Section 3

Modèles linéaires mixtes et modèles linéaires mixtes généralisés

Séance du 29 janvier 2027

Lectures : Wu (2010), ch. 2 (§ 2.2, 2.4, 2.6 et 2.7) ; compléments : Jiang et Nguyen (2021), § 3.5 ; Fitzmaurice et coll. (2011), ch. 15

Julien Miron – notes rédigées à partir du matériel de Lajmi Lakhal-Chaieb et Alexandre Bureau

Objectifs de la séance. À la fin de cette séance, vous devriez pouvoir :

1. écrire un modèle linéaire mixte (LMM) sous forme matricielle, en déduire la structure de covariance marginale et estimer ses paramètres par moindres carrés généralisés, maximum de vraisemblance (MV) et maximum de vraisemblance restreint (REML) ;
2. démontrer la formule de prédiction des effets aléatoires (BLUP, estimateur bayésien empirique) ;
3. définir un modèle linéaire mixte généralisé (GLMM), écrire sa vraisemblance marginale et expliquer les trois grandes familles de méthodes de calcul (quadrature de Gauss-Hermite, approximation de Laplace, quasi-vraisemblance pénalisée) ainsi que leurs limites ;
4. distinguer un effet individuel (conditionnel) d'un effet marginal (moyenné sur la population) et quantifier l'atténuation avec le lien logit ;
5. sélectionner les effets aléatoires (graphiques, test du rapport de vraisemblance à la frontière) et les covariables d'un modèle mixte ;
6. ajuster et interpréter des GLMM en R avec `lme4`.

3.1 Mise en contexte : la pleuropneumonie contagieuse bovine

La pleuropneumonie contagieuse bovine est une maladie respiratoire du bétail causée par un mycoplasme. Dans une étude de suivi menée en Éthiopie, on a prélevé environ tous les trois mois un échantillon sanguin dans chacun de 15 troupeaux de zébus et compté les nouveaux cas sérologiques (données `cbpp` du module `lme4`). Pour le troupeau i et la période j ($j = 1, \ldots, 4$), on observe le nombre de nouveaux cas y_{ij} parmi m_{ij} animaux.

La figure 1 suggère trois constats.

— Le niveau d'incidence diffère beaucoup d'un troupeau à l'autre : les observations d'un même

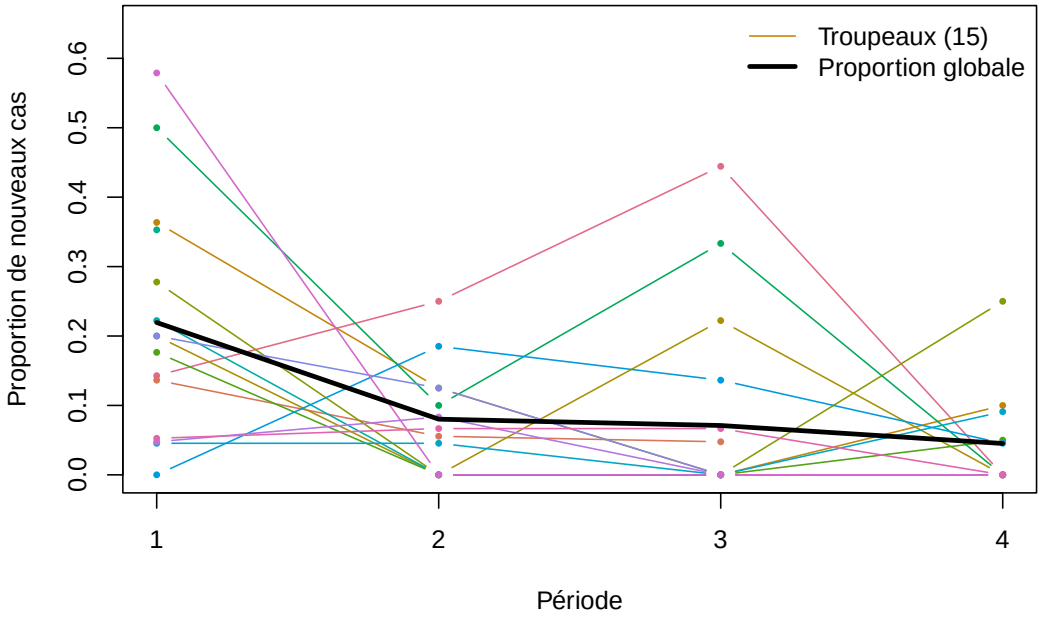


FIGURE 1 – Graphique « spaghetti » de la proportion de nouveaux cas par période pour les 15 troupeaux (données `cbpp`) ; trait épais : proportion globale par période.

troupeau sont corrélées, et un modèle linéaire généralisé (GLM) ordinaire sous-estimerait l'incertitude.

— L'évolution dans le temps varie elle aussi d'un troupeau à l'autre, mais moins nettement.

— Un troupeau n'est observé qu'une fois et un autre trois fois : des troupeaux ont été perdus au suivi. Un modèle fondé sur la vraisemblance utilise toutes les observations disponibles ; nous verrons aux séances 8 et 9 sous quelles conditions sur le mécanisme de perte au suivi cela donne des estimations sans biais.

Le modèle linéaire mixte généralisé répond aux deux premiers constats en ajoutant au prédicteur linéaire des *effets aléatoires* propres à chaque troupeau. Avant de le définir, rappelons le cas gaussien, où tous les calculs sont explicites.

3.2 Le modèle linéaire mixte : rappels

3.2.1 Formulation

On dispose de n grappes indépendantes (sujets, troupeaux, cliniques). La grappe i contient n_i observations, $N = \sum_{i=1}^{n} n_i$.

Définition 3.2.1 (Modèle linéaire mixte ; Wu, § 2.2)**.** Pour $i = 1, \dots, n$,

$$y_i = X_i\beta + Z_i b_i + e_i, \qquad b_i \sim \mathcal{N}_q(0, D), \qquad e_i \sim \mathcal{N}_{n_i}(0, R_i), \tag{1}$$

où $y_i \in \mathbb{R}^{n_i}$ est le vecteur réponse, X_i $(n_i \times p)$ et Z_i $(n_i \times q)$ sont des matrices de plan connues, $\beta \in \mathbb{R}^p$ est le vecteur des effets fixes et $b_i \in \mathbb{R}^q$ celui des effets aléatoires. Les b_i et les e_i sont

mutuellement indépendants. Les matrices D et R_i dépendent d'un vecteur de paramètres de variance η ; le plus souvent $R_i = \sigma^2 I_{n_i}$.

Les colonnes de Z_i sont en général un sous-ensemble des colonnes de X_i : l'effet fixe représente l'effet moyen dans la population et l'effet aléatoire, l'écart de la grappe i à cet effet moyen.

Proposition 3.2.2 (Loi marginale)**.** *Sous* (1), $y_i \sim \mathcal{N}_{n_i}(X_i\beta, V_i)$ *avec*

$$V_i = V_i(\eta) = Z_i D Z_i^\top + R_i. \tag{2}$$

Démonstration. y_i est une transformation affine du vecteur gaussien (b_i, e_i), donc gaussien. $\mathrm{E}(y_i) = X_i\beta$ et, par indépendance de b_i et e_i, $\mathrm{Var}(y_i) = Z_i \,\mathrm{Var}(b_i) Z_i^\top + \mathrm{Var}(e_i) = Z_i D Z_i^\top + R_i$. □

Exemple 3.2.3 (Ordonnée à l'origine aléatoire)**.** Avec $Z_i = \mathbf{1}_{n_i}$, $D = \sigma_b^2$ et $R_i = \sigma^2 I$, on obtient $V_i = \sigma^2 I_{n_i} + \sigma_b^2 \mathbf{1}\mathbf{1}^\top$: variance $\sigma^2 + \sigma_b^2$ sur la diagonale et covariance σ_b^2 entre deux observations de la même grappe. C'est la structure *échangeable* (symétrie composée), de corrélation intra-grappe $\rho = \sigma_b^2/(\sigma_b^2 + \sigma^2)$.

Le modèle (1) a donc deux lectures : une lecture *conditionnelle* ($y_i \mid b_i \sim \mathcal{N}(X_i\beta + Z_i b_i, R_i)$, interprétation propre à la grappe) et une lecture *marginale* ($y_i \sim \mathcal{N}(X_i\beta, V_i)$, interprétation dans la population). Avec le lien identité, β a la même interprétation dans les deux lectures ; ce ne sera plus le cas pour un GLMM (section 3.5).

3.2.2 Estimation de β : moindres carrés généralisés

En empilant les grappes, on écrit $y = X\beta + Zb + e$ avec $V = \mathrm{diag}(V_1, \ldots, V_n)$. La log-vraisemblance marginale est

$$\ell(\beta, \eta) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\Big\{\log|V_i| + (y_i - X_i\beta)^\top V_i^{-1}(y_i - X_i\beta)\Big\}. \tag{3}$$

Proposition 3.2.4 (Moindres carrés généralisés)**.** *Pour η fixé, $\ell(\beta, \eta)$ est maximisée en*

$$\hat\beta(\eta) = \Big(\sum_{i=1}^{n} X_i^\top V_i^{-1} X_i\Big)^{-1} \sum_{i=1}^{n} X_i^\top V_i^{-1} y_i = (X^\top V^{-1} X)^{-1} X^\top V^{-1} y, \tag{4}$$

et $\hat\beta(\eta) \sim \mathcal{N}_p(\beta, (X^\top V^{-1} X)^{-1})$. Lorsque η est connu, $\hat\beta(\eta)$ est le meilleur estimateur linéaire sans biais de β (théorème d'Aitken).

Démonstration. Le gradient de (3) en β est $X^\top V^{-1}(y - X\beta)$; il s'annule en $\hat\beta(\eta)$ et la hessienne $-X^\top V^{-1} X$ est définie négative si X est de plein rang. Comme $\hat\beta(\eta) = Gy$ avec $G = (X^\top V^{-1} X)^{-1} X^\top V^{-1}$, il est gaussien, d'espérance $GX\beta = \beta$ et de variance $GVG^\top = (X^\top V^{-1} X)^{-1}$. L'optimalité s'obtient en appliquant le théorème de Gauss-Markov au modèle transformé $V^{-1/2} y = V^{-1/2} X\beta + \varepsilon$, $\mathrm{Var}(\varepsilon) = I$. □

En pratique, η est inconnu : on le remplace par un estimateur $\hat\eta$ et on utilise $\hat\beta = \hat\beta(\hat\eta)$, avec la variance estimée $(X^\top V(\hat\eta)^{-1} X)^{-1}$, qui ignore l'incertitude sur $\hat\eta$ (d'où les corrections de degrés de liberté de type Kenward-Roger dans les petits échantillons).

3.2.3 Maximum de vraisemblance et maximum de vraisemblance restreint

MV. En substituant (4) dans (3), on obtient la log-vraisemblance profilée $\ell_P(\eta) = \ell(\hat{\beta}(\eta), \eta)$, que l'on maximise numériquement (Newton-Raphson, algorithme EM). L'estimateur du MV de η est biaisé vers le bas en échantillon fini parce qu'il traite β comme connu et égal à $\hat{\beta}$, sans « payer » les p degrés de liberté utilisés pour l'estimer. L'exemple le plus simple est la régression linéaire ordinaire ($V = \sigma^2 I$), où $\hat{\sigma}^2_{\text{MV}} = \text{SCR}/N$ alors que l'estimateur sans biais est $\text{SCR}/(N-p)$.

REML. L'idée du maximum de vraisemblance restreint (Wu, § 2.2) est de baser l'estimation de η sur des combinaisons linéaires des données dont la loi ne dépend pas de β : les *contrastes d'erreur*. Soit A une matrice $N \times (N-p)$ de rang $N-p$ telle que $A^\top X = 0$ (par exemple, $N-p$ colonnes linéairement indépendantes de $I - X(X^\top X)^{-1}X^\top$), et $z = A^\top y$. Alors $z \sim \mathcal{N}_{N-p}(0, A^\top V A)$ ne dépend pas de β : la log-vraisemblance de z, appelée log-vraisemblance restreinte, ne dépend que de η.

Le résultat suivant, repris du document « MVR » du cours H26, relie cette vraisemblance à celle de y.

Lemme 3.2.5. *Soit $G = V^{-1}X(X^\top V^{-1}X)^{-1}$, de sorte que $\hat{\beta} = G^\top y$. Alors*

(a) $z = A^\top y$ et $\hat{\beta}$ sont indépendants ;

(b) la matrice $T = [A \;\; G]$ $(N \times N)$ est inversible et $|\det T|$ ne dépend ni de β ni de η.

Démonstration. (a) Le vecteur $(z, \hat{\beta}) = T^\top y$ est gaussien et

$$\text{Cov}(A^\top y, G^\top y) = A^\top V G = A^\top X(X^\top V^{-1}X)^{-1} = 0.$$

Or des sous-vecteurs conjointement gaussiens et non corrélés sont indépendants.

(b) Comme $A^\top X = 0$ et $\text{rang}(A) = N-p$, les colonnes de A engendrent l'orthogonal de l'espace des colonnes de X, donc $[A \;\; X]$ est inversible. Puisque $G^\top X = I_p$,

$$T^\top [A \;\; X] = \begin{pmatrix} A^\top A & A^\top X \\ G^\top A & G^\top X \end{pmatrix} = \begin{pmatrix} A^\top A & 0 \\ G^\top A & I_p \end{pmatrix},$$

matrice triangulaire par blocs de déterminant $\det(A^\top A) \neq 0$. Donc $|\det T| = \det(A^\top A)/|\det[A \;\; X]|$, qui ne dépend que de A et X. □

Théorème 3.2.6 (Log-vraisemblance restreinte)**.** *Notons f_R la densité de $z = A^\top y$. À une constante près (ne dépendant que de A, X et N),*

$$\begin{aligned} \log f_R(z;\eta) &= \ell(\hat{\beta}(\eta), \eta) - \frac{1}{2}\log|X^\top V^{-1}X| + K' \\ &= -\frac{1}{2}\log|V| - \frac{1}{2}\log|X^\top V^{-1}X| - \frac{1}{2}(y - X\hat{\beta})^\top V^{-1}(y - X\hat{\beta}) + K. \end{aligned} \tag{5}$$

La log-vraisemblance restreinte est donc la log-vraisemblance profilée de y évaluée en $\beta = \hat{\beta}(\eta)$, *corrigée par le terme $-\frac{1}{2}\log|X^\top V^{-1}X|$.*

Démonstration. Notons g la densité de $\hat\beta \sim \mathcal{N}_p(\beta, (X^\top V^{-1} X)^{-1})$. Par le changement de variable $y \mapsto T^\top y = (z, \hat\beta)$ et le lemme 3.2.5,

$$f_y(y) = f_{z,\hat\beta}(A^\top y, G^\top y)\,|\det T| = f_R(z)\, g(\hat\beta)\,|\det T|,$$

d'où $\log f_R(z) = \log f_y(y) - \log g(\hat\beta) - \log|\det T|$. Or $\log f_y(y) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|V| - \frac{1}{2}(y - X\beta)^\top V^{-1}(y - X\beta)$ et

$$\log g(\hat\beta) = -\frac{p}{2}\log 2\pi + \frac{1}{2}\log|X^\top V^{-1} X| - \frac{1}{2}(\hat\beta - \beta)^\top X^\top V^{-1} X(\hat\beta - \beta).$$

En écrivant $y - X\beta = (y - X\hat\beta) + X(\hat\beta - \beta)$ et en utilisant les équations normales $X^\top V^{-1}(y - X\hat\beta) = 0$, le terme croisé s'annule :

$$(y - X\beta)^\top V^{-1}(y - X\beta) = (y - X\hat\beta)^\top V^{-1}(y - X\hat\beta) + (\hat\beta - \beta)^\top X^\top V^{-1} X(\hat\beta - \beta).$$

Les formes quadratiques en $\hat\beta - \beta$ se compensent exactement, ce qui confirme que $\log f_R$ ne dépend pas de β, et l'on obtient (5) avec $K = -\frac{N-p}{2}\log 2\pi - \log|\det T|$. □

Remarque 3.2.7.

(i) La constante K dépend du choix de A, mais pas de η : l'estimateur REML ne dépend pas du choix des contrastes.

(ii) Dans la régression linéaire ordinaire, maximiser (5) en σ^2 donne $\mathrm{SCR}/(N-p)$ (exercice 3.9.3).

(iii) Une autre lecture (Harville, 1974) : (5) est, à une constante près, le log de $\int L(\beta, \eta)\,\mathrm{d}\beta$, la vraisemblance intégrée sur β avec une loi a priori uniforme.

(iv) Les vraisemblances REML de deux modèles ayant des effets fixes *différents* ne sont pas comparables (les contrastes A diffèrent) : pour comparer des effets fixes, on utilise le MV.

Correction au document MVR (H26)

Dans la dernière égalité du document H26, le terme $\frac{1}{2}(y - X\hat\beta)' V^{-1}(y - X\hat\beta)$ doit être précédé d'un signe moins, comme dans (5). Par ailleurs, l'affirmation « $f_{z,\hat\beta}$ est proportionnelle à $f(y)$ » requiert que le jacobien de la transformation ne dépende pas de V : c'est l'objet du lemme 3.2.5(b).

3.2.4 Prédiction des effets aléatoires

Les effets aléatoires ne sont pas des paramètres, mais des variables aléatoires non observées : on les *prédit*. Le prédicteur naturel est l'espérance conditionnelle $\mathrm{E}(b_i \mid y_i)$, qui minimise l'erreur quadratique moyenne de prédiction.

Proposition 3.2.8 (BLUP ; Wu, éq. (2.13))**.** *Sous* (1), $\mathrm{Cov}(b_i, y_i) = DZ_i^\top$ *et*

$$\mathrm{E}(b_i \mid y_i) = DZ_i^\top V_i^{-1}(y_i - X_i\beta), \qquad \mathrm{Var}(b_i \mid y_i) = D - DZ_i^\top V_i^{-1} Z_i D. \tag{6}$$

En remplaçant (β, η) *par leurs estimations, on obtient l'estimateur bayésien empirique*

$$\hat b_i = \hat D Z_i^\top \hat V_i^{-1}(y_i - X_i\hat\beta).$$

Démonstration. D'abord, par bilinéarité et indépendance de b_i et e_i, $\operatorname{Cov}(b_i, y_i) = \operatorname{Cov}(b_i, X_i\beta + Z_i b_i + e_i) = \operatorname{Var}(b_i) Z_i^\top + \operatorname{Cov}(b_i, e_i) = D Z_i^\top$. Le vecteur (b_i, y_i) est gaussien :

$$\begin{pmatrix} b_i \\ y_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ X_i\beta \end{pmatrix}, \begin{pmatrix} D & DZ_i^\top \\ Z_i D & V_i \end{pmatrix} \right).$$

La formule de la loi conditionnelle d'un vecteur gaussien, $\operatorname{E}(U \mid W) = \operatorname{E} U + \operatorname{Cov}(U, W) \operatorname{Var}(W)^{-1}(W - \operatorname{E} W)$ et $\operatorname{Var}(U \mid W) = \operatorname{Var} U - \operatorname{Cov}(U,W) \operatorname{Var}(W)^{-1} \operatorname{Cov}(W, U)$, donne (6). □

Lorsque β est remplacé par $\hat\beta(\eta)$ avec η connu, $\hat b_i$ est le *meilleur prédicteur linéaire sans biais* (BLUP) de b_i ; $\hat\beta$ et $\hat b$ sont aussi les solutions des équations du modèle mixte de Henderson. Quand η est estimé, on parle de BLUP empirique ou d'estimateur bayésien empirique : $\hat b_i$ est la moyenne a posteriori de b_i sous la « loi a priori » $\mathcal{N}(0, D)$, avec des hyperparamètres estimés à partir des données.

Exemple 3.2.9 (Rétrécissement)**.** Dans le modèle à ordonnée aléatoire de l'exemple 3.2.3, $V_i \mathbf{1} = (\sigma^2 + n_i\sigma_b^2)\mathbf{1}$, donc $\mathbf{1}^\top V_i^{-1} = \mathbf{1}^\top/(\sigma^2 + n_i\sigma_b^2)$ et

$$\hat b_i = \sigma_b^2\, \mathbf{1}^\top V_i^{-1}(y_i - X_i\beta) = \underbrace{\frac{n_i\sigma_b^2}{\sigma^2 + n_i\sigma_b^2}}_{\in(0,1)} (\bar y_i - \bar x_i^\top \beta),$$

où $\bar y_i$ et $\bar x_i$ sont les moyennes de la grappe. Le prédicteur rétrécit le résidu moyen de la grappe vers 0, d'autant plus fortement que la grappe est petite ou que σ_b^2 est petit par rapport à σ^2. Les grappes peu informatives « empruntent » de l'information à l'ensemble.

À retenir

- LMM : $y_i = X_i\beta + Z_i b_i + e_i$, $V_i = Z_i D Z_i^\top + R_i$; β s'estime par moindres carrés généralisés à η fixé.
- REML = vraisemblance des contrastes $A^\top y$; elle vaut la log-vraisemblance profilée en $\hat\beta(\eta)$ moins $\frac{1}{2}\log|X^\top V^{-1} X|$, et corrige le biais du MV pour les variances.
- BLUP : $\hat b_i = D Z_i^\top V_i^{-1}(y_i - X_i\hat\beta)$, conséquence directe de $\operatorname{Cov}(b_i, y_i) = D Z_i^\top$ et de la loi conditionnelle gaussienne ; il rétrécit vers 0.

3.3 Le modèle linéaire mixte généralisé

3.3.1 Définition

Ce qui distingue un GLMM d'un GLM, c'est uniquement la présence d'effets aléatoires dans le prédicteur linéaire ; les autres ingrédients du GLM (loi de la famille exponentielle, fonction de lien, fonction de variance) demeurent. Rappelons qu'une loi de la famille exponentielle a une densité de la forme

$$f(y; \theta, \phi) = \exp\left\{ \frac{y\theta - \psi(\theta)}{a(\phi)} + c(y, \phi) \right\}, \qquad \operatorname{E} Y = \mu = \psi'(\theta), \qquad \operatorname{Var} Y = a(\phi)\, \psi''(\theta) = a(\phi)\, v(\mu).$$

Définition 3.3.1 (GLMM ; Wu, § 2.4)**.** Un modèle linéaire mixte généralisé repose sur deux éléments.

(1) *Indépendance conditionnelle et loi de la famille exponentielle.* Sachant b_i, les réponses $y_{i1}, \ldots, y_{in_i}$ sont indépendantes, de loi de la famille exponentielle, avec moyenne conditionnelle $\mu_{ij}^b = \mathrm{E}(y_{ij} \mid b_i)$ telle que

$$g(\mu_{ij}^b) = x_{ij}^\top \beta + z_{ij}^\top b_i, \tag{7}$$

où g est une fonction de lien et $\mathrm{Var}(y_{ij} \mid b_i) = a(\phi)\, v(\mu_{ij}^b)$.

(2) *Loi des effets aléatoires.* Les b_i sont indépendants, de loi $\mathcal{N}_q(0, D)$ (par défaut ; d'autres lois sont possibles).

Dans l'exemple `cbpp`, $y_{ij} \mid b_i \sim \text{Binomiale}(m_{ij}, \pi_{ij})$ avec $\log \pi_{ij} = \beta_0 + \beta_j + b_i$ (lien log, risques relatifs) ou $\mathrm{logit}\, \pi_{ij} = \beta_0 + \beta_j + b_i$ (lien logit, rapports de cotes), $b_i \sim \mathcal{N}(0, \sigma_b^2)$.

L'indépendance conditionnelle signifie que toute la dépendance intra-grappe passe par b_i : deux observations d'une même grappe partagent le même b_i et sont donc marginalement corrélées.

3.3.2 La vraisemblance marginale

Les b_i n'étant pas observés, la vraisemblance des paramètres $\theta = (\beta, \phi, D)$ est obtenue en les intégrant (Wu, éq. (2.42)) :

$$L(\theta) = \prod_{i=1}^{n} L_i(\theta), \qquad L_i(\theta) = \int_{\mathbb{R}^q} \prod_{j=1}^{n_i} f(y_{ij} \mid b_i; \beta, \phi)\; \varphi_q(b_i; 0, D)\, \mathrm{d}b_i, \tag{8}$$

où $\varphi_q(\cdot\,; 0, D)$ est la densité normale à q dimensions. Dans le cas gaussien avec lien identité, l'intégrande est l'exponentielle d'une forme quadratique en b_i et l'intégrale donne la densité $\mathcal{N}(X_i\beta, V_i)$ de la proposition 3.2.2. Dans tous les autres cas (binomial, Poisson, …), (8) n'a pas de forme explicite : chaque évaluation de la vraisemblance demande n intégrales de dimension q. On doit alors recourir soit à une méthode de calcul numérique de l'intégrale (quadrature de Gauss-Hermite, intégration Monte Carlo, algorithme EM), soit à une approximation analytique de la vraisemblance (Laplace, PQL). C'est l'objet de la section 3.4.

Le résultat suivant (Wu, p. 310) éclaire le lien avec l'algorithme EM : le score de la vraisemblance marginale est l'espérance a posteriori du score « complet ».

Proposition 3.3.2 (Score de la vraisemblance marginale)**.** *Sous des conditions permettant de dériver sous le signe d'intégrale,*

$$\frac{\partial \log L(\theta)}{\partial \beta} = \sum_{i=1}^{n} \mathrm{E}\left[\frac{\partial}{\partial \beta} \log f(y_i \mid b_i; \beta, \phi) \;\middle|\; y_i; \theta \right],$$

où l'espérance porte sur la loi a posteriori de b_i, $f(b_i \mid y_i; \theta) = f(y_i \mid b_i)\varphi_q(b_i; 0, D)/L_i(\theta)$.

Démonstration. Comme la loi de b_i ne dépend pas de β,

$$\frac{\partial \log L_i}{\partial \beta} = \frac{1}{L_i} \int \frac{\partial f(y_i \mid b_i)}{\partial \beta}\, \varphi_q(b_i)\, \mathrm{d}b_i = \int \frac{\partial \log f(y_i \mid b_i)}{\partial \beta}\, \frac{f(y_i \mid b_i)\varphi_q(b_i)}{L_i}\, \mathrm{d}b_i,$$

et la dernière fraction est la densité a posteriori de b_i. □

Les équations du MV demandent donc, elles aussi, des intégrales sur la loi a posteriori des effets aléatoires ; l'algorithme EM et ses variantes Monte Carlo exploitent directement cette représentation.

3.4 Calcul de la vraisemblance d'un GLMM

Pour alléger, on considère une grappe i et on note

$$h_i(b) = \log f(y_i \mid b; \beta, \phi) + \log \varphi_q(b; 0, D), \qquad L_i = \int_{\mathbb{R}^q} e^{h_i(b)}\, \mathrm{d}b,$$

la log-densité conjointe de (y_i, b_i), ou log-vraisemblance des « données complètes ». Soit $\tilde{b}_i = \arg\max_b h_i(b)$ le *mode conditionnel* (mode a posteriori) de b_i.

3.4.1 Quadrature de Gauss-Hermite

La quadrature de Gauss-Hermite à K points approxime une intégrale contre le poids e^{-x^2} par une somme pondérée :

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)\, \mathrm{d}x \approx \sum_{k=1}^{K} w_k f(x_k),$$

où les nœuds x_k sont les racines du polynôme d'Hermite de degré K et les w_k des poids tabulés. La règle est exacte lorsque f est un polynôme de degré au plus $2K-1$.

Version classique. Pour un effet aléatoire scalaire $b \sim \mathcal{N}(0, \sigma^2)$, le changement de variable $b = \sqrt{2}\,\sigma x$ donne

$$L_i = \int f(y_i \mid b)\, \frac{e^{-b^2/(2\sigma^2)}}{\sqrt{2\pi}\,\sigma}\, \mathrm{d}b = \frac{1}{\sqrt{\pi}} \int e^{-x^2} f(y_i \mid \sqrt{2}\,\sigma x)\, \mathrm{d}x \approx \frac{1}{\sqrt{\pi}} \sum_{k=1}^{K} w_k\, f(y_i \mid \sqrt{2}\,\sigma x_k).$$

Lorsque la loi a posteriori de b_i est concentrée loin de 0 ou beaucoup plus étroite que la loi a priori (grappes de grande taille), la plupart des nœuds tombent là où l'intégrande est négligeable, et il faut beaucoup de points.

Version adaptative. On centre et on met à l'échelle les nœuds selon la loi a posteriori : avec $\hat{\sigma}_i^2 = \{-h_i''(\tilde{b}_i)\}^{-1}$ et $b = \tilde{b}_i + \sqrt{2}\,\hat{\sigma}_i x$,

$$L_i = \sqrt{2}\,\hat{\sigma}_i \int e^{-x^2} \left[e^{x^2} e^{h_i(\tilde{b}_i + \sqrt{2}\hat{\sigma}_i x)} \right] \mathrm{d}x \approx \sqrt{2}\,\hat{\sigma}_i \sum_{k=1}^{K} w_k\, e^{x_k^2}\, e^{h_i(\tilde{b}_i + \sqrt{2}\hat{\sigma}_i x_k)}. \tag{9}$$

La fonction entre crochets est alors presque constante quand la loi a posteriori est proche d'une normale, et quelques points suffisent. C'est la quadrature de Gauss-Hermite adaptative (AGQ), utilisée par `glmer(..., nAGQ = K)` et par SAS `PROC NLMIXED`.

En augmentant K, l'approximation converge vers la valeur exacte de l'intégrale : on qualifie la méthode d'« exacte ». Sa limite est la dimension : une grille produit en dimension q demande K^q évaluations par grappe et par itération. En pratique, on l'utilise avec un, parfois deux effets aléatoires (`glmer` n'accepte `nAGQ > 1` que pour un seul effet aléatoire scalaire).

3.4.2 Approximation de Laplace

Théorème 3.4.1 (Approximation de Laplace ; Wu, éq. (2.78))**.** *Soit $p : \mathbb{R}^q \to \mathbb{R}$ deux fois dérivable, de minimum unique $\hat{v}$ à hessienne $p''(\hat{v}) = \partial^2 p / \partial v \, \partial v^\top |_{v=\hat{v}}$ définie positive. Alors, pour k grand,*

$$\int_{\mathbb{R}^q} e^{-k\,p(v)}\,\mathrm{d}v \approx \left(\frac{2\pi}{k}\right)^{q/2} |p''(\hat{v})|^{-1/2} e^{-k\,p(\hat{v})}, \tag{10}$$

avec une erreur relative $O(k^{-1})$.

Dérivation. Développons p à l'ordre 2 autour de son minimum. Comme $p'(\hat{v}) = 0$,

$$p(v) \approx p(\hat{v}) + \frac{1}{2}(v-\hat{v})^\top p''(\hat{v})(v-\hat{v}).$$

Donc

$$\int e^{-kp(v)}\,\mathrm{d}v \approx e^{-kp(\hat{v})} \int \exp\Big\{-\frac{1}{2}(v-\hat{v})^\top [k\,p''(\hat{v})](v-\hat{v})\Big\}\,\mathrm{d}v = e^{-kp(\hat{v})}\,(2\pi)^{q/2}\,|k\,p''(\hat{v})|^{-1/2},$$

car l'intégrande est, à la constante de normalisation près, la densité $\mathcal{N}_q(\hat{v}, [k\,p''(\hat{v})]^{-1})$. On conclut avec $|k\,p''| = k^q |p''|$. Lorsque k croît, la masse de e^{-kp} se concentre dans un voisinage de $\hat{v}$ de taille $O(k^{-1/2})$, où le développement de Taylor est de plus en plus précis ; les termes d'ordre 3 et 4 donnent l'erreur relative $O(k^{-1})$. □

Application au GLMM. On prend $k\,p(v) = -h_i(v)$ (chez Wu, p est la log-vraisemblance des données complètes changée de signe et divisée par $k = n_i$), d'où $\hat{v} = \tilde{b}_i$ et

$$\log L_i \approx h_i(\tilde{b}_i) + \frac{q}{2}\log(2\pi) - \frac{1}{2}\log\big|-h_i''(\tilde{b}_i)\big|. \tag{11}$$

On maximise ensuite $\sum_i$ (11) en (β, ϕ, D) comme s'il s'agissait de la vraisemblance (en recalculant les modes $\tilde{b}_i$ à chaque itération). Pour un lien canonique ($\theta_{ij} = x_{ij}^\top \beta + z_{ij}^\top b$), on a $\partial \mu_{ij}^b / \partial \theta_{ij} = v(\mu_{ij}^b)$ et

$$-h_i''(b) = \frac{1}{a(\phi)} \sum_{j=1}^{n_i} v(\mu_{ij}^b)\, z_{ij} z_{ij}^\top + D^{-1}. \tag{12}$$

Proposition 3.4.2 (Propriétés de l'approximation de Laplace)**.**

(a) Elle est exacte pour le LMM gaussien (l'intégrande est exactement gaussienne en b).

(b) Elle coïncide avec la quadrature de Gauss-Hermite adaptative à un point.

(c) Elle devient exacte lorsque $D \to 0$ et elle est d'autant meilleure que la variance des effets aléatoires est petite ou que n_i est grand.

Démonstration. (a) Si h_i est quadratique en b, le développement de Taylor d'ordre 2 est exact. (b) Avec $K = 1$, le nœud est $x_1 = 0$ et le poids $w_1 = \sqrt{\pi}$; (9) devient $\sqrt{2\pi}\,\hat{\sigma}_i\, e^{h_i(\tilde{b}_i)}$, c'est-à-dire (11) avec $q = 1$. (c) Quand $D \to 0$, le terme D^{-1} domine (12) et $\tilde{b}_i \to 0$; alors $|-h_i''(\tilde{b}_i)|^{-1/2} \approx |D|^{1/2}$ et $e^{h_i(\tilde{b}_i)} \approx f(y_i \mid 0)\,|2\pi D|^{-1/2}$, de sorte que le membre de droite de l'approximation tend vers $f(y_i \mid 0) = L_i$. Plus généralement, l'approximation suppose la loi a posteriori de b_i concentrée près de son mode, ce qui est le cas quand la loi a priori est étroite (variance petite) ou quand la grappe apporte beaucoup d'information (erreur relative $O(n_i^{-1})$). □

À l'inverse, l'approximation se dégrade pour des réponses binaires avec peu d'observations par grappe et une grande variance des effets aléatoires : la loi a posteriori de b_i est alors asymétrique et étalée.

3.4.3 Quasi-vraisemblance pénalisée (PQL)

La quasi-vraisemblance pénalisée (Breslow et Clayton, 1993 ; Wu, § 2.6) pousse l'approximation de Laplace plus loin, de façon à se ramener à l'ajustement répété d'un LMM. Les approximations successives sont (voir l'encadré « Niveau doctorat » pour les formules) :

(1) remplacer $\log f(y_{ij} \mid b)$ par une quasi-vraisemblance, fonction seulement de la moyenne et de la variance conditionnelles (via la déviance) ;

(2) appliquer l'approximation de Laplace à cette quasi-vraisemblance ;

(3) dans la hessienne, négliger les termes d'espérance nulle (proportionnels à $y_{ij} - \mu^b_{ij}$), ce qui la remplace par $Z_i^\top W_i Z_i + D^{-1}$, où W_i contient les poids itératifs du GLM ;

(4) supposer que W_i varie lentement avec la moyenne, de sorte que le terme $\log|Z_i^\top W_i Z_i + D^{-1}|$ (le « $D(l, b)$ » de Wu) ne dépend pas de β et peut être ignoré pour estimer β.

Il reste à maximiser conjointement en (β, b) la quasi-log-vraisemblance *pénalisée* par $-\frac{1}{2}\sum_i b_i^\top D^{-1} b_i$. En pratique, on itère :

a. à partir des valeurs courantes, former la pseudo-réponse (linéarisation de g) $\tilde{y}_{ij} = x_{ij}^\top \beta + z_{ij}^\top b_i + g'(\mu^b_{ij})(y_{ij} - \mu^b_{ij})$ et les poids $w_{ij} = \{a(\phi)\, v(\mu^b_{ij})\, g'(\mu^b_{ij})^2\}^{-1}$;

b. ajuster le LMM pondéré $\tilde{y}_i = X_i\beta + Z_i b_i + \varepsilon_i$, $\mathrm{Var}(\varepsilon_i) = W_i^{-1}$, par MV ou REML ; obtenir $\hat{\beta}$, $\hat{D}$ et les BLUP $\hat{b}_i$;

c. recommencer jusqu'à convergence.

C'est ce que fait `MASS::glmmPQL`, qui appelle `nlme::lme` à chaque itération.

Biais du PQL

Les estimateurs PQL ne sont pas convergents lorsque $n \to \infty$ à n_i fixé : le biais ne disparaît pas avec le nombre de grappes. Il est prononcé pour des réponses binaires avec peu d'observations par grappe (le cas extrême étant les paires appariées), où les composantes de variance et, par ricochet, les effets fixes sont sous-estimés en valeur absolue (Breslow et Lin, 1995 ; Rodríguez et Goldman, 1995). Le PQL est plus fiable pour des comptes de Poisson élevés ou des binomiales à grand dénominateur, où la loi conditionnelle est proche de la normale.

3.4.4 Monte Carlo et algorithme EM

— *Intégration Monte Carlo.* $L_i \approx M^{-1}\sum_{m=1}^{M} f(y_i \mid b^{(m)})$ avec $b^{(m)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, D)$, ou mieux par échantillonnage préférentiel centré sur le mode $\tilde{b}_i$. L'erreur décroît en $M^{-1/2}$ quelle que soit la dimension q, mais le calcul est lourd et l'estimation est bruitée.

— *Algorithme EM.* On traite les b_i comme des données manquantes. L'étape E demande l'espérance a posteriori de la log-vraisemblance complète (proposition 3.3.2), elle-même une intégrale ; on

l'approxime par Monte Carlo (MCEM) ou par échantillonnage de Gibbs ou de Metropolis-Hastings. L'EM est stable mais lent, et ne fournit pas directement les erreurs types.

— *Approche bayésienne.* Les mêmes outils MCMC donnent directement la loi a posteriori de tous les paramètres.

3.4.5 Comparaison

Le tableau 1 résume les avantages et inconvénients des trois méthodes principales (d'après Wu, § 2.6 et 2.7.2 ; Jiang et Nguyen, § 3.5.2 ; Fitzmaurice et coll., ch. 15).

TABLE 1 – Méthodes de calcul de la vraisemblance d'un GLMM.

Méthode	Avantages	Inconvénients
Gauss-Hermite (adaptative)	Approxime la vraisemblance avec une précision arbitraire ; estimateurs du MV convergents quelle que soit la variance des effets aléatoires ; tests du RV valides.	Coût exponentiel en la dimension (K^q) : un, parfois deux effets aléatoires ; plus lente.
Laplace	Pas d'intégrale ; dimension modérée possible ; effets aléatoires emboîtés ou croisés ; exacte si $D = 0$, bonne si la variance est petite ou si n_i est grand.	Biais si la variance des effets aléatoires est grande et les grappes petites (binaire) ; calcul plus lourd quand le nombre d'effets aléatoires augmente.
PQL	Ne requiert que la moyenne et la variance conditionnelles (pas de loi complète) ; se ramène à des LMM itérés ; rapide ; grand nombre d'effets aléatoires possible.	Estimateurs non convergents ; biais important pour les réponses binaires avec peu d'observations par grappe ; pas de vraisemblance véritable (pas de test du RV ni d'AIC).

Exemple 3.4.3 (Les trois méthodes sur les données `cbpp`)**.** Ajustons le modèle logistique logit $\pi_{ij} = \beta_0 + \beta_j + b_i$, $b_i \sim \mathcal{N}(0, \sigma_b^2)$, par Laplace (`nAGQ = 1`), par AGQ à 5, 9 et 25 points, et par PQL.

```
library(lme4); library(MASS)
gl1 <- glmer(cbind(incidence, size - incidence) ~ period + (1 | herd),
             data = cbpp, family = binomial)                  # Laplace
gl9 <- update(gl1, nAGQ = 9)                                   # Gauss-Hermite
    adaptative
pq  <- glmmPQL(cbind(incidence, size - incidence) ~ period, random = ~ 1 | herd,
               data = cbpp, family = binomial, verbose = FALSE)
```

Méthode	$\hat\beta_0$	$\hat\beta_2$	$\hat\beta_3$	$\hat\beta_4$	$\hat\sigma_b$
Laplace (`nAGQ = 1`)	−1,398	−0,992	−1,128	−1,580	0,642
AGQ, $K = 5$	−1,399	−0,991	−1,128	−1,580	0,647
AGQ, $K = 9$ et $K = 25$	−1,399	−0,991	−1,128	−1,579	0,648
PQL	−1,327	−1,016	−1,150	−1,605	0,556

L'AGQ a convergé dès $K = 5$; Laplace en est très proche (variance modérée, binomiales de taille m_{ij} de 2 à 34). Le PQL sous-estime nettement σ_b et déplace l'ordonnée à l'origine.

Comparer des log-vraisemblances dans `glmer`

Pour des réponses binomiales, `glmer` ne calcule pas la log-vraisemblance avec les mêmes constantes selon que `nAGQ = 1` ou `nAGQ > 1` (sur l'exemple ci-dessus : $-92{,}0$ contre $-50{,}0$, pour des estimations presque identiques). Les valeurs obtenues par Laplace coïncident avec celles de `glm` et de `glmmTMB`. On ne compare donc par test du RV, AIC ou BIC que des modèles ajustés avec la même méthode, et l'on évite de mélanger `glm` et `glmer(nAGQ > 1)`.

Niveau doctorat : des approximations de Laplace au PQL

On suit Jiang et Nguyen (2021, § 3.5.2), en notant $\alpha = b$ (leur notation), $G = D$, et en supposant q effets aléatoires pour l'ensemble des données (le raisonnement est le même grappe par grappe).

Étape 1 (quasi-vraisemblance). On remplace $\log f(y_j \mid \alpha)$ par $-\frac{1}{2a(\phi)}\, d_j(y_j, \mu_j^\alpha)$, où $d(y,\mu) = 2\int_\mu^y \frac{y-u}{v(u)}\,\mathrm{d}u$ est la contribution à la déviance. Il vient

$$L(\beta, G) \propto |G|^{-1/2} \int e^{-q(\alpha)}\,\mathrm{d}\alpha, \qquad q(\alpha) = \frac{1}{2a(\phi)} \sum_j d_j(y_j, \mu_j^\alpha) + \frac{1}{2}\alpha^\top G^{-1}\alpha.$$

Seuls $\mathrm{E}(y \mid \alpha)$ et $\mathrm{Var}(y \mid \alpha)$ interviennent.

Étape 2 (Laplace). Avec $\tilde\alpha = \arg\min q$, $\log L \approx -\frac{1}{2}\log|G| - \frac{1}{2}\log|q''(\tilde\alpha)| - q(\tilde\alpha)$ (leur éq. (3.15) ; le terme q'' correspond au $D(l, b)$ de Wu, p. 87).

Étape 3 (hessienne approchée). Puisque $\partial d_j/\partial\mu = -2(y_j - \mu)/v(\mu)$, on obtient

$$q'(\alpha) = -\sum_j \frac{y_j - \mu_j^\alpha}{a(\phi) v(\mu_j^\alpha) g'(\mu_j^\alpha)}\, z_j + G^{-1}\alpha, \qquad q''(\alpha) = Z^\top W Z + G^{-1} + \mathcal{R},$$

avec $W = \mathrm{diag}\left[\{a(\phi)v(\mu_j^\alpha)g'(\mu_j^\alpha)^2\}^{-1}\right]$ et un reste $\mathcal{R}$ proportionnel aux résidus $y_j - \mu_j^\alpha$, d'espérance nulle. On néglige $\mathcal{R}$ (leur éq. (3.16)) : $|q''| \approx |G|^{-1}\,|I + Z^\top W Z G|$, donc $\log L \approx -\frac{1}{2}\log|I + Z^\top W Z G| - q(\tilde\alpha)$.

Étape 4 (poids constants). On suppose que W varie lentement en fonction de la moyenne, de sorte que $\log|I + Z^\top W Z G|$ ne dépend (presque) pas de β. Pour estimer β, il reste à maximiser $-q(\alpha)$ conjointement en (β, α) : c'est la quasi-vraisemblance pénalisée (leur éq. (3.18)). Les équations de score de $-q$ en (β, α) sont exactement les équations du modèle mixte de Henderson pour la pseudo-réponse $\tilde y$ et les poids W, d'où l'algorithme de la section précédente ; G est ensuite estimée par (RE)ML du LMM de travail.

Conséquences. Les étapes 3 et 4 sont celles qui ne s'améliorent pas quand $n \to \infty$: le PQL n'est pas convergent. Il est valide asymptotiquement quand chaque grappe est grande ($n_i \to \infty$, loi conditionnelle proche de la normale) ou quand $G \to 0$ (Breslow et Lin, 1995, proposent une correction de biais pour G petit). L'approximation de Laplace (étapes 1–2 sans quasi-vraisemblance) donne des estimateurs dont le biais est $O(n_i^{-1})$, et la quadrature adaptative rend ce biais arbitrairement petit. Fitzmaurice et coll. (2011, ch. 15) recommandent pour cette raison le MV par quadrature adaptative lorsque la dimension le permet, et Laplace sinon ; le PQL sert surtout à obtenir rapidement des valeurs initiales ou à traiter de très nombreux effets aléatoires.

3.5 Effets individuels et effets marginaux

3.5.1 Deux paramètres différents

Dans un GLMM, β décrit l'effet d'une covariable *pour une grappe donnée* (même b_i) : c'est un effet *individuel* ou *conditionnel* (« subject-specific »). L'effet *marginal* ou *moyenné sur la population* (« population-averaged ») décrit plutôt la modification de la moyenne dans la population :

$$\mu_{ij}^M = \mathrm{E}(y_{ij}) = \mathrm{E}\{\mathrm{E}(y_{ij} \mid b_i)\} = \int g^{-1}(x_{ij}^\top \beta + z_{ij}^\top b)\, \varphi_q(b; 0, D)\, \mathrm{d}b.$$

Si l'on écrit un modèle marginal $g(\mu_{ij}^M) = x_{ij}^\top \beta^M$ (estimé par exemple par équations d'estimation généralisées, séance 6), β^M et β ne coïncident en général pas. Tout dépend de la fonction de lien.

Proposition 3.5.1 (Selon la fonction de lien)**.**

(a) Lien identité. $\mu_{ij}^M = x_{ij}^\top \beta$ *: effets individuels et marginaux identiques.*

(b) Lien log. *Si* $b_i \sim \mathcal{N}(0, D)$, $\mu_{ij}^M = \exp\{x_{ij}^\top \beta + \frac{1}{2} z_{ij}^\top D z_{ij}\}$. *Avec une ordonnée aléatoire seulement* ($z_{ij} = 1$), $\log \mu_{ij}^M = (\beta_0 + \sigma_b^2/2) + \tilde{x}_{ij}^\top \tilde{\beta}$ *: seule l'ordonnée à l'origine change, et les effets des variables* sans effet aléatoire *sont identiques. Pour une variable qui a un effet aléatoire,* $z_{ij}^\top D z_{ij}$ *dépend de la variable, et l'effet marginal n'est plus égal à l'effet individuel (ni même linéaire).*

(c) Lien logit (et autres liens non linéaires). *Les effets marginaux sont atténués vers 0 (proposition 3.5.3).*

Démonstration. (a) est immédiat. (b) découle de la fonction génératrice des moments de la loi normale : $\mathrm{E}\, e^{z^\top b} = e^{z^\top D z/2}$. Par exemple, avec $z_{ij} = (1, t_{ij})^\top$ et $D = \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}$, $\log \mu_{ij}^M = x_{ij}^\top \beta + \frac{1}{2}(\sigma_0^2 + 2\sigma_{01} t_{ij} + \sigma_1^2 t_{ij}^2)$, quadratique en t_{ij}. □

3.5.2 Atténuation avec le lien logit

Lemme 3.5.2. *Si* $b \sim \mathcal{N}(0, \sigma^2)$ *et* $c > 0$, *alors* $\mathrm{E}\{\Phi(c(a+b))\} = \Phi(ca/\sqrt{1 + c^2\sigma^2})$, *où* Φ *est la fonction de répartition* $\mathcal{N}(0,1)$.

Démonstration. Soit $U \sim \mathcal{N}(0,1)$ indépendante de b. Alors $\mathrm{E}\{\Phi(c(a+b))\} = \mathbb{P}(U \leq ca + cb) = \mathbb{P}(U - cb \leq ca)$ et $U - cb \sim \mathcal{N}(0, 1 + c^2\sigma^2)$. □

Proposition 3.5.3 (Zeger, Liang et Albert, 1988)**.** *Dans le modèle logistique à ordonnée aléatoire* $\operatorname{logit} \mathbb{P}(y_{ij} = 1 \mid b_i) = x_{ij}^\top \beta + b_i$, $b_i \sim \mathcal{N}(0, \sigma^2)$,

$$\operatorname{logit} \mathbb{P}(y_{ij} = 1) \approx \frac{x_{ij}^\top \beta}{\sqrt{1 + c^2\sigma^2}}, \qquad c = \frac{16\sqrt{3}}{15\pi} \approx 0{,}588, \tag{13}$$

c'est-à-dire $\beta^M \approx \beta/\sqrt{1 + 0{,}346\,\sigma^2}$, *avec* $|\beta^M| < |\beta|$ *dès que* $\sigma^2 > 0$.

Esquisse. La fonction logistique $\operatorname{expit}(u) = 1/(1+e^{-u})$ est bien approchée par $\Phi(cu)$; la valeur de c ci-dessus donne une excellente approximation. Par le lemme 3.5.2, $\mathrm{E}\{\operatorname{expit}(\eta+b)\} \approx \mathrm{E}\{\Phi(c(\eta+b))\} = \Phi(c\eta/\sqrt{1 + c^2\sigma^2}) \approx \operatorname{expit}\left(\eta/\sqrt{1 + c^2\sigma^2}\right)$. □

L'approximation est bonne pour des variances modérées et devient grossière quand σ est grand. Avec $\beta_0 = -1$ et $\beta_1 = 1$ pour une covariable binaire, la différence exacte des log-cotes marginales (calculée par intégration numérique) et l'approximation (13) donnent :

σ	0,5	1	2	3
β^M exact	0,947	0,832	0,609	0,461
Approximation (13)	0,959	0,862	0,648	0,493

Exemple 3.5.4 (Propension à une maladie avant et après traitement ; Fitzmaurice et coll., 2011, p. 405–408)**.** Trois sujets ont des probabilités de base très différentes ; le traitement multiplie la cote de chacun par le même facteur, $e^{-0,70} \approx 0,5$.

Sujet	Avant	Après	Différence	Log-RC individuel
A	0,80	0,67	−0,13	−0,68
B	0,50	0,33	−0,17	−0,71
C	0,20	0,11	−0,09	−0,70
Moyenne	0,50	0,37	−0,13	−0,70
Log-cote de la moyenne	0	−0,53		

Le log-RC individuel est d'environ −0,70 pour chacun, mais le log-RC calculé sur les probabilités moyennes de la population est $\text{logit}(0,37) - \text{logit}(0,50) = -0,53$: l'effet marginal est atténué, sans qu'aucun sujet ne réponde moins bien. Si l'on reprend l'exemple avec un risque relatif commun de 0,60 (A : 0,80 → 0,48 ; B : 0,50 → 0,30 ; C : 0,20 → 0,12), le risque relatif des moyennes vaut aussi $0,30/0,50 = 0,60$: avec le lien log, les effets individuels et marginaux coïncident lorsque les risques relatifs individuels sont (à peu près) égaux, comme le prévoit la proposition 3.5.1(b).

Remarque 3.5.5 (Choisir le paramètre selon la question). Aucun des deux paramètres n'est « le bon » : l'effet individuel répond à la question « que se passe-t-il pour un sujet si on modifie son exposition ? » (pertinent pour une covariable qui varie dans la grappe, par exemple un traitement en chassé-croisé), alors que l'effet marginal répond à une question de santé publique (« comment change la prévalence dans la population ? »). La séance 5 (Neuhaus et McCulloch, 2006) approfondira la distinction entre effets intra et inter-grappes.

Remarque 3.5.6 (Exemple 2.11 de Wu : détresse psychologique). Dans l'exemple 2.11 de Wu, le score de dépression est dichotomisé (au-dessus ou au-dessous de la moyenne) et analysé par un GLM et un GLMM logistiques (tableau 2.3). D'après ce qui précède, on s'attendrait à ce que les estimations du GLM (effets marginaux) soient atténuées par rapport à celles du GLMM (effets individuels), et à ce que les erreurs types du GLMM, qui tient compte des mesures répétées, soient plus grandes que celles du GLM pour les covariables fixes dans le temps (groupe de traitement, genre). Le tableau 2.3 montre l'inverse : lisez-le d'un œil critique.

3.6 Sélection du modèle

3.6.1 Sélection des effets aléatoires

Outils graphiques (Wu, § 2.7). Pour une étude longitudinale, on trace la réponse (ou une transformation appropriée, par exemple la log-cote empirique) en fonction du temps pour chaque

sujet (graphique « spaghetti », figure 1). Des niveaux très dispersés suggèrent une ordonnée à l'origine aléatoire ; des pentes très variables, une pente aléatoire. On peut aussi ajuster une droite de régression par sujet (`lmList` dans `lme4`) et examiner la dispersion des ordonnées et des pentes estimées.

Test du rapport de vraisemblance à la frontière. Considérons le test de $H_0 : \sigma_b^2 = 0$ contre $H_1 : \sigma_b^2 > 0$. Le paramètre testé est sur la frontière de l'espace des paramètres ($\sigma_b^2 \geq 0$) : une condition de régularité du théorème de Wilks (paramètre intérieur) n'est pas satisfaite, et la statistique $Q = 2(\ell_1 - \ell_0)$ ne suit pas une loi χ^2_1 sous H_0.

Théorème 3.6.1 (Self et Liang, 1987)**.** *Sous $H_0 : \sigma_b^2 = 0$ (un seul paramètre de variance testé, les autres paramètres intérieurs),*

$$Q \xrightarrow{\mathcal{L}} \tfrac{1}{2}\chi^2_0 + \tfrac{1}{2}\chi^2_1,$$

où χ^2_0 est la masse ponctuelle en 0. La valeur p correcte est donc $\frac{1}{2}\mathbb{P}(\chi^2_1 > q_{obs})$: la moitié de la valeur p usuelle.

Esquisse. Sans contrainte, l'estimateur $\hat{\tau}$ du paramètre (prolongé aux valeurs négatives) serait asymptotiquement normal, et la statistique du RV vaudrait Z^2 avec $Z \sim \mathcal{N}(0,1)$ (statistique de Wald standardisée). Avec la contrainte $\tau \geq 0$, l'estimateur contraint vaut $\max(\hat{\tau}, 0)$ et la statistique devient $Z^2\,\mathbf{1}(Z > 0)$: elle est nulle avec probabilité $\frac{1}{2}$ et, sinon, distribuée comme χ^2_1. □

Cas plus généraux (Stram et Lee, 1994 ; Wu, § 2.7) :

— passer de k à $k+1$ effets aléatoires corrélés (on teste une variance et k covariances) : $Q \xrightarrow{\mathcal{L}} \frac{1}{2}\chi^2_k + \frac{1}{2}\chi^2_{k+1}$; par exemple, ajouter une pente aléatoire corrélée à une ordonnée aléatoire donne $\frac{1}{2}\chi^2_1 + \frac{1}{2}\chi^2_2$;

— tester plusieurs variances simultanément mène à une loi « khi-deux barre », mélange de lois χ^2 dont les poids dépendent de la matrice d'information ;

— utiliser la loi χ^2 usuelle donne un test conservateur (valeur p trop grande) ; `anova()` dans `lme4` ne fait pas la correction.

Ces résultats supposent que les deux modèles sont ajustés par la même méthode (MV ou REML ; même `nAGQ`) et que le nombre de grappes est assez grand ; on peut sinon recourir à un bootstrap paramétrique (simulation sous H_0).

3.6.2 Sélection des covariables

— *Analyse exploratoire ou prédictive.* On peut examiner graphiquement les effets aléatoires prédits à partir d'un modèle nul en fonction des covariables candidates, et comparer des modèles par AIC ou BIC (vraisemblance du MV, pas REML, pour des effets fixes différents).

— *Analyse confirmatoire.* Quand l'objectif est d'estimer l'effet d'une exposition, une sélection automatique invalide l'inférence. On se restreint plutôt aux variables potentiellement confondantes, choisies a priori à l'aide d'un graphe orienté acyclique (DAG) représentant les relations causales supposées entre exposition, réponse et covariables ; on évite d'ajuster sur des médiateurs ou des « colliders ».

3.7 Exemples traités en R

3.7.1 Pleuropneumonie contagieuse bovine : modèle log-binomial mixte

On veut estimer le risque relatif de nouveaux cas d'une période à l'autre ; on utilise donc le lien log.

```
library(lme4)
glog1  <- glm(cbind(incidence, size - incidence) ~ period,
              data = cbpp, family = binomial(link = "log"))       # GLM sans effet alé
    atoire
gmlog1 <- glmer(cbind(incidence, size - incidence) ~ period + (1 | herd),
                data = cbpp, family = binomial(link = "log"))   # Laplace
summary(gmlog1)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
 Family: binomial  ( log )
     AIC      BIC   logLik deviance df.resid
   193.3    203.4    -91.6    183.3       51
Random effects:
 Groups Name        Variance Std.Dev.
 herd   (Intercept) 0.2784   0.5276
Number of obs: 56, groups:  herd, 15
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.6495     0.1899  -8.685  < 2e-16 ***
period2      -0.8415     0.2648  -3.178 0.001483 **
period3      -0.9452     0.2879  -3.283 0.001026 **
period4      -1.3884     0.3925  -3.537 0.000404 ***
```

À titre de comparaison, le GLM donne $\hat{\beta}_4 = -1{,}581$ (erreur type 0,386) et une ordonnée à l'origine de $-1{,}517$ (erreur type 0,113) : l'erreur type de l'ordonnée à l'origine, qui dépend fortement de la variabilité entre troupeaux, est nettement sous-estimée par le GLM.

Quadrature de Gauss-Hermite. Avec `nAGQ = 9`, les estimations changent peu ($\hat{\beta}_0 = -1{,}593$, $\hat{\beta}_4 = -1{,}397$, $\hat{\sigma}_b = 0{,}527$), mais `glmer` avertit que la hessienne calculée par différences finies n'est pas définie positive et la log-vraisemblance affichée vaut $-\infty$. C'est une particularité du lien log avec une réponse binomiale : en certains nœuds de quadrature éloignés, $\beta_0 + \beta_j + b$ devient positif, la « probabilité » e^{η} dépasse 1 et la log-vraisemblance conditionnelle n'est plus définie. Le lien log impose la contrainte $\eta < 0$, que la loi normale des effets aléatoires ne respecte pas ; on retient ici l'ajustement de Laplace.

Surdispersion : effet aléatoire par observation. Un effet aléatoire par troupeau ne capte pas forcément toute la variabilité : on ajoute un effet aléatoire par observation (troupeau × période), emboîté dans le troupeau, qui joue le rôle d'un terme de surdispersion.

```
cbpp$obs <- 1:nrow(cbpp)
gmlog2 <- glmer(cbind(incidence, size - incidence) ~ period + (1 | herd/obs),
                family = binomial(link = "log"), data = cbpp)
```

```
Random effects:
 Groups   Name        Variance Std.Dev.
 obs:herd (Intercept) 0.48144  0.6939
```

```
 herd     (Intercept) 0.03759  0.1939
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.7480     0.2388  -7.321 2.45e-13 ***
period2      -0.9939     0.3929  -2.530  0.01141 *
period3      -1.0596     0.4031  -2.628  0.00858 **
period4      -1.5678     0.5062  -3.097  0.00195 **
```

La variabilité se déplace presque entièrement vers le niveau de l'observation, et les erreurs types des effets de période augmentent sensiblement (de 0,39 à 0,51 pour la période 4) : ignorer la surdispersion rendait l'inférence trop optimiste.

Test de la composante de variance. Le modèle `gmlog2` ajoute une seule variance à `gmlog1` ; sous H_0, elle est nulle, sur la frontière.

```
anova(gmlog1, gmlog2)                       # valeur p du khi-deux usuel (incorrecte ici)

Q <- as.numeric(2 * (logLik(gmlog2) - logLik(gmlog1)))
pchisq(Q, df = 1, lower.tail = FALSE) / 2  # valeur p corrigée (frontière)
```

```
       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
gmlog1    5 193.27 203.40 -91.637   183.27
gmlog2    6 187.49 199.64 -87.745   175.49 7.7829  1   0.005274 **
[1] 0.002637192
```

La valeur p correcte, 0,0026, est la moitié de celle d'`anova` : on conserve l'effet aléatoire par observation. L'AIC et le BIC vont dans le même sens.

Risques relatifs.

```
exp(cbind(RR = fixef(gmlog2), confint(gmlog2, parm = "beta_")))   # IC par profil
```

```
                RR 2.5 % 97.5 %
(Intercept) 0.174 0.103  0.272
period2     0.370 0.160  0.797
period3     0.347 0.145  0.756
period4     0.209 0.070  0.531
```

Le risque de nouveau cas à la période 4 est environ cinq fois plus faible qu'à la période 1 (RR = 0,21, IC à 95 % : 0,07 à 0,53). Comme la période n'a pas d'effet aléatoire et que le lien est log, ces risques relatifs ont aussi une interprétation marginale (proposition 3.5.1(b)).

Prédictions. La commande `exp(predict(gmlog1))` donne les prédictions conditionnelles $\exp(x_{ij}^\top \hat\beta + \tilde b_i)$. Dans un GLMM, `ranef()` renvoie les modes conditionnels $\tilde b_i$, analogues non gaussiens du BLUP. La figure 2 montre que le modèle à effet de troupeau seul impose la même évolution relative à tous les troupeaux, alors que l'effet par observation permet aux prédictions de suivre de près les données (en les rétrécissant vers la tendance commune).

3.7.2 Essai d'une crème anti-infectieuse : effet individuel ou marginal

Dans un essai multicentrique, les patients de huit cliniques reçoivent une crème active ($t = 1$) ou un placebo ($t = 0$) ; on compte les patients dont l'infection s'est résorbée (x parmi n). Il y a 16 lignes

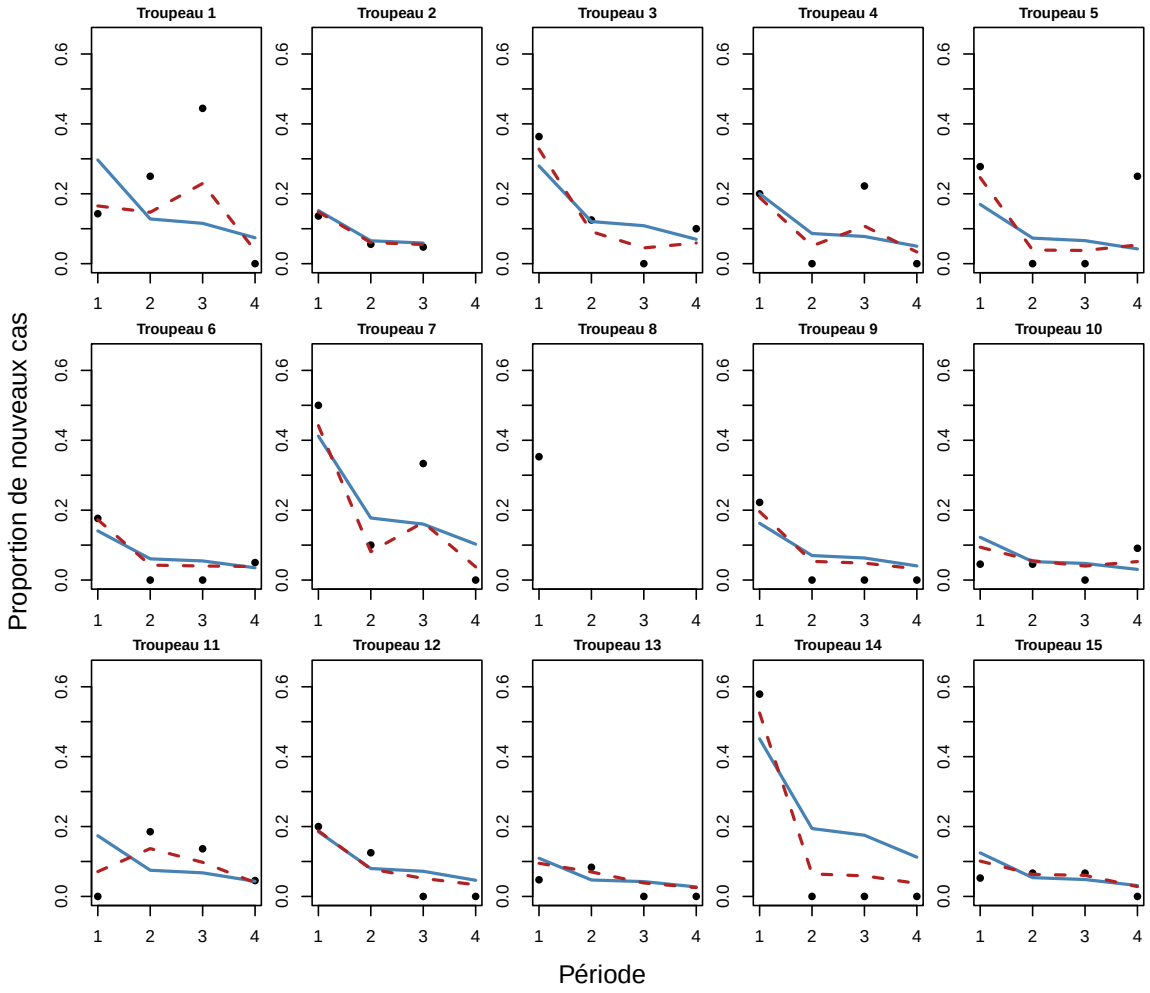


FIGURE 2 – Proportions observées (points) et prédictions conditionnelles des modèles `gmlog1` (effet aléatoire du troupeau ; trait plein bleu) et `gmlog2` (troupeau et observation ; tirets rouges).

(fichier `infection.txt`).

```
infection <- read.table("infection.txt", header = TRUE)
model1 <- glmer(cbind(x, n - x) ~ t + (1 | clinic), family = binomial,
                nAGQ = 9, data = infection)
summary(model1)
exp(cbind(OR = fixef(model1), confint(model1, parm = "beta_")))
```

```
Random effects:
 Groups Name        Variance Std.Dev.
 clinic (Intercept) 1.962    1.401
Number of obs: 16, groups:  clinic, 8
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.1976     0.5565  -2.152   0.0314 *
t             0.7385     0.3004   2.458   0.0140 *
               OR 2.5 % 97.5 %
(Intercept) 0.302 0.083  0.985
t           2.093 1.171  3.819
```

Dans une clinique donnée, la crème double environ la cote de guérison (RC individuel 2,09). Avec Laplace, on obtient pratiquement la même chose ($\hat{\beta}_1 = 0{,}738$, $\hat{\sigma}_b = 1{,}389$) ; le PQL donne $\hat{\beta}_1 = 0{,}721$ et $\hat{\sigma}_b = 1{,}323$, un peu plus petits, comme attendu.

Comparons au modèle marginal estimé par GEE (séance 6), qui estime directement β^M :

```
library(geepack)
model1.gee <- geeglm(cbind(x, n - x) ~ t, id = clinic, family = binomial,
                     data = infection)
```

```
            Estimate Std.err  Wald Pr(>|W|)
(Intercept)  -0.7142  0.4505 2.514    0.113
t             0.4040  0.2462 2.693    0.101
```

Le RC marginal est $e^{0{,}404} = 1{,}50$ (IC de Wald à 95 % : 0,92 à 2,43). L'atténuation est importante parce que la variance entre cliniques est grande : l'approximation (13) donne $0{,}7385/\sqrt{1 + 0{,}346 \times 1{,}962} = 0{,}57$ et l'intégration numérique des probabilités moyennes implicites du GLMM, $\text{logit}(0{,}417) - \text{logit}(0{,}293) = 0{,}55$. L'estimation GEE (0,40) est plus petite encore ; avec huit cliniques de tailles inégales, les deux estimateurs du paramètre marginal (implicite au GLMM, ou direct par GEE) peuvent différer sensiblement, mais tous deux confirment que l'effet marginal est plus faible que l'effet individuel.

Avec le lien log, en revanche (risques relatifs, ordonnée aléatoire seulement), le GLMM donne $\hat{\beta}_1 = 0{,}254$ (erreur type 0,130) et les GEE 0,253 (erreur type robuste 0,168) : les effets individuel et marginal coïncident, comme le prévoit la proposition 3.5.1(b).

Effet du traitement aléatoire ? Le modèle `(1 + t | clinic)`, ajusté par Laplace (`nAGQ > 1` n'est pas disponible avec deux effets aléatoires), donne $Q = 2(-36{,}831 + 37{,}101) = 0{,}539$. Sous H_0 (variance de la pente et covariance nulles), la loi asymptotique est $\frac{1}{2}\chi^2_1 + \frac{1}{2}\chi^2_2$, d'où une valeur p de $\frac{1}{2}(0{,}463) + \frac{1}{2}(0{,}764) = 0{,}61$ (au lieu de 0,76 avec χ^2_2) : rien n'indique que l'effet de la crème varie d'une clinique à l'autre. Inversement, le test de l'ordonnée aléatoire elle-même (GLMM ajusté par Laplace contre GLM) donne $Q = 55{,}1$, $p < 10^{-13}$.

3.8 Synthèse

Grille méthodologique

1. *Objet d'intérêt* : les effets fixes β, interprétés conditionnellement aux effets aléatoires (effets individuels) ; parfois les composantes de variance elles-mêmes (par exemple, une part de variance attribuable aux établissements).
2. *Nuisance* : les effets aléatoires b_i (non observés) et, le plus souvent, les composantes de variance D et ϕ.
3. *Traitement de la nuisance* : intégrer les effets aléatoires par rapport à leur loi ; dans le LMM, REML élimine en plus β pour estimer les variances sans biais.
4. *Critère* : vraisemblance marginale (8) (MV), vraisemblance restreinte (LMM), ou quasi-vraisemblance pénalisée.
5. *Calcul* : formes explicites dans le LMM ; quadrature de Gauss-Hermite adaptative (une

ou deux dimensions), approximation de Laplace (dimension modérée), PQL (rapide mais biaisée), Monte Carlo et EM.

6. *Inférence* : tests et intervalles de Wald ou par profil de vraisemblance ; test du RV, avec la correction de frontière $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$ pour une composante de variance ; bootstrap paramétrique dans les petits échantillons.

À retenir

— Un GLMM = un GLM conditionnel aux effets aléatoires + une loi pour ces effets ; sa vraisemblance est une intégrale sans forme explicite hors du cas gaussien.

— Laplace = développement de Taylor d'ordre 2 de la log-densité conjointe autour du mode des effets aléatoires ; exacte si la variance est nulle (et pour le LMM), bonne si elle est petite ; c'est la quadrature adaptative à un point.

— Gauss-Hermite adaptative = référence pour un effet aléatoire ; PQL = LMM itérés sur une pseudo-réponse, biaisée pour des réponses binaires avec de petites grappes.

— Avec le lien logit, $|\beta^M| < |\beta|$: $\beta^M \approx \beta/\sqrt{1 + 0{,}346\,\sigma^2}$. Avec le lien identité, $\beta^M = \beta$; avec le lien log, égalité pour les variables sans effet aléatoire.

— Tester une variance nulle : diviser la valeur p du χ^2_1 par 2.

La séance suivante (section 4) étend le prédicteur linéaire à des fonctions lisses des covariables : nous verrons qu'un modèle additif généralisé pénalisé s'écrit lui-même comme un modèle mixte, dont le paramètre de lissage est une composante de variance.

3.9 Exercices

Exercice 3.9.1 (MLM, H26, exercice 1)**.**

(a) Sous le modèle linéaire mixte (1), démontrez que $\mathrm{Cov}(b_i, y_i) = DZ_i^\top$.

(b) À partir de cette expression et de celles de $\mathrm{Var}(y_i)$ et $\mathrm{Var}(b_i)$, dérivez l'estimateur bayésien empirique $\hat{b}_i = \mathrm{E}(b_i \mid y_i, \hat{\theta})$ (Wu, éq. (2.13)). Refaites la démonstration de la proposition 3.2.8 sans consulter les notes, puis montrez que $\mathrm{Var}(b_i \mid y_i) \preceq D$ au sens des matrices et interprétez.

Exercice 3.9.2 (Pratique d'examen H23, question 2)**.** On mesure la tension artérielle systolique y_{ij} du sujet i à k_i occasions, à l'âge t_{ij} : $y_{ij} = \beta_0 + \beta_1 t_{ij} + \alpha_i + \epsilon_{ij}$, $\alpha_i \sim \mathcal{N}(0, \sigma^2)$, $\epsilon_{ij} \sim \mathcal{N}(0, \tau^2)$, tous indépendants.

(a) Écrivez y_i, X_i, Z_i et ϵ_i dans la forme $y_i = X_i\beta + Z_i\alpha_i + \epsilon_i$.

(b) Sachant que $(V_i^{-1})_{jj} = \dfrac{\tau^2 + (k_i - 1)\sigma^2}{\tau^2(\tau^2 + k_i\sigma^2)}$ et $(V_i^{-1})_{jl} = \dfrac{-\sigma^2}{\tau^2(\tau^2 + k_i\sigma^2)}$ pour $j \neq l$, écrivez une formule explicite pour $\mathrm{E}(\alpha_i \mid y_i)$ à partir de la formule générale du BLUP, et comparez à l'exemple 3.2.9.

(c) Donnez une expression pour la prédiction de la tension artérielle de l'individu i à 50 ans.

Exercice 3.9.3 (Complément au document MVR, H26)**.** Dans la régression linéaire $y = X\beta + e$,

$e \sim \mathcal{N}_N(0, \sigma^2 I)$, X de rang p, utilisez le théorème 3.2.6 pour montrer que l'estimateur REML de σ^2 est SCR$/(N-p)$. Comparez à l'estimateur du MV.

Exercice 3.9.4 (Examen final H23, question 2)**.** On observe que la capacité respiratoire de patients atteints d'une maladie chronique varie selon une loi normale à la grandeur du Québec. Une étude s'intéresse aux sources de cette variation parmi les établissements de santé, les médecins traitants et les patients eux-mêmes. Quelques facteurs sociodémographiques (âge, sexe, ...) doivent être pris en compte sans être l'objet de l'étude.

(a) Identifiez les objets statistiques d'intérêt et les paramètres de nuisance.

(b) Si l'estimation sans biais de l'objet d'intérêt est primordiale, quelle méthode d'estimation proposez-vous ? Expliquez informellement ce qu'il advient des paramètres de nuisance.

Exercice 3.9.5 (MLMG, H26, exercice 1)**.**

(a) Dérivez l'approximation de Laplace, équation (2.78) de Wu (théorème 3.4.1). Pour simplifier, vous pouvez poser $k = 1$.

(b) *Défi.* Dérivez une expression pour la hessienne $\partial^2 p(v)/\partial v\, \partial v^\top$ de l'équation (2.78) pour un GLMM de fonction de dispersion $a(\phi)$, de fonction de variance $v(\mu)$ et de lien g quelconque. Vérifiez que vous retrouvez (12) (au facteur k près) avec le lien canonique, et identifiez le terme supplémentaire qui apparaît avec un lien non canonique.

Exercice 3.9.6 (Examen final H26, question 1)**.** Une étude épidémiologique du mélanome au Canada s'intéresse à la variation géographique de l'incidence et à l'exposition aux rayons ultraviolets B (UVB). On dispose du nombre de cas de mélanome diagnostiqués pendant une année dans chaque localité i, dont la population est connue par le dernier recensement. Les localités sont nichées dans des régions j, elles-mêmes nichées dans des provinces k. Le rayonnement UVB est mesuré au niveau des régions.

(a) Proposez un GLMM du taux d'incidence du mélanome en fonction du rayonnement UVB qui tient compte des sources de variation géographique. Définissez clairement les variables observées et les lois des variables aléatoires.

(b) Proposez un critère d'estimation et une méthode de calcul, en les justifiant par la complexité du modèle. Décrivez précisément l'approximation utilisée (formules à l'appui ; exigences plus élevées au doctorat).

Exercice 3.9.7 (MLMG, H26, exercice de niveau doctoral)**.**

(1) À la suite de la lecture du § 3.5.2 de Jiang et Nguyen (2021), décrivez les approximations ajoutées à l'approximation de Laplace de la vraisemblance d'un GLMM pour obtenir la quasi-vraisemblance pénalisée (PQL).

(2) Discutez les avantages et les inconvénients de l'approche PQL, de l'approximation de Laplace et de la quadrature de Gauss-Hermite, selon le § 3.5.2 de Jiang et Nguyen (2021), les § 2.6 et 2.7.2 de Wu (2010) et le ch. 15 de Fitzmaurice et coll. (2011).

Exercice 3.9.8 (Exercice pratique, devoir 1 H26)**.** Le fichier `Dev1Ex3.txt` (*fichier à fournir* : il ne fait pas partie du matériel H26 récupéré) contient les données de suivi médical de $I = 60$ patients répartis en 3 groupes. Chaque groupe reçoit un médicament, puis chaque patient est suivi pendant 4 mois. Chaque mois, on note le nombre d'épisodes d'inconfort vécus par le patient (variable réponse `y`).

Analysez ces données à l'aide d'un GLMM avec une réponse de Poisson et le lien log (lien canonique, qui favorise la convergence).

(a) Sélectionnez les effets aléatoires du modèle. Pour ce faire, incluez dans les modèles comparés des effets fixes pour le temps, le groupe et leur interaction.

(b) Répondez ensuite aux questions suivantes sur les effets fixes : y a-t-il une interaction entre le médicament et le temps ? Y a-t-il une différence entre les médicaments ? Si oui, lesquels diffèrent ? Le temps est-il significativement associé au nombre d'épisodes d'inconfort ?

(c) Écrivez le modèle retenu sous forme matricielle, $\log \mathrm{E}(Y_i \mid X_i, b_i) = X_i\beta + Z_i b_i$, $i = 1, \ldots, 60$. Donnez les dimensions de Y_i, X_i et Z_i et écrivez explicitement b_i et β. Quels sont les paramètres du modèle retenu ?

Références

— Breslow, N. E. et Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9–25.

— Breslow, N. E. et Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika*, 82, 81–91.

— Fitzmaurice, G. M., Laird, N. M. et Ware, J. H. (2011). *Applied Longitudinal Analysis*, 2e éd. Wiley.

— Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, 61, 383–385.

— Jiang, J. et Nguyen, T. (2021). *Linear and Generalized Linear Mixed Models and Their Applications*, 2e éd. Springer.

— Rodríguez, G. et Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society A*, 158, 73–89.

— Self, S. G. et Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82, 605–610.

— Stram, D. O. et Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50, 1171–1177.

— Wu, L. (2010). *Mixed Effects Models for Complex Data*. Chapman & Hall/CRC, ch. 2.

— Zeger, S. L., Liang, K.-Y. et Albert, P. S. (1988). Models for longitudinal data : a generalized estimating equation approach. *Biometrics*, 44, 1049–1060.