, puis avec le sous-modèle **b)** sans la variable ArrivalTime:Sex.

```
> res.mla = update( res.ml, .~. - FoodTreatment:SexParent )
> res.mlb = update( res.ml, .~. - SexParent:ArrivalTime )
> anova( res.ml,res.mla )
> anova( res.ml,res.mlb )
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
res.ml	1	8	-0.7484292	34.41366	8.374215			
res.mla	2	7	-2.6246932	28.14214	8.312347	1 vs 2	0.123736	0.725

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
res.ml	1	8	-0.7484292	34.41366	8.374215			
res.mlb	2	7	-2.6103305	28.15650	8.305165	1 vs 2	0.1380986	0.7102

On accepte (**H0**) dans les deux tests car les *p-values* des tests sont supérieurs à 0.05. Et ensuite on valide l'utilisation du sous-modèle **a**) car la valeur du critère d'AIC de celui-ci est la plus faible.

Etape 2 : Modèle ajusté à 4 variables explicatives

```
> form = formula( LogNeg~FoodTreatment + SexParent + ArrivalTime +
ArrivalTime*SexParent )
> res.ml = lme( form, random= ~ 1 | Nest, method="ML", data=Owls )
> summary( res.ml )
```

```
Fixed effects: list(form)
              Value Std.Error DF   t-value p-value
(Intercept)    1.1181384 0.19459620 568    5.745942  0.0000
FoodTreatmentSatiated -0.1734020 0.02003822 568   -8.653565  0.0000
SexParentMale     0.1171775 0.25297131 568    0.463205  0.6434
ArrivalTime      -0.0289695 0.00781773 568   -3.705613  0.0002
SexParentMale:ArrivalTime -0.0039054 0.01019215 568   -0.383176  0.7017
```

On remarque que les variables SexParentMale, ArrivalTime:Sexe ne sont pas significatives, on va comparer le modèle ajusté avec d'abord le sous-modèle **a**) sans la variable Sex, puis avec le sous-modèle **b**) sans la variable ArrivalTime:Sex.

```
> res.mla = update( res.ml, .~. - SexParent )
> res.mlb = update( res.ml, .~. - SexParent:ArrivalTime )
> anova( res.ml,res.mla )
> anova( res.ml,res.mlb )
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
res.ml	1	7	-2.624693	28.14214	8.312347			
res.mla	2	6	-4.408607	21.96296	8.204303	1 vs 2	0.2160864	0.642

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
res.ml	1	7	-2.624693	28.14214	8.312347			
res.mlb	2	6	-4.476920	21.89465	8.238460	1 vs 2	0.1477732	0.7007

On accepte (**H0**) dans les deux tests car les *p-values* des tests sont supérieurs à 0.05. Et ensuite on valide l'utilisation du sous-modèle **b**) car la valeur du critère d'AIC de celui-ci

est la plus faible.

Etape 3 : Modèle ajusté à 3 variables explicatives

```
> form = formula( LogNeg ~ FoodTreatment + ArrivalTime + SexParent )
> res.ml = lme( form, random= ~ 1 | Nest, method="ML", data=Owls )
> summary( res.ml )
```

```
Fixed effects: list(form)
              Value Std.Error DF   t-value p-value
(Intercept)    1.1737693 0.12916231 569   9.087553   0.000
FoodTreatmentSatiated -0.1732631 0.02002247 569  -8.653433   0.000
ArrivalTime     -0.0312285 0.00512391 569  -6.094661   0.000
SexParentMale     0.0206291 0.02129175 569   0.968877   0.333
```

On remarque que la variables SexParentMale n'est pas significative, on la retire du modèle.

Etape 4 : Modèle selectionné

```
> form = formula( LogNeg ~ FoodTreatment + ArrivalTime )
> res.ml = lme( form, random= ~ 1 | Nest, method="ML", data=Owls )
> summary( res.ml )
```

```
Fixed effects: list(form)
              Value Std.Error DF   t-value p-value
(Intercept)    1.1813102 0.12896296 570   9.160074     0
FoodTreatmentSatiated -0.1747941 0.01997134 570  -8.752245     0
ArrivalTime     -0.0309974 0.00511511 570  -6.059964     0
```

Toutes les variables sont très significatives, le traitement alimentaire et le temps d'arrivée du parent dans le nid ont une influence sur la variable LogNeg.

5.2.2 Interprétation des résultats

```
> form = formula( LogNeg ~ FoodTreatment + ArrivalTime )
> res.ml = lme( form, random = ~ 1 | Nest, method="ML" , data=Owls )
> summary( res.ml )
```

```
Linear mixed-effects model fit by maximum likelihood
Data: Owls
      AIC      BIC    logLik
-5.545145 16.43116 7.772572

Random effects:
Formula: ~1 | Nest
      (Intercept)  Residual
StdDev:  0.09156867 0.2313028

Fixed effects: list(form)
              Value Std.Error DF   t-value p-value
(Intercept)    1.1813102 0.12896296 570   9.160074      0
FoodTreatmentSatiated -0.1747941 0.01997134 570  -8.752245      0
ArrivalTime     -0.0309974 0.00511511 570  -6.059964      0
Correlation:
              (Intr) FdTrtS
FoodTreatmentSatiated -0.112
ArrivalTime          -0.984  0.039

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-2.21779287 -0.78414613 -0.07500037  0.68936409  3.28871880

Number of Observations: 599
Number of Groups: 27
```

Avant d'aller plus loin, vérifions que nos hypothèses sont raisonnables en examinant la droite de Henry des résidus et celle pour les effets aléatoires.

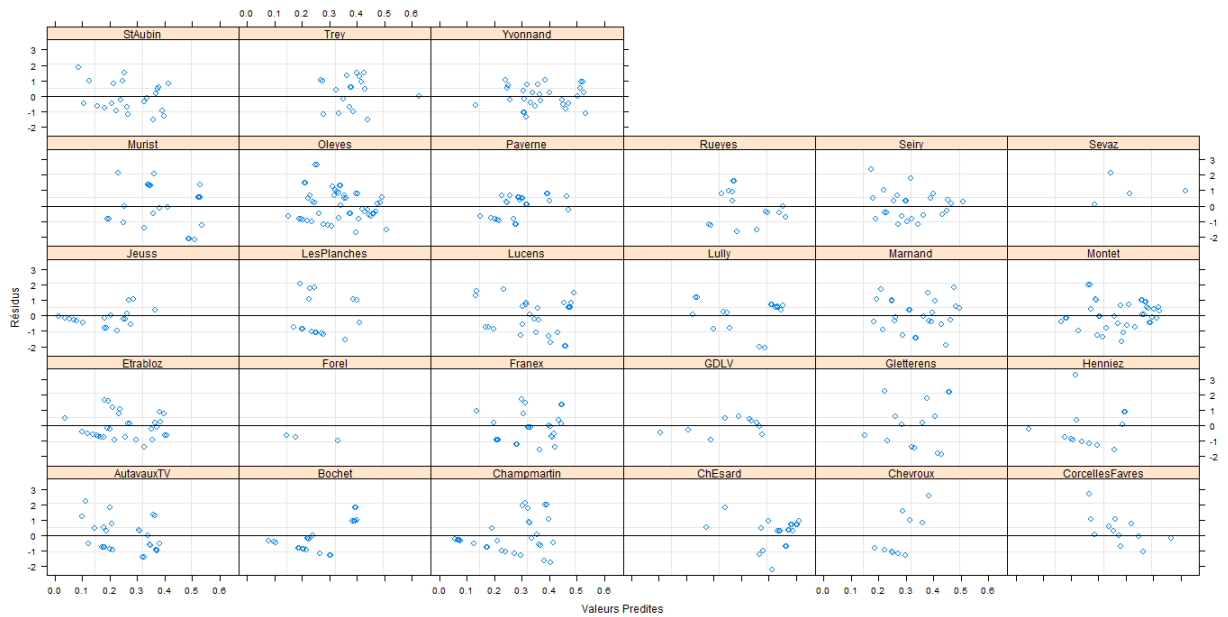


Figure 4- Les résidus en fonction des valeurs prédites par nid

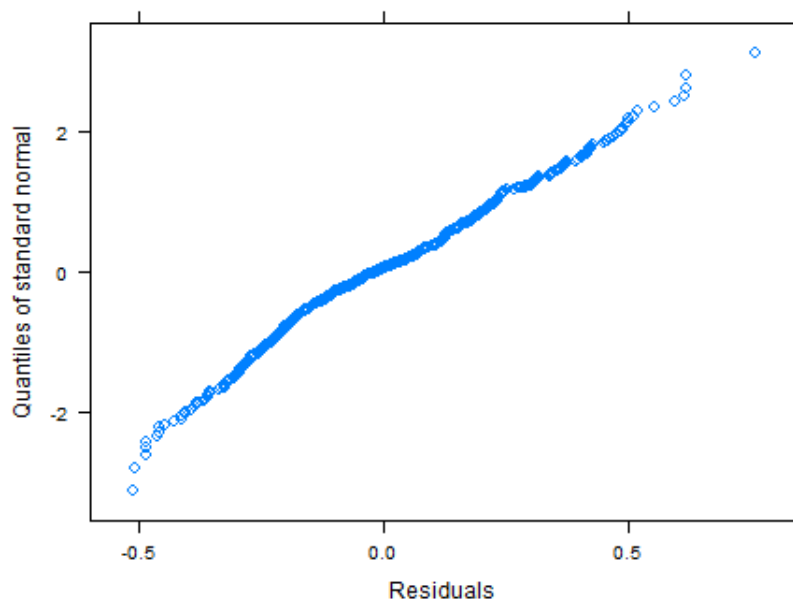


Figure 5- Droite de Henry des résidus

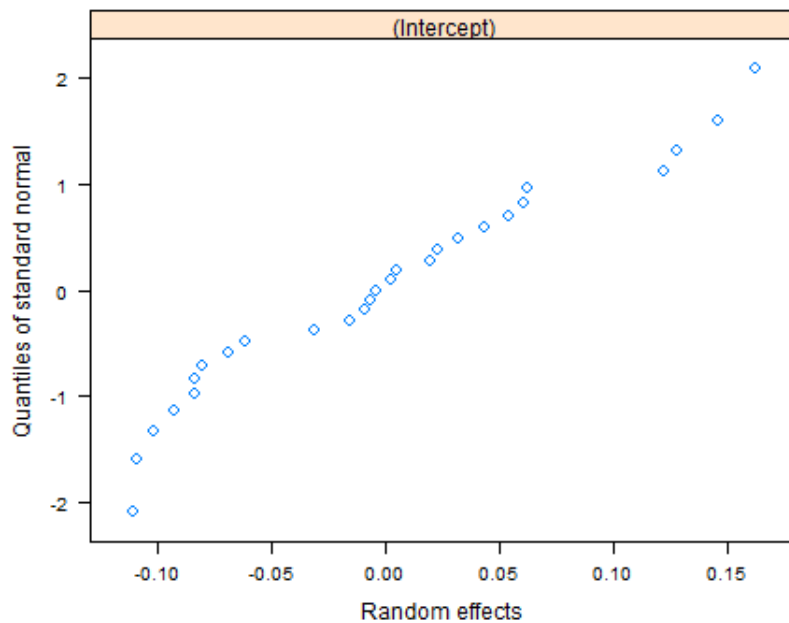


Figure 6- Droite de Henry des effets aléatoires

Les résidus semblent bien centrés sur zéro et ils se situent dans une bande pour chaque nid (voir Fig. 4). Mis à part les effets de bords inévitables, les droites de Henry (Fig. 5 et 6) montrent des points globalement alignés. Les hypothèses de normalité semblent donc raisonnables. Les hypothèses que nous avons faites semblent donc réalistes.

Intervalles de confiance approximatifs pour les paramètres du modèle linéaire à effets mixtes

On utilise une approximation normale de la distribution des estimateurs du maximum de vraisemblance pour construire les intervalles de confiance approximatifs.

```
> intervals(res.ml)
```

Approximate 95% confidence intervals

Fixed effects:

	lower	est.	upper
(Intercept)	0.92864470	1.18131019	1.4339757
FoodTreatmentSatiated	-0.21392215	-0.17479409	-0.1356660
ArrivalTime	-0.04101894	-0.03099737	-0.0209758

```
attr(,"label")
[1] "Fixed effects:"
```

Random Effects:

Level: Nest

	lower	est.	upper
sd((Intercept))	0.06092036	0.09156867	0.1376358

Within-group standard error:

	lower	est.	upper
	0.2182027	0.2313028	0.2451895

5.3 Illustration de la régression linéaire mixte avec la méthode REML

```
> form = formula( LogNeg ~ FoodTreatment + ArrivalTime )
> res.ml = lme( form, random = ~ 1 | Nest, method="ML" , data=Owls )
> summary( res.ml )
```

```
Linear mixed-effects model fit by REML
Data: Owls
      AIC      BIC    logLik
15.07383 37.02503 -2.536915

Random effects:
Formula: ~1 | Nest
      (Intercept)  Residual
StdDev:  0.09468769 0.2316398

Fixed effects: list(form)
              Value Std.Error DF   t-value p-value
(Intercept)    1.1821386 0.12897491 570   9.165648     0
FoodTreatmentSatiated -0.1750754 0.01996606 570  -8.768650     0
ArrivalTime      -0.0310214 0.00511232 570  -6.067954     0
Correlation:
              (Intr) FdTrtS
FoodTreatmentSatiated -0.112
ArrivalTime          -0.984  0.039

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-2.22283609 -0.78307304 -0.07461892  0.68690000  3.29183331

Number of Observations: 599
Number of Groups: 27
```

Les résultats obtenus sont très proches de ceux que nous avons trouvés lors de l'utilisation de la méthode ML. Une interprétation complémentaire est proposée p. 138-139 du livre "Mixed Effects Models and Extensions in Ecology with R" (réf Alain F.Zuur,...^[1]).



Bibliographie

- [1] Alain F.Zuur, Elena N.Ieno, Neil J.Walker, Anatoly A.Saveliev, Graham M.Smith, *Mixed Effects and Extensions in Ecology with R*. Springer, 129-142, 2009.
- [2] Laird, N.M and Ware,J.H, *Random effects models for longitudinal data*. Biometrics, Vol. 38, No. 4, 963-974, 1982.
- [3] Laurent DAVEZIES, *Modèles à effets fixes, à effets aléatoires, modèles mixtes ou multi-niveaux : propriétés et mises en oeuvre des modélisations de l'hétérogénéité dans le cas de données groupées*. 60-70, G 2011 / 03. <http://www.crest.fr/ckfinder/userfiles/files/Pageperso/ldavezies/WorkingPaperINSEE/G2011-03.pdf>
- [4] Geert Verbeke, Geert Molenberghs *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics, QA279, V458, 2000, 42-46, 2009.
- [5] Henderson C.R, Kempthorne O, Searle S.R, von Krosigk C.M, *The estimation of environmental and genetic trends from records subject to culling*. Biometrics, Vol. 15, No. 2. 192-218, 1959.
- [6] Jean-Louis Foulley, *Le modèle linéaire mixte*. 35-40, 2003. http://pbil.univ-lyon1.fr/members/fpicard/franckpicard_fichiers/pdf/cours.foulley.pdf
- [7] V. Monbet, *Notes de cours Statistique inférentielle : Tests Master Statistique et Économétrie*. Univ. Rennes 1, Master 1, 20-21 2013.
- [8] Z. Griliches and M.D. Intriligator, *Handbook of Econometrics, Volume II*. Elsevier Science Publishers BV, Chapter 13, 773-785, 1984.