Section 5

Étude de cas 1 : effets intra et inter-grappes

Séance du 12 février 2027
Article : Neuhaus, J. M. et McCulloch, C. E. (2006). Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society B*, 68(5), 859–872.
Lectures complémentaires : Wu, § 2.4 et 2.6 (GLMM) ; notes de la séance 3 ; Breslow et Day (1980), ch. 7

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

Objectifs et déroulement de la séance

Ces notes présentent la théorie nécessaire à l'étude de cas. Les questions à traiter en équipe, les consignes et le barème se trouvent dans le *document de consignes de l'étude de cas 1* ; on n'y revient pas ici. À la fin de la séance, vous devriez être en mesure de :

1. expliquer pourquoi une corrélation entre l'effet aléatoire de grappe et une covariable biaise l'estimateur du modèle mixte standard, et quantifier ce biais dans le modèle linéaire ;
2. écrire et interpréter le modèle qui décompose une covariable en composantes intra-grappe (β_W) et inter-grappes (β_B), ainsi que sa reparamétrisation de Mundlak ;
3. dériver la vraisemblance conditionnelle qui élimine l'effet de grappe dans les modèles linéaire et logistique, et la relier à la régression logistique conditionnelle des études cas-témoins appariées ;
4. comparer les trois approches (modèle mixte standard, décomposition, vraisemblance conditionnelle) quant à la robustesse, à l'efficacité et au calcul ;
5. estimer par bootstrap par grappes l'erreur type d'une différence de coefficients.

1 Mise en contexte : âge maternel et poids de naissance

Neuhaus et McCulloch (2006) illustrent leurs méthodes avec des données de naissances de l'État de Géorgie : pour chaque mère, on dispose des poids de naissance de plusieurs enfants successifs et de l'âge de la mère à chaque naissance. La question est l'effet de l'âge maternel sur le poids de naissance (ou sur le risque de faible poids, ≤ 2500 g). Le fichier `naisga.csv` utilisé en classe en contient une version plus complète : 8834 naissances de 1770 mères (presque toutes avec cinq naissances), avec les variables `matid` (identifiant de la mère), `poidsnais` (g), `mat_age` (ans), `dprenat`, `sexe`, `matrace` et `fumeuse` (cette dernière très souvent manquante).

L'âge maternel varie à deux niveaux :

— *entre les mères* : certaines ont leurs enfants jeunes, d'autres plus tard (écart type de l'âge moyen par mère : 3,6 ans) ;

— *à l'intérieur d'une mère* : l'âge augmente d'une naissance à l'autre (écart type des écarts à la moyenne de la mère : 2,8 ans).

Or, de nombreux facteurs propres à la mère (statut socio-économique, scolarité, génétique, habitudes de vie, accès aux soins) influencent le poids de ses bébés *et* sont associés à l'âge auquel elle a ses enfants. Dans un modèle mixte, ces facteurs non mesurés se retrouvent dans l'effet aléatoire de la mère b_i, qui devient alors corrélé avec l'âge. La comparaison entre mères est donc sujette à de la confusion au niveau de la mère, alors que la comparaison de naissances d'une même mère contrôle automatiquement tous les facteurs constants chez cette mère, mesurés ou non. C'est la logique des études appariées.

La structure se retrouve dans de nombreux contextes : patients groupés par médecin ou par hôpital, travailleurs par usine, élèves par école, mesures répétées chez un sujet (où la composante inter-sujets d'une covariable variant dans le temps peut refléter des caractéristiques stables du sujet).

2 Le problème : effet aléatoire corrélé avec une covariable

2.1 Le modèle standard

Pour la réponse Y_{ij} de l'unité $j = 1, \ldots, n_i$ de la grappe $i = 1, \ldots, K$ et une covariable X_{ij}, le modèle linéaire généralisé mixte (GLMM) à ordonnée à l'origine aléatoire s'écrit

$$h\{\mathrm{E}(Y_{ij} \mid b_i, \mathbf{X}_i)\} = \beta_0 + \beta X_{ij} + b_i, \qquad b_i \sim \mathcal{N}(0, \sigma_b^2), \tag{1}$$

où $\mathbf{X}_i = (X_{i1}, \ldots, X_{in_i})^\top$. L'ajustement standard (par `lmer` ou `glmer`) suppose implicitement que

$$b_i \perp\!\!\!\perp \mathbf{X}_i, \qquad \text{donc en particulier} \qquad \mathrm{E}(b_i \mid \mathbf{X}_i) = 0.$$

C'est cette hypothèse qui permet d'intégrer b_i contre sa loi marginale $\mathcal{N}(0, \sigma_b^2)$ pour obtenir la vraisemblance. Si b_i dépend de $\mathbf{X}_i$, la loi qu'il faudrait intégrer est la loi conditionnelle de b_i sachant $\mathbf{X}_i$; la vraisemblance standard est alors mal spécifiée et l'estimateur de β n'est pas convergent en général.

2.2 Le modèle linéaire : l'estimateur mixte est une moyenne pondérée

Dans le cas linéaire (h identité, $Y_{ij} = \beta_0 + \beta X_{ij} + b_i + \varepsilon_{ij}$, $\varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$), on peut décrire exactement ce qui se passe. Définissons les sommes de carrés et de produits intra- et inter-grappes

$$W_{xy} = \sum_{i,j} (X_{ij} - \bar{X}_i)(Y_{ij} - \bar{Y}_i), \qquad B_{xy} = \sum_i n_i (\bar{X}_i - \bar{X})(\bar{Y}_i - \bar{Y}),$$

et de même W_{xx}, B_{xx}, où $\bar{X}_i$ est la moyenne de la grappe i et $\bar{X}$ la moyenne globale. L'estimateur *intra* (ou à effets fixes de grappe) et l'estimateur *inter* (régression des moyennes) sont

$$\hat{\beta}_W = W_{xy}/W_{xx}, \qquad \hat{\beta}_B = B_{xy}/B_{xx}.$$

Proposition 2.1. *Supposons des grappes de même taille n et des composantes de variance connues. L'estimateur des moindres carrés généralisés (MCG) de β dans le modèle linéaire mixte* (1) *est*

$$\hat{\beta}_{\text{MCG}} = w\,\hat{\beta}_W + (1-w)\,\hat{\beta}_B, \qquad w = \frac{W_{xx}}{W_{xx}+\theta B_{xx}}, \qquad \theta = \frac{\sigma^2}{\sigma^2 + n\sigma_b^2} \in (0,1].$$

Démonstration. La matrice de covariance de $\mathbf{y}_i$ est $V = \sigma^2 I_n + \sigma_b^2 J_n$, où $J_n = \mathbf{1}\mathbf{1}^\top$. Écrivons $P = J_n/n$ (projection sur les constantes) et $Q = I_n - P$ (projection sur les écarts à la moyenne). Comme $V = \sigma^2 Q + (\sigma^2 + n\sigma_b^2)P$ et que P et Q sont des projections orthogonales complémentaires,

$$V^{-1} = \sigma^{-2}Q + (\sigma^2 + n\sigma_b^2)^{-1}P = \sigma^{-2}(Q + \theta P).$$

L'équation d'estimation MCG pour l'ordonnée à l'origine, $\sum_i \mathbf{1}^\top V^{-1}(\mathbf{y}_i - \beta_0\mathbf{1} - \beta\mathbf{X}_i) = 0$, donne $\hat{\beta}_0 = \bar{Y} - \beta\bar{X}$ (car $\mathbf{1}^\top Q = 0$). En substituant, et en remarquant que $Q\mathbf{X}_i = \mathbf{X}_i - \bar{X}_i\mathbf{1}$ et $P(\mathbf{X}_i - \bar{X}\mathbf{1}) = (\bar{X}_i - \bar{X})\mathbf{1}$, l'équation pour β devient

$$\sum_i (\mathbf{X}_i - \bar{X}\mathbf{1})^\top (Q+\theta P)\{\mathbf{y}_i - \bar{Y}\mathbf{1} - \beta(\mathbf{X}_i - \bar{X}\mathbf{1})\} = 0 \iff W_{xy} + \theta B_{xy} = \beta\,(W_{xx} + \theta B_{xx}).$$

Ainsi $\hat{\beta}_{\text{MCG}} = (W_{xx}\hat{\beta}_W + \theta B_{xx}\hat{\beta}_B)/(W_{xx} + \theta B_{xx})$. □

Les cas limites éclairent le résultat : si $\sigma_b^2 = 0$, $\theta = 1$ et l'on obtient l'estimateur des moindres carrés ordinaires, qui pondère les deux sources d'information selon leur variabilité brute ; si $\sigma_b^2 \to \infty$ ou $n \to \infty$, $\theta \to 0$ et l'estimateur mixte tend vers l'estimateur intra. Pour des grappes de tailles inégales, le même raisonnement vaut avec $\theta_i = \sigma^2/(\sigma^2 + n_i\sigma_b^2)$ propre à chaque grappe (et un centrage pondéré) ; on vérifiera numériquement la formule sur les données de naissances à la section 7.

Corollaire 2.2 (Biais de l'estimateur mixte)**.** *Supposons que* $\mathrm{E}(b_i \mid \mathbf{X}_i) = \delta(\bar{X}_i - \mu_X)$ *(l'effet de grappe dépend linéairement de la moyenne de la covariable dans la grappe). Alors, conditionnellement aux covariables,*

$$\mathrm{E}(\hat{\beta}_W) = \beta, \qquad \mathrm{E}(\hat{\beta}_B) = \beta + \delta, \qquad \mathrm{E}(\hat{\beta}_{\text{MCG}}) = \beta + (1-w)\,\delta.$$

Démonstration. Dans les écarts à la moyenne de grappe, $Y_{ij} - \bar{Y}_i = \beta(X_{ij} - \bar{X}_i) + (\varepsilon_{ij} - \bar{\varepsilon}_i)$: b_i disparaît, donc $\mathrm{E}(W_{xy} \mid \mathbf{X}) = \beta W_{xx}$ quelle que soit la relation entre b_i et $\mathbf{X}_i$. Pour les moyennes, $\bar{Y}_i = \beta_0 + \beta\bar{X}_i + b_i + \bar{\varepsilon}_i$; en prenant l'espérance conditionnelle et en centrant, $\mathrm{E}(B_{xy} \mid \mathbf{X}) = \beta B_{xx} + \delta\sum_i n(\bar{X}_i - \bar{X})(\bar{X}_i - \mu_X) = (\beta + \delta)B_{xx}$. On conclut par la proposition 2.1. □

L'estimateur intra est toujours sans biais ; l'estimateur mixte standard hérite d'une fraction $1 - w$ du biais de l'estimateur inter. Le biais est d'autant plus grand que la variabilité inter-grappes de la covariable est importante (B_{xx} grand) et que la corrélation intra-grappe est faible (θ proche de 1). L'article exprime de la même façon la limite de l'estimateur standard comme une combinaison des effets intra et inter.

2.3 Le cas non linéaire

Dans un GLMM logistique, il n'y a pas de forme explicite, mais le même phénomène se produit : l'estimateur du maximum de vraisemblance de β dans le modèle standard converge vers une valeur qui mélange les informations intra et inter, et il est biaisé dès que b_i est associé à $\mathbf{X}_i$. S'ajoute

une difficulté propre aux liens non linéaires : la spécification de la *loi* de b_i (normalité) influence les estimateurs des effets de niveau grappe, alors qu'elle n'influençait pas les estimateurs MCG du modèle linéaire. Une petite simulation à la section 8 illustre ces deux sources de biais.

3 Décomposition intra/inter et modèle de Mundlak

3.1 Le modèle décomposé

L'idée de la méthode de partition (*partitioning*) est d'écrire $X_{ij} = \bar{X}_i + (X_{ij} - \bar{X}_i)$ et de donner à chaque composante son propre coefficient (Neuhaus et Kalbfleisch, 1998) :

$$h\{\mathrm{E}(Y_{ij} \mid b_i, \mathbf{X}_i)\} = \beta_0 + \beta_W (X_{ij} - \bar{X}_i) + \beta_B \bar{X}_i + b_i, \qquad b_i \sim \mathcal{N}(0, \sigma_b^2). \tag{2}$$

— β_W (*within*) est l'effet d'une variation de la covariable *à l'intérieur d'une grappe* : chez une même mère, la différence de log-cote (ou de poids moyen) associée à une naissance survenue un an plus tard. C'est un effet conditionnel à la grappe.

— β_B (*between*) compare des grappes dont les moyennes de covariable diffèrent d'une unité. Il combine l'effet de la covariable et l'effet de tous les facteurs de grappe associés à $\bar{X}_i$.

— Le modèle standard (1) est le cas particulier $\beta_W = \beta_B = \beta$. Tester $H_0 : \beta_W = \beta_B$ est donc un diagnostic de l'hypothèse $b_i \perp\!\!\!\perp \mathbf{X}_i$ (sous une forme linéaire en $\bar{X}_i$).

3.2 La reparamétrisation de Mundlak

Mundlak (1978) part du modèle (1) et modélise explicitement la dépendance de l'effet aléatoire envers les covariables par une régression auxiliaire $b_i = \alpha \bar{X}_i + b_i^\dagger$, avec $b_i^\dagger \perp\!\!\!\perp \mathbf{X}_i$. En substituant :

$$h\{\mathrm{E}(Y_{ij} \mid b_i^\dagger, \mathbf{X}_i)\} = \beta_0 + \beta X_{ij} + \alpha \bar{X}_i + b_i^\dagger.$$

Ce modèle est une simple reparamétrisation de (2) : en développant $\beta_W(X_{ij} - \bar{X}_i) + \beta_B \bar{X}_i = \beta_W X_{ij} + (\beta_B - \beta_W)\bar{X}_i$, on obtient

$$\beta = \beta_W, \qquad \alpha = \beta_B - \beta_W.$$

Le coefficient de X_{ij} dans le modèle de Mundlak est l'effet intra ; celui de $\bar{X}_i$ est la *différence* entre les effets inter et intra, et son test de nullité est le test de $H_0 : \beta_W = \beta_B$. Mundlak a montré que, dans le modèle linéaire, l'estimateur MCG de β dans ce modèle augmenté est exactement l'estimateur intra ; la proposition suivante en donne la raison.

Proposition 3.1. *Dans le modèle linéaire mixte à ordonnée à l'origine aléatoire avec les covariables $(X_{ij} - \bar{X}_i)$ et $\bar{X}_i$, l'estimateur MCG de β_W est égal à l'estimateur intra $\hat{\beta}_W = W_{xy}/W_{xx}$, quelles que soient les valeurs (connues ou estimées) des composantes de variance et la taille des grappes.*

Démonstration. Pour la grappe i, $V_i^{-1} = \sigma^{-2}(Q_i + \theta_i P_i)$ comme dans la preuve de la proposition 2.1. La colonne des écarts $\mathbf{d}_i = Q_i \mathbf{X}_i$ vérifie $P_i \mathbf{d}_i = 0$, alors que les colonnes $\mathbf{1}$ et $\bar{X}_i \mathbf{1}$ sont dans l'image de P_i et vérifient $Q_i \mathbf{1} = 0$. Donc $\mathbf{d}_i^\top V_i^{-1} \mathbf{1} = \mathbf{d}_i^\top V_i^{-1} \bar{X}_i \mathbf{1} = 0$: la colonne des écarts est orthogonale aux deux autres pour le produit scalaire défini par V^{-1}. Les équations MCG se découplent, et l'équation pour β_W est $\sum_i \mathbf{d}_i^\top V_i^{-1}(\mathbf{y}_i - \beta_W \mathbf{d}_i) = 0$, soit $\sigma^{-2} \sum_i \mathbf{d}_i^\top Q_i (\mathbf{y}_i - \beta_W \mathbf{d}_i) = 0$, qui ne dépend ni de θ_i ni de σ^2 et dont la solution est W_{xy}/W_{xx}. □

3.3 Interprétation par un effet aléatoire modifié

On peut réécrire (2) comme

$$h\{\mathrm{E}(Y_{ij} \mid b_i^*, \mathbf{X}_i)\} = \beta_0 + \beta_W(X_{ij} - \bar{X}_i) + b_i^*, \qquad b_i^* = b_i + \beta_B \bar{X}_i.$$

Toute la variation entre grappes, celle qui provient de la covariable comme celle qui provient de facteurs non mesurés, est regroupée dans le nouvel effet de grappe b_i^*. Celui-ci est corrélé avec la covariable par sa composante de grappe $\bar{X}_i$, tandis que la composante intra $X_{ij} - \bar{X}_i$ n'en dépend pas. C'est exactement le type de structure que considèrent Neuhaus et McCulloch : une covariable qui s'écrit comme une fonction de l'effet de grappe plus une composante propre à l'unité, indépendante de cet effet (la question 2 des consignes demande d'en donner la forme explicite). Dans une telle structure,

— l'information sur β_W provient des écarts $X_{ij} - \bar{X}_i$, qui sont (approximativement) indépendants de l'effet de grappe : l'estimation de β_W est donc protégée contre l'association entre la covariable et l'effet de grappe ;

— toute l'association entre la covariable et l'effet de grappe est absorbée par le terme $\beta_B \bar{X}_i$, dont l'interprétation causale est perdue.

Remarque 3.2. Dans le modèle linéaire, la protection de $\hat{\beta}_W$ est exacte (proposition 3.1). Dans un GLMM, elle n'est qu'approximative, pour deux raisons : la moyenne observée $\bar{X}_i$ n'est qu'une approximation bruitée de la composante de grappe (surtout pour de petites grappes), et l'association réelle entre b_i et $\bar{X}_i$ peut ne pas être linéaire. L'article montre néanmoins, analytiquement et par simulation, que l'estimateur de β_W par décomposition reste très proche de l'estimateur conditionnel (section suivante), et que ce dernier est sans biais. Par contre, l'estimateur de β_B dépend de l'hypothèse sur la loi de l'effet aléatoire et peut être biaisé si elle est fausse.

À retenir

— Si l'effet de grappe est corrélé avec une covariable, le modèle mixte standard mélange un effet intra non biaisé et un effet inter biaisé ; dans le cas linéaire, $\hat{\beta}_{\mathrm{MCG}} = w\hat{\beta}_W + (1-w)\hat{\beta}_B$.

— Le modèle décomposé estime séparément β_W (comparaison intra-grappe, protégée de la confusion de niveau grappe) et β_B (comparaison entre grappes, sujette à confusion).

— Modèle de Mundlak : coefficient de $X_{ij} = \beta_W$, coefficient de $\bar{X}_i = \beta_B - \beta_W$.

— $b_i^* = b_i + \beta_B \bar{X}_i$ regroupe toute la variation de grappe ; la composante intra de la covariable ne dépend pas de l'effet de grappe.

4 Vraisemblance conditionnelle

4.1 Principe

Plutôt que de modéliser la loi de b_i, on peut l'éliminer. Traitons $\alpha_i = \beta_0 + b_i$ comme un paramètre de nuisance propre à la grappe. S'il existe une statistique T_i exhaustive pour α_i lorsque β est fixé, alors la loi conditionnelle $f(\mathbf{y}_i \mid T_i; \beta)$ ne dépend plus de α_i. La *vraisemblance conditionnelle*

$$L_C(\beta) = \prod_{i=1}^{K} f(\mathbf{y}_i \mid T_i = t_i, \mathbf{X}_i; \beta)$$

ne fait alors aucune hypothèse sur les α_i : ni sur leur loi, ni sur leur relation avec les covariables. Pour un GLM à lien canonique, $f(y_{ij} \mid \alpha_i) \propto \exp[\{y_{ij}(\alpha_i + \beta X_{ij}) - c(\alpha_i + \beta X_{ij})\}/\phi]$, et la factorisation de Neyman–Fisher montre que $T_i = \sum_j Y_{ij}$ est exhaustive pour α_i. Conditionner sur T_i revient à ne garder que l'information contenue dans la répartition de la réponse *à l'intérieur* de la grappe.

Remarque 4.1. Il est essentiel de distinguer cette approche du maximum de vraisemblance conjoint en $(\alpha_1, \ldots, \alpha_K, \beta)$: avec un nombre de paramètres de nuisance qui croît avec K et des grappes de taille fixe, l'estimateur conjoint de β n'est pas convergent dans le modèle logistique (problème de Neyman–Scott ; pour des paires appariées, il converge vers 2β). La vraisemblance conditionnelle évite ce problème.

4.2 Modèle linéaire

Proposition 4.2. *Dans le modèle* $Y_{ij} = \alpha_i + \beta X_{ij} + \varepsilon_{ij}$, $\varepsilon_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$, *la vraisemblance conditionnelle sachant* $\bar{Y}_i$ *est, à une constante près,*

$$\log L_C(\beta, \sigma^2) = -\frac{N-K}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i,j} \{(Y_{ij} - \bar{Y}_i) - \beta(X_{ij} - \bar{X}_i)\}^2, \qquad N = \sum_i n_i,$$

et son maximiseur en β *est l'estimateur intra* $\hat{\beta}_W$.

Démonstration. Pour la grappe i, $\mathbf{y}_i \sim \mathcal{N}(\alpha_i \mathbf{1} + \beta \mathbf{X}_i, \sigma^2 I)$. Les vecteurs $P_i \mathbf{y}_i$ (qui ne dépend que de $\bar{Y}_i$) et $Q_i \mathbf{y}_i$ (écarts) sont des transformations linéaires orthogonales d'un vecteur gaussien à covariance sphérique, donc indépendants. Ainsi $f(\mathbf{y}_i \mid \bar{Y}_i)$ est la densité de $Q_i \mathbf{y}_i$, qui suit une loi normale dégénérée de moyenne $\beta Q_i \mathbf{X}_i$, de covariance $\sigma^2 Q_i$ et de rang $n_i - 1$; α_i a disparu puisque $Q_i \mathbf{1} = 0$. Sa log-densité sur l'image de Q_i est $-\frac{n_i - 1}{2} \log \sigma^2 - \|Q_i \mathbf{y}_i - \beta Q_i \mathbf{X}_i\|^2 / (2\sigma^2)$. En sommant sur i on obtient l'expression annoncée ; le maximum en β est atteint en W_{xy}/W_{xx}. □

On remarque que le maximiseur $\hat{\sigma}^2 = \mathrm{SC}_{\text{rés}}/(N-K)$ corrige le biais de l'estimateur du MV conjoint $\mathrm{SC}_{\text{rés}}/N$, non convergent quand $K \to \infty$ (l'estimateur exactement sans biais est $\mathrm{SC}_{\text{rés}}/(N-K-1)$, qui tient compte aussi de l'estimation de β) : la vraisemblance conditionnelle joue ici le rôle d'une vraisemblance restreinte pour les α_i. Les effets des covariables constantes dans la grappe (comme `matrace`) ne sont pas estimables : ils disparaissent avec α_i.

4.3 Modèle logistique

Proposition 4.3. *Dans le modèle* $\operatorname{logit} \mathbb{P}(Y_{ij} = 1 \mid \alpha_i, \mathbf{X}_i) = \alpha_i + \beta X_{ij}$ *avec indépendance conditionnelle, et pour* $s_i = \sum_j y_{ij}$,

$$\mathbb{P}\Big(\boldsymbol{Y}_i = \mathbf{y}_i \;\Big|\; \textstyle\sum_j Y_{ij} = s_i, \mathbf{X}_i\Big) = \frac{\exp\big(\beta \sum_j y_{ij} X_{ij}\big)}{\sum_{\mathbf{u} \in \mathcal{R}_i} \exp\big(\beta \sum_j u_j X_{ij}\big)},$$

où $\mathcal{R}_i = \{\mathbf{u} \in \{0,1\}^{n_i} : \sum_j u_j = s_i\}$ *contient les* $\binom{n_i}{s_i}$ *façons de répartir* s_i *succès dans la grappe.*

Démonstration. On a

$$\mathbb{P}(\boldsymbol{Y}_i = \mathbf{u} \mid \alpha_i) = \prod_j \frac{\exp\{u_j(\alpha_i + \beta X_{ij})\}}{1 + \exp(\alpha_i + \beta X_{ij})} = \frac{\exp(\alpha_i \sum_j u_j) \exp(\beta \sum_j u_j X_{ij})}{C_i},$$

où $C_i = \prod_j \{1 + \exp(\alpha_i + \beta X_{ij})\}$ ne dépend pas de $\mathbf{u}$. Alors

$$\mathbb{P}\Big(\textstyle\sum_j Y_{ij} = s_i \;\Big|\; \alpha_i\Big) = \sum_{\mathbf{u}\in\mathcal{R}_i} \frac{\exp(\alpha_i s_i)\exp(\beta \sum_j u_j X_{ij})}{C_i}.$$

Dans le rapport, C_i et $\exp(\alpha_i s_i)$ se simplifient. □

Conséquences :

— les grappes où $s_i = 0$ ou $s_i = n_i$ (toutes les réponses identiques) ont une contribution égale à 1 : elles n'apportent aucune information sur β ;

— seule la variation intra-grappe de X intervient : une covariable constante dans la grappe se simplifie ;

— l'évaluation de L_C ne demande aucune intégrale ; avec des grappes de cinq naissances, $\mathcal{R}_i$ compte au plus $\binom{5}{2} = 10$ éléments. Pour de grandes strates, des algorithmes récursifs évitent l'énumération (Gail, Lubin et Rubinstein, 1981). La maximisation par Newton–Raphson est rapide.

En R, `survival::clogit(y ~ x + strata(grappe))` maximise cette vraisemblance (la méthode par défaut, `method = "exact"`, calcule exactement la vraisemblance de la proposition 4.3 ; les options `"efron"` ou `"breslow"` en donnent des approximations plus rapides pour de grandes strates, exactes lorsque $s_i = 1$).

4.4 Régression logistique conditionnelle en cas-témoins appariés

La même vraisemblance est l'outil standard des études cas-témoins appariées (Breslow et Day, 1980, ch. 7). Chaque strate (paire) contient un cas ($Y = 1$) et un témoin ($Y = 0$) appariés sur des facteurs de confusion ; l'intercept de strate α_i représente ces facteurs. Avec une exposition binaire X, une paire 1:1 peut prendre quatre configurations :

I	$X=1$	$X=0$	Total	II	$X=1$	$X=0$	Total
Cas ($Y=1$)	1	0	1	Cas	1	0	1
Témoin ($Y=0$)	1	0	1	Témoin	0	1	1
III	$X=1$	$X=0$	Total	**IV**	$X=1$	$X=0$	Total
Cas	0	1	1	Cas	0	1	1
Témoin	1	0	1	Témoin	0	1	1

On résume l'échantillon de paires par le tableau croisé des expositions du cas et du témoin :

		Témoin exposé	
		1	0
Cas exposé	1	a_2 (type I)	a_1 (type II)
	0	b_1 (type III)	b_0 (type IV)

Conditionnons sur le nombre de cas et de témoins dans la paire ($m_1 = m_0 = 1$) et sur les expositions observées. Pour une paire discordante, dont un membre est exposé et l'autre non, les deux seules

configurations compatibles sont II et III, et la proposition 4.3 donne

$$\mathbb{P}(\text{II} \mid m_1 = 1, m_0 = 1, \{X\} = \{0,1\}) = \frac{\mathbb{P}(\text{II})}{\mathbb{P}(\text{II}) + \mathbb{P}(\text{III})} = \frac{e^{\beta}}{1 + e^{\beta}}.$$

En général, pour un cas et un témoin de covariables $\mathbf{x}^{(\text{cas})}$ et $\mathbf{x}^{(\text{témoin})}$, la contribution est

$$\frac{e^{\boldsymbol{\beta}^\top \mathbf{x}^{(\text{cas})}}}{e^{\boldsymbol{\beta}^\top \mathbf{x}^{(\text{cas})}} + e^{\boldsymbol{\beta}^\top \mathbf{x}^{(\text{témoin})}}}.$$

Les paires concordantes (types I et IV) contribuent un facteur 1. La vraisemblance conditionnelle est

$$L_C(\beta) = \Big(\frac{e^{\beta}}{1 + e^{\beta}}\Big)^{a_1} \Big(\frac{1}{1 + e^{\beta}}\Big)^{b_1},$$

qui ne dépend pas des intercepts de strate. En dérivant $a_1\beta - (a_1 + b_1)\log(1 + e^{\beta})$, on obtient $e^{\hat{\beta}} = a_1/b_1$: le rapport de cotes estimé est le rapport des nombres de paires discordantes des deux types. Une variance asymptotique est $\widehat{\text{Var}}(\hat{\beta}) = 1/a_1 + 1/b_1$, et le test du score de $H_0 : \beta = 0$ est le test de McNemar.

Proposition 4.4. *En appariement* 1:1*, l'estimateur du maximum de vraisemblance conditionnelle* a_1/b_1 *est égal à l'estimateur de Mantel–Haenszel du rapport de cotes commun, calculé en traitant chaque paire comme une strate.*

Démonstration. Pour la strate k, de tableau 2×2 d'effectifs $(A_k, B_k; C_k, D_k)$ (cas exposés, cas non exposés ; témoins exposés, témoins non exposés) et de taille $N_k = 2$, l'estimateur de Mantel–Haenszel est $\sum_k A_k D_k/N_k \,/\, \sum_k B_k C_k/N_k$. Une paire de type II a $A_k = D_k = 1$ et contribue $1/2$ au numérateur, rien au dénominateur ; une paire de type III a $B_k = C_k = 1$ et contribue $1/2$ au dénominateur ; les paires concordantes ont $A_k D_k = B_k C_k = 0$. Le rapport vaut donc $(a_1/2)/(b_1/2) = a_1/b_1$. □

Pour un appariement 1:M avec $M > 1$, les deux estimateurs diffèrent (tous deux restent convergents). Enfin, la validité de l'analyse conditionnelle pour un échantillonnage *rétrospectif* (on échantillonne selon Y) découle de ce que le modèle logistique avec intercept de strate est préservé par un tel échantillonnage, seuls les intercepts étant modifiés (Prentice et Pyke, 1979) ; or ceux-ci sont éliminés par le conditionnement.

Lien avec les données en grappes. Une mère de l'étude de Géorgie joue le rôle d'une strate : la régression logistique conditionnelle stratifiée par mère compare les naissances de faible poids et de poids normal d'une même mère, exactement comme une étude appariée compare un cas à ses témoins.

À retenir

- On élimine l'effet de grappe en conditionnant sur sa statistique exhaustive $\sum_j Y_{ij}$ (lien canonique). Aucune hypothèse n'est faite sur la loi de α_i ni sur sa relation avec les covariables.
- Linéaire : la vraisemblance conditionnelle donne l'estimateur intra. Logistique : la régression logistique conditionnelle ; seules les grappes discordantes en Y contribuent.
- Cas-témoins 1:1 : $e^{\hat{\beta}} = a_1/b_1$ = Mantel–Haenszel ; ce n'est plus vrai en 1:M.

5 Propriétés comparées des approches

Le tableau 1 résume les propriétés discutées par Neuhaus et McCulloch (2006) et dans les sections précédentes.

TABLE 1 – Comparaison des trois approches pour l'effet d'une covariable variant dans les grappes.

	GLMM standard	GLMM décomposé	Vraisemblance conditionnelle
Hypothèse $b_i \perp\!\!\!\perp \mathbf{X}_i$	requise	remplacée par une relation via $\bar{X}_i$	aucune
Loi de b_i	normale (spécifiée)	normale (spécifiée)	aucune
Robustesse de l'effet intra	biaisé si b_i corrélé avec $\mathbf{X}_i$	exacte (linéaire), approximative (GLMM)	exacte (lien canonique)
Effets de niveau grappe	estimables	estimables (β_B, sensible à la loi de b_i)	non estimables
Information utilisée	intra + inter	intra pour β_W	intra seulement (grappes discordantes)
Efficacité si le modèle standard est vrai	maximale	perte (intra seulement)	perte similaire
Calcul	intégrale (quadrature, Laplace)	intégrale	aucune intégrale ; rapide
Liens non canoniques	oui	oui	non (pas de statistique exhaustive)

Robustesse. L'approche conditionnelle est la plus robuste : elle ne suppose rien sur les α_i. L'approche par décomposition est robuste dans le modèle linéaire (proposition 3.1, pour toute loi de b_i) et approximativement robuste dans les GLMM pour l'effet intra. Les deux approches protègent l'effet intra à la fois contre la corrélation entre l'effet aléatoire et la covariable et contre une mauvaise spécification de la loi de l'effet aléatoire.

Efficacité. La robustesse a un prix : on renonce à l'information inter-grappes. Dans le modèle linéaire avec les hypothèses standards vraies, la proposition 2.1 donne $\mathrm{Var}(\hat{\beta}_{\mathrm{MCG}}) = \sigma^2/(W_{xx}+\theta B_{xx})$ et $\mathrm{Var}(\hat{\beta}_W) = \sigma^2/W_{xx}$, d'où

$$\frac{\mathrm{Var}(\hat{\beta}_{\mathrm{MCG}})}{\mathrm{Var}(\hat{\beta}_W)} = w.$$

La perte est faible quand la covariable varie surtout à l'intérieur des grappes ($w \approx 1$) ou que la corrélation intra-grappe est forte ($\theta \approx 0$), et importante dans le cas contraire. Dans un modèle logistique, la perte de l'approche conditionnelle est accentuée par l'élimination des grappes concordantes.

Calcul. La vraisemblance conditionnelle logistique s'évalue exactement sans intégration numérique ; la maximisation est rapide et stable. Les GLMM, standard ou décomposé, demandent une quadrature de Gauss–Hermite ou une approximation de Laplace (séance 3).

Quand préférer la décomposition ? Quand on veut aussi estimer des effets de niveau grappe (covariables constantes dans la grappe), des effets inter-grappes, des prédictions ou des effets

marginaux (qui demandent la loi de b_i), ou quand le lien n'est pas canonique. Quand on veut seulement l'effet intra le plus robuste possible, l'approche conditionnelle est préférable.

Attention : effet intra ne veut pas dire effet causal

L'effet intra est protégé contre la confusion par des facteurs *constants* dans la grappe, pas contre les facteurs qui varient dans la grappe. Dans l'exemple des naissances, l'âge de la mère augmente mécaniquement avec le rang de naissance : l'effet intra de l'âge confond l'effet de l'âge et celui de la parité (et de l'intervalle entre les naissances). L'interprétation de β_W exige donc la même réflexion sur la confusion qu'une analyse ordinaire, pour les facteurs qui varient d'une naissance à l'autre.

6 Comparer deux estimateurs : bootstrap non paramétrique par grappes

Une façon naturelle de juger si la corrélation entre l'effet aléatoire et la covariable est importante est de comparer $\hat{\beta}_{\text{std}}$ (modèle standard) et $\hat{\beta}_W$ (décomposé ou conditionnel). Sous l'hypothèse que le modèle standard est correct, les deux estimateurs sont convergents pour le même paramètre et leur différence tend vers 0 ; sinon elle reflète le biais du modèle standard. Pour juger de l'ampleur de la différence, il faut son erreur type. Or les deux estimateurs sont calculés sur les mêmes données et sont fortement corrélés : $\text{Var}(\hat{\beta}_{\text{std}} - \hat{\beta}_W) = \text{Var}(\hat{\beta}_{\text{std}}) + \text{Var}(\hat{\beta}_W) - 2\,\text{Cov}(\hat{\beta}_{\text{std}}, \hat{\beta}_W)$, et les logiciels ne fournissent pas la covariance.

Lien avec le test de Hausman. Si $\hat{\beta}_{\text{std}}$ est efficace sous H_0, on a $\text{Cov}(\hat{\beta}_{\text{std}}, \hat{\beta}_W) = \text{Var}(\hat{\beta}_{\text{std}})$ (un estimateur efficace n'est pas corrélé avec sa différence avec un autre estimateur convergent), d'où $\text{Var}(\hat{\beta}_W - \hat{\beta}_{\text{std}}) = \text{Var}(\hat{\beta}_W) - \text{Var}(\hat{\beta}_{\text{std}})$ (Hausman, 1978). Cette formule exige que le modèle standard soit entièrement correct, y compris la loi de b_i, ce qu'on cherche justement à remettre en question.

Bootstrap par grappes. Le bootstrap non paramétrique évite ces hypothèses. L'unité d'échantillonnage indépendante est la grappe, pas l'observation : il faut rééchantillonner les grappes entières pour préserver la corrélation intra-grappe.

Algorithme. Pour $r = 1, \ldots, R$:

1. tirer K grappes avec remise parmi les K grappes observées, en gardant toutes les observations de chaque grappe tirée ;
2. donner un *nouvel identifiant* à chaque grappe tirée (une grappe tirée deux fois doit compter comme deux grappes distinctes, sinon le modèle mixte la traiterait comme une seule grappe de taille double) ;
3. recalculer les moyennes $\bar{X}_i$ et les écarts, ajuster les deux modèles et calculer $D^{(r)} = \hat{\beta}^{(r)}_{\text{std}} - \hat{\beta}^{(r)}_W$.

L'erreur type bootstrap est l'écart type de $D^{(1)}, \ldots, D^{(R)}$; on peut former $z = D/\widehat{\text{ET}}_{\text{boot}}$ ou un intervalle percentile.

Le bootstrap demande $2R$ ajustements de modèles mixtes ; quelques centaines de répétitions suffisent

pour une erreur type. Il est prudent de compter les répétitions où l'ajustement n'a pas convergé. La question 4 des consignes applique cette procédure au modèle logistique.

7 Illustration en R : poids de naissance continu

Pour illustrer les notions sans faire l'analyse logistique demandée dans l'étude de cas, on étudie ici le poids de naissance *continu* en fonction de l'âge maternel avec des modèles linéaires mixtes.

```
library(lme4)
naisga <- read.csv("naisga.csv")
naisga <- subset(naisga, !is.na(mat_age))          # 13 âges manquants
naisga$age_moy <- ave(naisga$mat_age, naisga$matid) # moyenne par mère
naisga$age_dev <- naisga$mat_age - naisga$age_moy   # écart intra-mère
c(sd_inter = sd(tapply(naisga$mat_age, naisga$matid, mean)), sd_intra = sd(naisga$age_
    dev))
```

```
sd_inter sd_intra
3.603778 2.799686
```

Modèle standard, modèle décomposé et modèle de Mundlak.

```
m_std <- lmer(poidsnais ~ mat_age + (1 | matid), data = naisga)
m_dec <- lmer(poidsnais ~ age_dev + age_moy + (1 | matid), data = naisga)
m_mun <- lmer(poidsnais ~ mat_age + age_moy + (1 | matid), data = naisga)
summary(m_std)$coefficients; summary(m_dec)$coefficients; summary(m_mun)$coefficients
```

```
             Estimate Std. Error  t value
(Intercept) 2803.7307  31.821677 88.10757
mat_age       16.7391   1.399126 11.96397
             Estimate Std. Error   t value
(Intercept) 2508.28924  59.458620 42.185460
age_dev       11.87463   1.628319  7.292568
age_moy       30.39766   2.711486 11.210701
             Estimate Std. Error   t value
(Intercept) 2508.28924  59.458620 42.185460
mat_age       11.87463   1.628319  7.292568
age_moy       18.52304   3.162844  5.856449
```

Les écarts types estimés sont de 366 g pour l'effet de mère et de 428 g pour le résidu dans le modèle standard (363 et 428 g dans le modèle décomposé). Chez une même mère, une naissance survenue un an plus tard pèse en moyenne 11,9 g de plus ; entre mères, une différence d'un an d'âge moyen est associée à une différence de 30,4 g. Le modèle de Mundlak retrouve $\beta_W = 11{,}87$ et $\beta_B - \beta_W = 30{,}40 - 11{,}87 = 18{,}52$. Le modèle standard donne 16,7 g, entre les deux.

Vérifications des propositions. L'estimateur intra par effets fixes de mère, ou par régression des écarts, coïncide avec $\hat{\beta}_W$ du modèle décomposé (proposition 3.1) :

```
coef(lm(poidsnais ~ mat_age + factor(matid), data = naisga))["mat_age"]
moy <- aggregate(cbind(poidsnais, mat_age) ~ matid, data = naisga, FUN = mean)
coef(lm(poidsnais ~ mat_age, data = moy))["mat_age"]        # estimateur inter
```

```
 mat_age
11.87463
 mat_age
30.43671
```

Calculons le poids w de la proposition 2.1 avec les composantes de variance du modèle standard et des θ_i propres à chaque mère :

```
vc <- as.data.frame(VarCorr(m_std)); s2b <- vc$vcov[1]; s2 <- vc$vcov[2]
ni <- as.numeric(table(naisga$matid)); theta_i <- s2 / (s2 + ni * s2b)
xbar_i <- tapply(naisga$mat_age, naisga$matid, mean)
Wxx <- sum(naisga$age_dev^2)
Bxx <- sum(ni * theta_i * (xbar_i - mean(naisga$mat_age))^2)
w <- Wxx / (Wxx + Bxx)
c(w = w, combinaison = w * fixef(m_dec)[["age_dev"]] + (1 - w) * fixef(m_dec)[["age_moy
    "]])
```

```
         w combinaison
 0.7373869  16.7390205
```

La combinaison $0{,}737 \times 11{,}87 + 0{,}263 \times 30{,}40 = 16{,}739$ reproduit l'estimation du modèle standard (16,739) à la quatrième décimale près. Le biais du modèle standard par rapport à l'effet intra est donc $(1-w)(\hat\beta_B - \hat\beta_W) \approx 0{,}263 \times 18{,}5 \approx 4{,}9$ g par année. De plus, $(1{,}399/1{,}628)^2 = 0{,}738 \approx w$: le rapport des variances estimées est bien le poids w (section précédente).

Test de $\beta_W = \beta_B$ et bootstrap. Les modèles standard et décomposé sont emboîtés (contrainte $\beta_W = \beta_B$) ; on les compare par un test du rapport de vraisemblance avec des ajustements par ML, puisque les parties fixes diffèrent :

```
anova(update(m_std, REML = FALSE), update(m_dec, REML = FALSE))
```

```
                          npar    AIC    BIC logLik deviance  Chisq Df Pr(>Chisq)
update(m_std, REML = FALSE)    4 134669 134697 -67330   134661
update(m_dec, REML = FALSE)    5 134637 134672 -67313   134627 34.144  1  5.118e-09 ***
```

L'hypothèse $\beta_W = \beta_B$ est fortement rejetée. Le bootstrap par grappes de la différence $\hat\beta_{\text{std}} - \hat\beta_W$ ($R = 200$, 1770 mères tirées avec remise à chaque répétition) donne :

```
set.seed(2027)
ids <- unique(naisga$matid); par_mere <- split(naisga, naisga$matid)
diff_boot <- replicate(200, {
  tirage <- sample(ids, replace = TRUE)
  db <- do.call(rbind, lapply(seq_along(tirage), function(k) {
    dk <- par_mere[[as.character(tirage[k])]]; dk$grappe <- k; dk }))
  db$age_moy <- ave(db$mat_age, db$grappe); db$age_dev <- db$mat_age - db$age_moy
  fixef(lmer(poidsnais ~ mat_age + (1 | grappe), data = db))["mat_age"] -
    fixef(lmer(poidsnais ~ age_dev + age_moy + (1 | grappe), data = db))["age_dev"]
})
c(diff = 16.7391 - 11.87463, et_boot = sd(diff_boot))
```

```
    diff  et_boot
4.864470 0.941355
```

La différence de 4,86 g correspond à $z \approx 5{,}2$ erreurs types. L'erreur type de Hausman, $(1{,}628^2 - 1{,}399^2)^{1/2} = 0{,}83$, est un peu plus petite que l'erreur type bootstrap (0,94), ce qui est cohérent avec le fait qu'elle repose sur l'efficacité du modèle standard. (Une des 200 répétitions a produit un avertissement de convergence de `lmer`, sans conséquence ici.) La figure 1 représente les deux sources d'information.

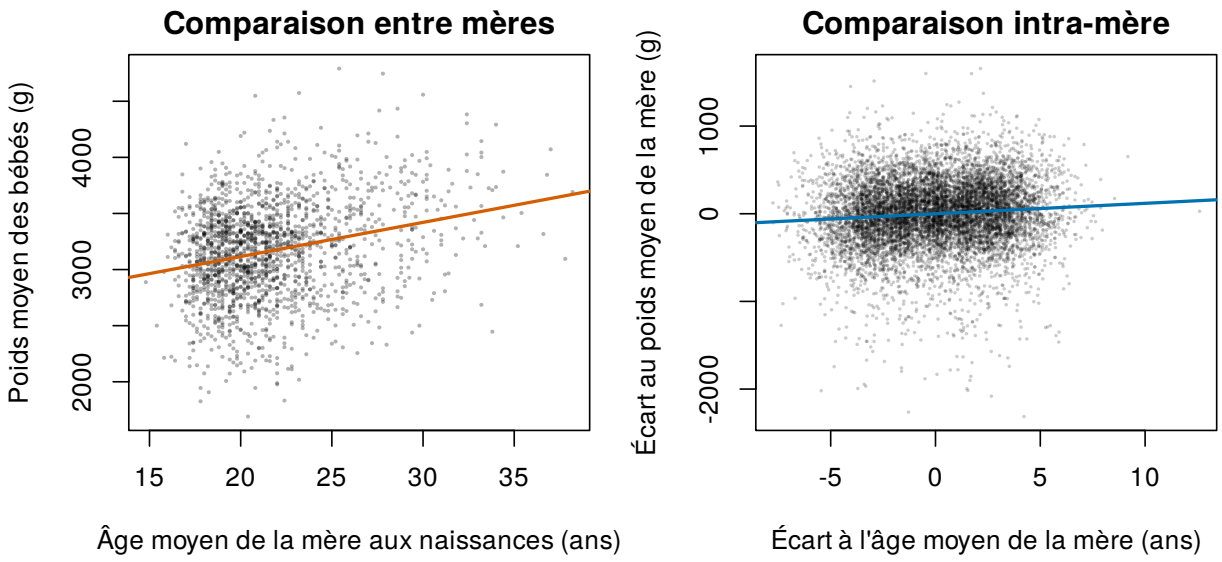


FIGURE 1 – Données `naisga`. Gauche : poids moyen des bébés en fonction de l'âge moyen de la mère (une observation par mère) ; la pente (30,4 g/an) est l'estimateur inter. Droite : écarts des poids à la moyenne de la mère en fonction des écarts d'âge ; la pente (11,9 g/an) est l'estimateur intra.

8 Illustration par simulation : cas logistique

Pour visualiser les propriétés du tableau 1 dans un modèle logistique, on simule $K = 300$ grappes de taille 5 selon $\operatorname{logit} \mathbb{P}(Y_{ij} = 1 \mid b_i) = -0{,}5 + 0{,}5 X_{ij} + b_i$, avec $X_{ij} = a\, b_i + \eta_{ij}$, $\eta_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, dans trois scénarios : (1) $a = 0$ et $b_i \sim \mathcal{N}(0{,}1{,}5^2)$ (hypothèses standards vraies) ; (2) $a = 0{,}6$ et b_i normal (covariable corrélée avec l'effet aléatoire) ; (3) $a = 0{,}6$ et $b_i = 1{,}5(E_i - 1)$ avec $E_i \sim$ exponentielle(1) (corrélation et loi asymétrique). Pour chacun, on ajuste le GLMM standard, le GLMM décomposé (coefficient de $X_{ij} - \bar{X}_i$) et la régression logistique conditionnelle, sur 150 jeux de données.

```
un_jeu <- function(scen, K = 300, ni = 5, beta = 0.5) {
  b <- if (scen == 3) (rexp(K) - 1) * 1.5 else rnorm(K, sd = 1.5)
  a <- if (scen == 1) 0 else 0.6
  id <- rep(1:K, each = ni); x <- a * b[id] + rnorm(K * ni)
  y <- rbinom(K * ni, 1, plogis(-0.5 + beta * x + b[id]))
  d <- data.frame(id, x, y); d$xm <- ave(d$x, d$id); d$xd <- d$x - d$xm
  c(standard  = fixef(glmer(y ~ x + (1 | id), family = binomial, data = d))["x"],
    decompose = fixef(glmer(y ~ xd + xm + (1 | id), family = binomial, data = d))["xd"],

    condition = coef(survival::clogit(y ~ x + strata(id), data = d)))
}
```

Moyennes et écarts types des 150 estimations (vraie valeur $\beta = 0{,}5$) :

```
Scenario 1          standard.x decompose.xd condition.x
           moyenne       0.506        0.505       0.508
           et            0.069        0.072       0.073
Scenario 2 moyenne       1.014        0.506       0.508
           et            0.073        0.081       0.083
Scenario 3 moyenne       0.971        0.508       0.514
           et            0.075        0.082       0.083
```

Quand les hypothèses standards sont vraies (scénario 1), les trois estimateurs sont sans biais, et le GLMM standard est un peu plus précis (écart type 0,069 contre 0,072 et 0,073) : c'est le prix de la robustesse. Dès que l'effet aléatoire est lié à la covariable (scénarios 2 et 3), le GLMM standard estime environ le double de la vraie valeur, alors que l'effet intra du modèle décomposé et l'estimateur conditionnel restent sans biais, y compris lorsque la loi de l'effet aléatoire est fortement asymétrique (scénario 3). L'estimation décomposée et l'estimation conditionnelle sont très proches l'une de l'autre, comme le soulignent Neuhaus et McCulloch (figure 2).

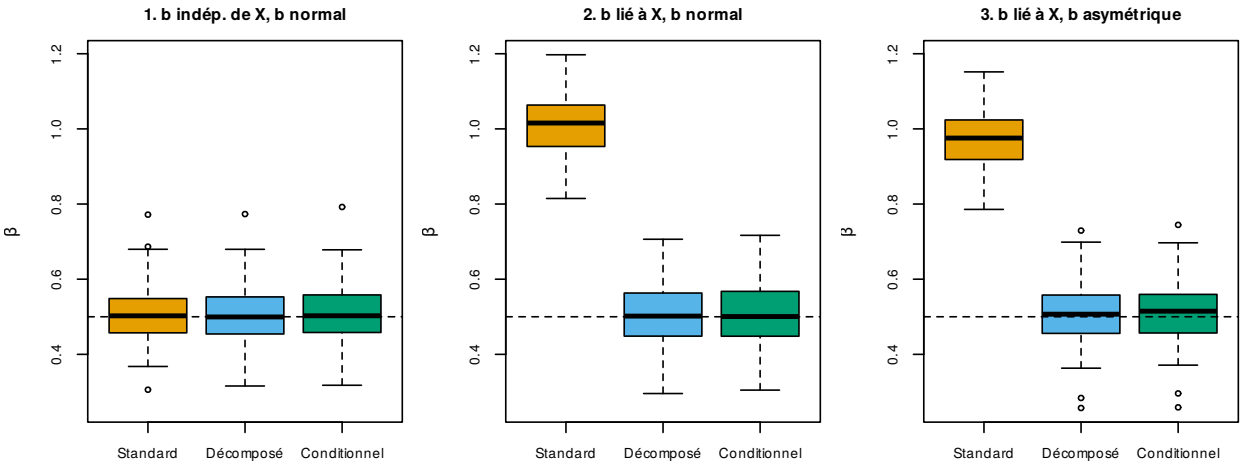


FIGURE 2 – Simulation (150 répétitions par scénario) : estimations de $\beta = 0{,}5$ (ligne pointillée) par le GLMM standard, le GLMM décomposé (effet intra) et la régression logistique conditionnelle.

Références

— Begg, M. D. et Parides, M. K. (2003). Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Statistics in Medicine*, 22, 2591–2602.

— Breslow, N. E. et Day, N. E. (1980). *Statistical Methods in Cancer Research, vol. 1 : The Analysis of Case-Control Studies*. CIRC, Lyon, ch. 7.

— Gail, M. H., Lubin, J. H. et Rubinstein, L. V. (1981). Likelihood calculations for matched case-control studies and survival studies with tied death times. *Biometrika*, 68, 703–707.

— Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46, 1251–1271.

— Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46, 69–85.

— Neuhaus, J. M. et Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, 54, 638–645.

— Neuhaus, J. M. et McCulloch, C. E. (2006). Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society B*, 68, 859–872.

— Prentice, R. L. et Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, 66, 403–411.

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

Synthèse après la séance

À lire après la séance : cet encadré correspond à la fiche des caractéristiques méthodologiques remplie pour l'article. Essayez d'abord de la remplir vous-même pendant l'étude de cas.

Synthèse après la séance : fiche méthodologique de Neuhaus et McCulloch (2006)

1. **Objet statistique d'intérêt** : l'effet intra-grappe β_W d'une covariable variant dans les grappes, dans un modèle conditionnel à la grappe (p. ex. log-rapport de cotes de faible poids de naissance par année d'âge maternel chez une même mère) ; secondairement l'effet inter β_B et la différence $\beta_B - \beta_W$.
   Justification : la comparaison intra-grappe est protégée contre la confusion par les caractéristiques constantes de la grappe, que l'on soupçonne d'être liées à la covariable.
2. **Paramètres nuisibles, sources de biais** : les effets de grappe b_i (ou intercepts α_i) et leur loi (σ_b^2) ; source de biais principale : la corrélation entre b_i et la covariable (confusion de niveau grappe) ; source secondaire : mauvaise spécification de la loi de b_i (normalité).
3. **Approche pour traiter la nuisance** : (a) conditionnement sur la statistique exhaustive $\sum_j Y_{ij}$ de l'effet de grappe, qui l'élimine ; (b) décomposition de la covariable en $\bar{X}_i$ et $X_{ij} - \bar{X}_i$, qui absorbe l'association dans $\beta_B \bar{X}_i$.
4. **Critère d'estimation** : (a) vraisemblance conditionnelle (régression logistique conditionnelle) ; (b) vraisemblance marginale du GLMM décomposé (maximum de vraisemblance).
   Justification : (a) aucune hypothèse sur la loi de b_i ni sur sa relation avec $\mathbf{X}_i$; (b) permet d'estimer aussi β_B et des effets de niveau grappe, et l'effet intra reste approximativement robuste.
5. **Méthode d'optimisation** : (a) Newton–Raphson sur la vraisemblance conditionnelle, évaluée exactement sans intégration (rapide) ; (b) maximisation de la vraisemblance du GLMM avec intégration numérique de l'effet aléatoire (quadrature de Gauss–Hermite adaptative ou Laplace).
6. **Obtention de la distribution des statistiques** : normalité asymptotique des estimateurs du maximum de vraisemblance (conditionnelle ou marginale), erreurs types par l'inverse de l'information observée ; tests de Wald ou du rapport de vraisemblance de $\beta_W = \beta_B$; bootstrap non paramétrique par grappes pour l'erreur type de la différence entre l'estimateur standard et l'estimateur intra.