Modélisation jointe de la précision et du temps de réponse à des tâches de reconnaissance chez la personne âgée

Barbara Azzopardi & Jacques Juhel

LPE/CRPCC (E.A. 1285), Université Rennes 2

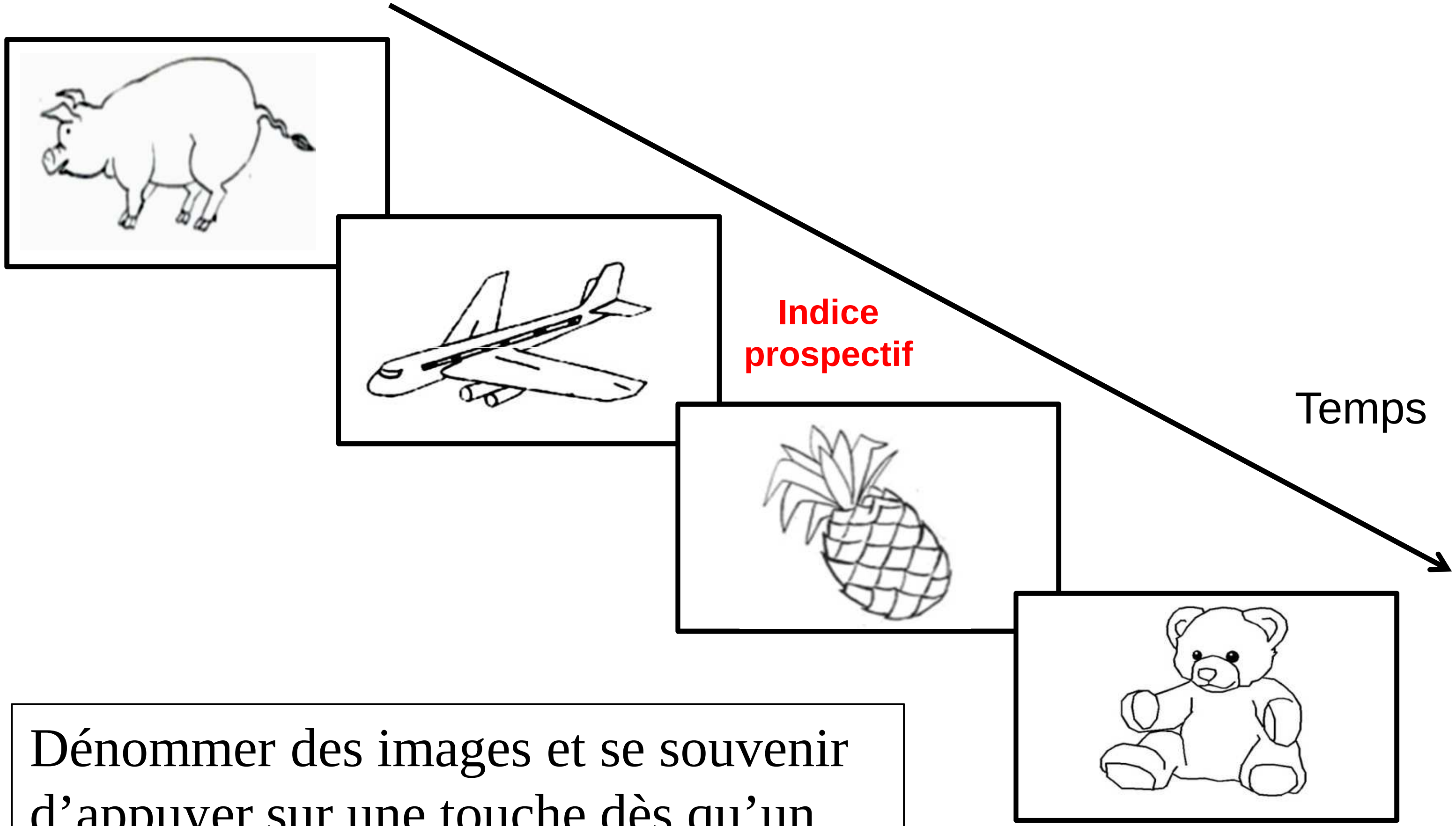
28-30 juin 2011

IIIème atelier Modeviia - Le Markstein

Objectifs de la recherche

- Mise en place d'un programme de remédiation cognitive visant à améliorer la performance de mémoire prospective chez la personne âgée en bonne santé.
- Objectif intermédiaire : identification des mécanismes pouvant sous-tendre le déclin des performances observé chez la personne âgée dans ce domaine.

Tâche *event-based* → basée sur l'évènement



Dénommer des images et se souvenir d'appuyer sur une touche dès qu'un fruit est présenté à l'écran.

Données analysées : caractéristiques de l'échantillon

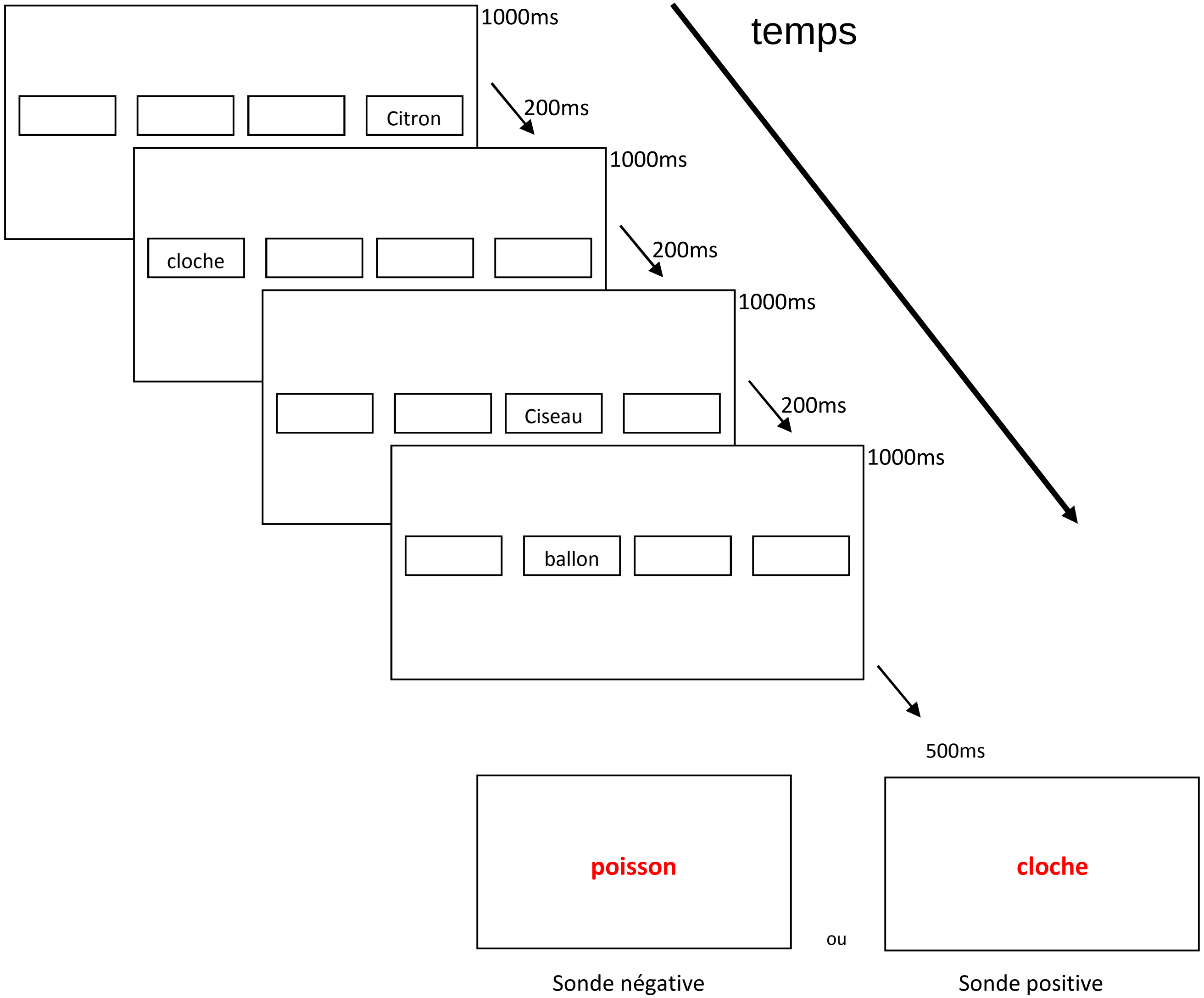
Participants

- 110 personnes âgées de 65 à 92 ans (M=74,30, E.T.=7.62) vivant à domicile ou en foyer-logement.
- Niveau d'études : 11 ans en moyenne.
- Score supérieur ou égal à 27 au *MMSE* (M=29.22; E.T.= 0.82)

Matériel et procédure

- Epreuves évaluant le *binding* (adaptées d'Oberauer, 2005)
  - Sternberg modifiée
  - Reconnaissance globale
  - Reconnaissance locale
- Deux épreuves *event-based* de mémoire prospective.

Matériel et procédure



Epreuve de reconnaissance globale (adaptée d'Oberauer, 2005).

Question posée

- Les DI d'efficacité dans les processus de couplage rendent-elles compte des DI dans le domaine de la mémoire prospective?
-
  - Premières analyses réalisées sur la base de calcul d'indices de discrimination: absence de prise en compte de la variabilité
  - Autres analyses: modèles à coefficients aléatoires

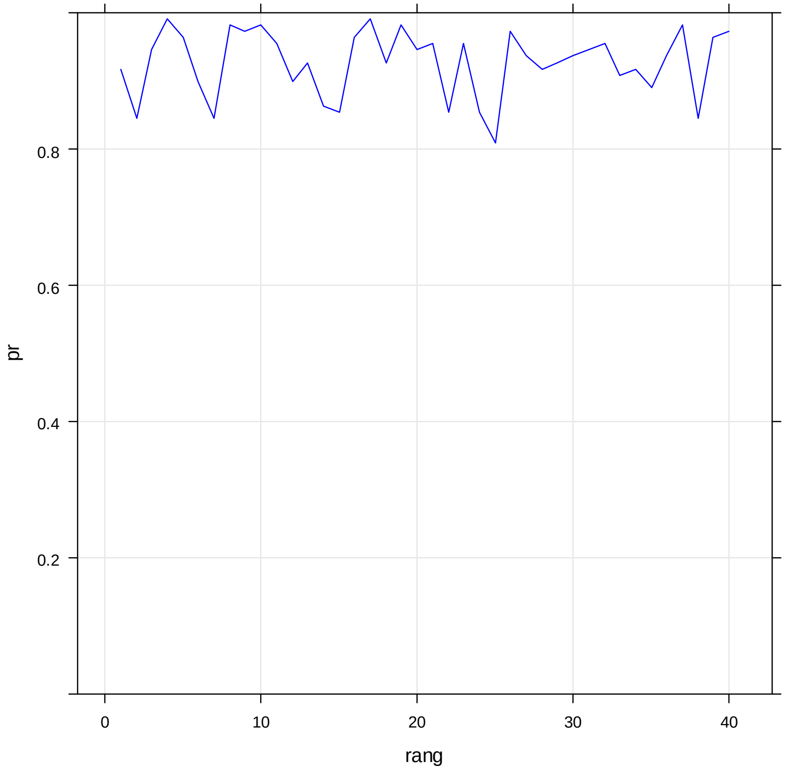
Position du problème

- Analyse des effets de manipulations expérimentales théoriquement « fondées » sur la performance de personnes âgées : inférences *subject-specific*.
- Plusieurs mesures répétées au cours du temps : précision (PR) et temps de réponse (TR), nombre *réduit* d'essais.
- Assez classiquement : modélisation de la précision d'une part, du temps de réponse d'autre part.
- Modéliser *conjointement* les relations entre les mesures effectuées (binaires et continues) pour extraire des mesures individuelles mises ensuite en relation avec des variables distales.

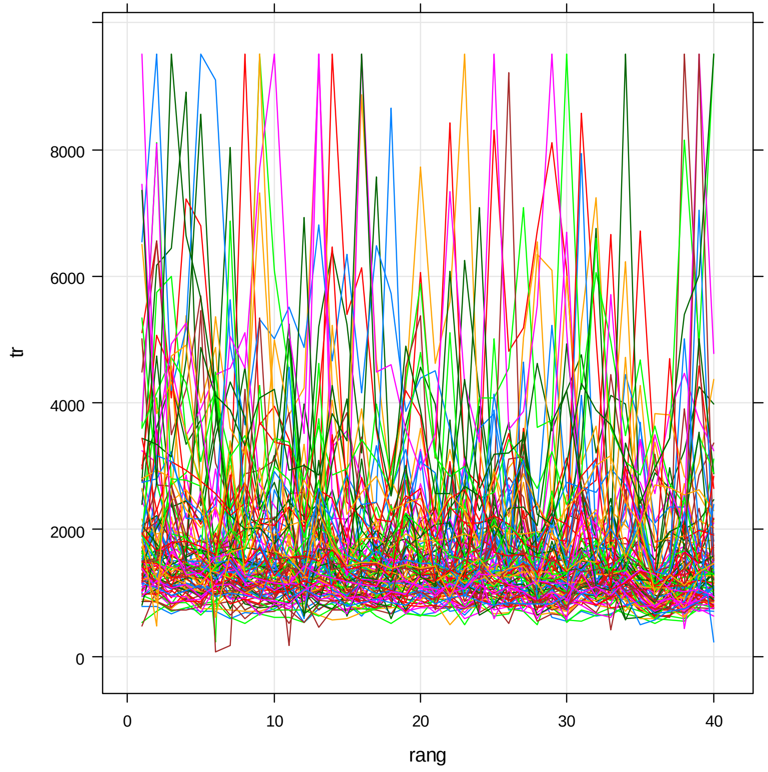
Performance à la tâche de reconnaissance globale: données binaires et continues

Evolution de la précision moyenne et du temps de réponse individuel au cours des 40 essais (N=110).

Précision moyenne



Temps de réponse (msec)



1
Modélisation de données répétées continues

Modèles pour données répétées

Vecteur de *n* réponses $\boldsymbol{Y}_{ij} = (Y_{i1}, Y_{i2}, \ldots, Y_{in})'$ du sujet *i*.

• Modèle *marginal* : modélisation des distributions marginales des composantes de ***Y*** à partir d'un ensemble ***X*** de variables explicatives (temps, antécédents individuels, caractéristiques situationnelles).

• Modèle *à effets aléatoires* : variables explicatives de ***X*** + vecteur aléatoire $\boldsymbol{\upsilon}$ (vecteur d'effets aléatoires non observés propres à chaque sujet).

⇨ Interprétation *hiérarchique* : l'interprétation des paramètres des effets fixes est conditionnée par un niveau constant du paramètre d'effet aléatoire.

1) Données continues : modèle linéaire mixte ;

2) Données discrètes : modèle mixte linéaire généralisé.

Modèle linéaire mixte (LMM)

Modèle linéaire classique pour données répétées :

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \varepsilon_{ij}$$

Modèle linéaire mixte : addition d'effet(s) aléatoire(s).

- Modèle à intercept aléatoire :

$$Y_{ij} = (\beta_0 + \upsilon_{i0}) + \beta_1 t_{ij} + \varepsilon_{ij}$$

υ_{i0} écart individuel à l'intercept moyen (après prise en compte des effets des covariables).

- Modèle à intercept et pente aléatoires

$$Y_{ij} = (\beta_0 + \upsilon_{i0}) + (\beta_1 + \upsilon_{i1}) t_{ij} + \varepsilon_{ij}$$

➢ effets aléatoires υ_{i0} et υ_{i1} de moyenne nulle et normalement distribués.

Modèle linéaire mixte (LMM)

Représentation multi-niveaux du LMM à intercept aléatoire

Les mesures répétées (niveau 1) sont emboîtées dans chaque individu (niveau 2).

- Intra-sujets ou niveau 1 :

$$y_{ij} = b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

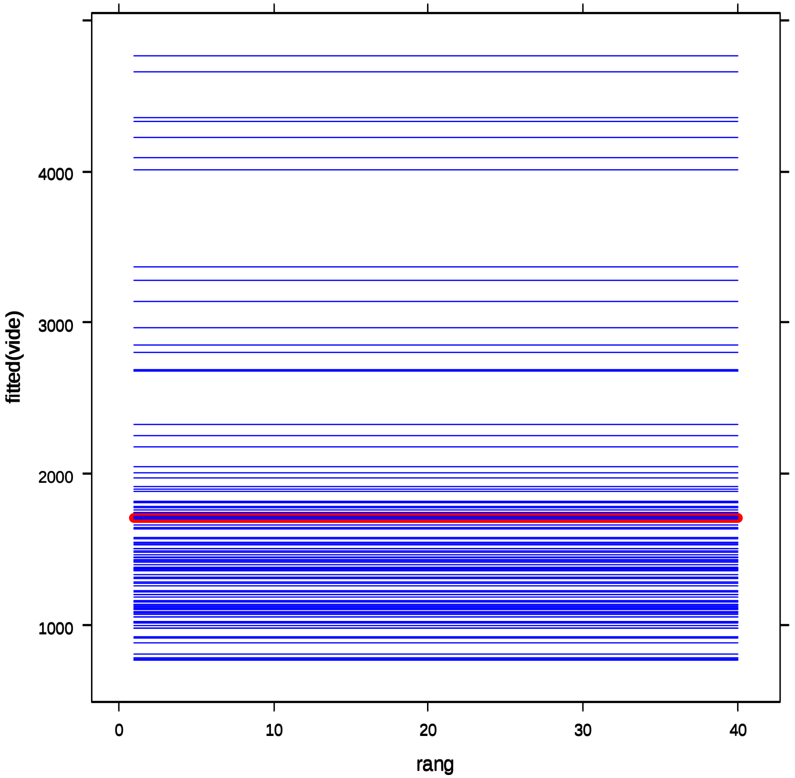
- Inter-sujets ou niveau 2 :

$$b_{0i} = \beta_0 + \upsilon_{i0},$$

$$b_{1i} = \beta_1,$$

$$\upsilon_{i0} \sim \mathcal{N}(0, \sigma_\upsilon^2)$$

ε_{ij} et υ_{i0} indépendants.



Modèle linéaire mixte (LMM)

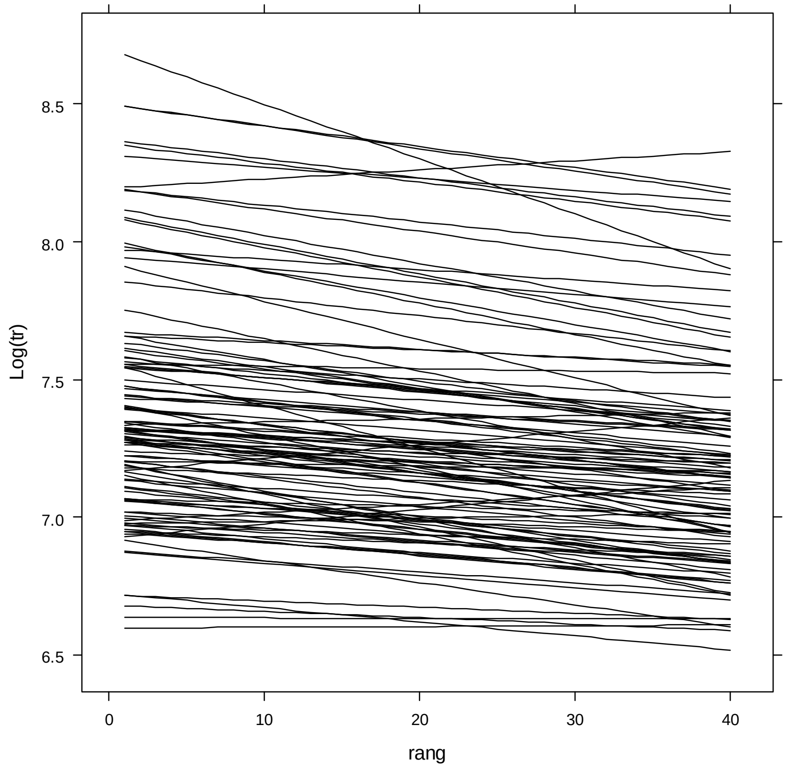
Représentation multi-niveaux du LMM à intercept et pente aléatoires

- Intra-sujets ou niveau 1 :

$$y_{ij} = b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij}$$
$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

- Inter-sujets ou niveau 2 :

$$b_{0i} = \beta_0 + \upsilon_{i0},$$
$$b_{1i} = \beta_1 + \upsilon_{i1},$$
$$\mathcal{N}(0, \Sigma_\upsilon) \quad \Sigma_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_0 \upsilon_1} \\ \sigma_{\upsilon_1 \upsilon_0} & \sigma^2_{\upsilon_1} \end{bmatrix}$$



Modèle linéaire mixte (LMM)

Formulation matricielle du LMM à intercept aléatoire

$$\boldsymbol{Y}_{ij} = \boldsymbol{X}_{ij}\beta + \boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij}$$

- Profil de réponse moyen :

$$E(\boldsymbol{Y}_{ij}) = \mu_{ij} = \boldsymbol{X}_{ij}\beta$$

- Trajectoires individuelles :

$$E(\boldsymbol{Y}_{ij} \mid \boldsymbol{v}_i) = \boldsymbol{X}_{ij}\beta + \boldsymbol{v}_i$$

Modèle linéaire mixte (LMM)

LMM à intercept aléatoire : coefficient intra-classe

$$\begin{aligned}\mathrm{Var}(\boldsymbol{y}_{ij}) &= \mathrm{Var}(\boldsymbol{X}_{ij}\beta + \boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij}) \\ &= \mathrm{Var}(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij}) \\ &= \mathrm{Var}(\boldsymbol{v}_i) + \mathrm{Var}(\boldsymbol{\varepsilon}_{ij}) \\ &= \sigma_v^2 + \sigma^2.\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{y}_{ij}, \boldsymbol{y}_{ik}) &= \mathrm{Cov}(\boldsymbol{X}_{ij}\beta + \boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij}, \boldsymbol{X}_{ik}\beta + \boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ik}) \\ &= \mathrm{Cov}(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij}, \boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ik}) \\ &= \mathrm{Cov}(\boldsymbol{v}_i, \boldsymbol{v}_i) \\ &= \mathrm{Var}(\boldsymbol{v}_i) \\ &= \sigma_v^2.\end{aligned}$$

$$\mathrm{ICC} = \mathrm{Corr}(\boldsymbol{y}_{ij}, \boldsymbol{y}_{ik}) = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2}$$

ICC : degré d’association intra des données, proportion de variance attribuable aux individus.

Modèle linéaire mixte (LMM)

<u>Généralisation</u> (Laird & Ware, 1982) :

$$\boldsymbol{Y}_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}\boldsymbol{\upsilon}_i + \boldsymbol{\varepsilon}_i,$$
$$\boldsymbol{\upsilon}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\upsilon),$$
$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{\Omega}_i)$$

$\boldsymbol{\beta}$: vecteur de paramètres inconnus des effets fixes et $\boldsymbol{X}_{ij}$ sa matrice d'incidence (fait correspondre à chaque observation la combinaison des effets fixes qui la concerne),

$\boldsymbol{\upsilon}_i$: vecteur de paramètres inconnus des effets aléatoires et $\boldsymbol{Z}_{ij}$ sa matrice d'incidence (associe à chaque observation l'effet aléatoire de l'individu qui a fourni l'observation).

$\boldsymbol{\varepsilon}_i$: vecteur aléatoire d'erreurs.

SAS : $\boldsymbol{\upsilon}_i \sim \mathcal{N}(0, G)$, $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, R)$.

Modèle linéaire mixte (LMM)

Sous ces hypothèses :

moyenne conditionnelle $E(\boldsymbol{Y}_i \mid \boldsymbol{v}_i) = \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{v}_i$

variance conditionnelle $Var(\boldsymbol{Y}_i \mid \boldsymbol{v}_i) = \boldsymbol{R}_i = \sigma^2 \mathrm{I}_{n_i}$

moyenne marginale $E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\beta$

variance marginale $Var(\boldsymbol{Y}_i) = \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}_i^{'} + \boldsymbol{R}_i$

La variance de $\boldsymbol{Y}_i$ est modélisée en spécifiant $\boldsymbol{Z}_i$ et les matrices de covariance $\mathbf{G}$ (effets aléatoires, inter) et $\mathbf{R}_i$ (erreurs de mesure, intra).

E($\boldsymbol{Y}_i$)=$\boldsymbol{X}_i\beta$ => les effets fixes peuvent être interprétés marginalement.

Modèle linéaire mixte (LMM)

Estimation des paramètres

L'inférence est basée sur le modèle marginalisé de $\boldsymbol{Y}_i$ et s'obtient par intégration : la distribution marginale peut être obtenue analytiquement.

Estimation des effets fixes β et des composantes de la variance σ^2 et $\boldsymbol{\Sigma}_\upsilon$ par :

ML : maximisation de la vraisemblance des données d'échantillon (ajustement des effets fixes et aléatoires du modèle).

REML : maximisation de la vraisemblance des résidus d'échantillon (après estimation par MCO ou généralisés des effets fixes) afin de corriger la sous-estimation des effets aléatoires par ML.

Prédictions spécifiques à chaque personne : les estimations bayésiennes empiriques des effets aléatoires ou BLUP (*Best Linear Unbiased Predictor*) sont dérivées de la moyenne ou du mode de la distribution *a posteriori* de $\boldsymbol{\upsilon}_i$ sachant $\boldsymbol{y}_i$.

Modèle linéaire mixte (LMM)

Stratégie de modélisation

• Choix d'un modèle « maximal » de la moyenne (sur la base du modèle saturé si c'est possible).

• Modélisation de la structure de covariance (en commençant par une matrice non structurée pour G et R=σ^2I).

• Comparaison de modèles s'appuyant sur :

Déviance = indice de « mauvais » ajustement

$$\text{Déviance} = -2\left[LL_{\text{modèle estimé}} - LL_{\text{modèle saturé}}\right]$$

Critères pénalisant le log de la vraisemblance par le nombre de paramètres introduits : AIC = deviance + 2 (#par) et BIC = deviance + ln N (#par).

Test de différence de déviance :

chi-deux = $Dev_{H0} - Dev_{H1}$, ddl = $\#par_{H0} - \#par_{H1}$.

LMM appliqué aux TR

<u>Illustration</u> : log(tr) : modèle à intercept aléatoire

Code SAS

```
PROC MIXED DATA=nom METHOD=REML;
CLASS id;
MODEL tr = / SOLUTION S;
RANDOM INTERCEPT / VCORR G TYPE=UN SUBJECT=id S;
RUN;
```

ICC = 0,543

BIC = 1242.3

NB – Option REPEATED pour modéliser des résidus corrélés ou l’hétérogénéité de la variance.

LMM appliqué aux TR

Illustration : log(tr) : modèle à intercept aléatoire, tous les effets fixes

Code SAS

```
PROC MIXED DATA=nom METHOD=REML;
CLASS id;
MODEL tr = rang age mat rang*mat rang*age age*mat
rang*age*mat / SOLUTION S;
RANDOM INTERCEPT / VCORR G TYPE=UN SUBJECT=id S;
RUN;
```

BIC = 1072

LMM appliqué aux TR

Illustration : log(tr) : tous les effets fixes sauf rang*age rang*age*mat, intercept aléatoire.

Code SAS

```
PROC MIXED DATA=RGmultivar NOCLPRINT NOITPRINT
COVTEST METHOD=REML;
class id ;
MODEL tr = rang age mat rang*mat age*mat
/SOLUTION S ;
RANDOM  intercept/ TYPE=un SUBJECT=id G S;
RUN;
```

BIC = 1045

LMM appliqué aux TR

Illustration : log(tr) : tous les effets fixes sauf rang*age rang*age*mat age*mat, intercept aléatoire.

Code SAS

```
PROC MIXED DATA=RGmultivar NOCLPRINT NOITPRINT
COVTEST METHOD=REML;
class id ;
MODEL tr = rang age mat rang*mat /SOLUTION S ;
RANDOM  intercept/ TYPE=un SUBJECT=id G S;
RUN;
```

BIC = 1034

LMM appliqué aux TR

Illustration : log(tr) : tous les effets fixes sauf rang*age rang*age*mat,age*mat, intercept et pente_rang aléatoires.

Code SAS

```
PROC MIXED DATA=RGmultivar NOCLPRINT NOITPRINT
COVTEST METHOD=REML;
class id;
MODEL tr = rang age mat rang*mat /SOLUTION S ;
RANDOM  intercept rang / TYPE=un SUBJECT=id G S;
RUN;
```

BIC = 999

LMM appliqué aux TR

Illustration : log(tr) : tous les effets fixes sauf rang*age rang*age*mat,age*mat, intercept, pente_rang et pente_mat aléatoires.

Code SAS

```
PROC MIXED DATA=RGmultivar NOCLPRINT NOITPRINT
COVTEST METHOD=REML;
class id;
MODEL tr = rang age mat rang*mat/SOLUTION S ;
RANDOM  intercept rang mat / TYPE=un SUBJECT=id G
S;
RUN;
```

BIC = 932

LMM appliqué aux TR

<u>Illustration</u> : log(tr) : tous les effets fixes sauf rang*age rang*age*mat,age*mat, intercept, pente_rang, pente_mat et pente_rang*mat aléatoires.

Code SAS

```
PROC MIXED DATA=RGmultivar NOCLPRINT NOITPRINT
COVTEST METHOD=REML;
class id;
MODEL tr = rang age mat rang*mat/SOLUTION S ;
RANDOM  intercept rang mat rang*mat / TYPE=un
SUBJECT=id G S;
RUN;
```

BIC = 916

LMM appliqué aux TR

Illustration : log(tr) : tous les effets fixes sauf rang*age rang*age*mat,age*mat et mat; intercept, pente_rang, pente_mat et pente_rang*mat aléatoires.

Code SAS

```
PROC MIXED DATA=RGmultivar NOCLPRINT NOITPRINT
COVTEST METHOD=REML;
class id ;
MODEL tr = rang age rang*mat /SOLUTION S ;
RANDOM  intercept rang mat rang*mat / TYPE=un
SUBJECT=id G S;
RUN;
```

BIC = 913

LMM appliqué aux TR

The Mixed Procedure

Informations sur le modèle

Data Set	WORK.RGMULTIVAR
Dependent Variable	tr
Covariance Structure	Unstructured
Subject Effect	id
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Dimensions

Covariance Parameters	11
Columns in X	4
Columns in Z Per Subject	4
Subjects	110
Max Obs Per Subject	40
Observations Used	4193
Observations Not Used	207
Total Observations	4400

LMM appliqué aux TR

Covariance Parameter Estimates

Cov Parm	Subject	Estimation	Erreur type	Z Value	Pr Z
UN(1,1)	id	0.1034	0.01638	6.31	<.0001
UN(2,1)	id	-0.00068	0.000269	-2.54	0.0112
UN(2,2)	id	0.000022	6.562E-6	3.35	0.0004
UN(3,1)	id	-0.01947	0.009041	-2.15	0.0313
UN(3,2)	id	0.000225	0.000181	1.24	0.2140
UN(3,3)	id	0.03190	0.008186	3.90	<.0001
UN(4,1)	id	0.000618	0.000342	1.80	0.0711
UN(4,2)	id	-0.00001	7.442E-6	-1.93	0.0535
UN(4,3)	id	-0.00083	0.000287	-2.87	0.0040
UN(4,4)	id	0.000037	0.000012	3.09	0.0010
Residual		0.06013	0.001389	43.29	<.0001

$\boldsymbol{G}$

$\boldsymbol{R} = \sigma^2 \boldsymbol{I}$

Fit Statistics

-2 Res Log Likelihood	861.3
AIC (smaller is better)	883.3
AICC (smaller is better)	883.4
BIC (smaller is better)	913.0

LMM appliqué aux TR

Solution for Fixed Effects

Effet	Estimation	Erreur type	DDL	Valeur du test t	Pr > |t|
Intercept	5.4932	0.2649	109	20.74	<.0001
rang	-0.00300	0.000593	109	-5.06	<.0001
age	0.02479	0.003536	3752	7.01	<.0001
rang*mat	-0.00239	0.000516	109	-4.63	<.0001

Solution for Random Effects

Effet	id	Estimation	Std Err Pred	DDL	Valeur du test t	Pr > |t|
Intercept	1	-0.5543	0.08978	3752	-6.17	<.0001
rang	1	0.002263	0.003303	3752	0.68	0.4934
mat	1	0.01842	0.1131	3752	0.16	0.8706
rang*mat	1	0.003691	0.004452	3752	0.83	0.4072
Intercept	23	0.3027	0.09191	3752	3.29	0.0010
rang	23	-0.00203	0.003303	3752	-0.61	0.5392
mat	23	-0.2653	0.1131	3752	-2.35	0.0190
rang*mat	23	0.007150	0.004453	3752	1.61	0.1084

LMM appliqué aux TR

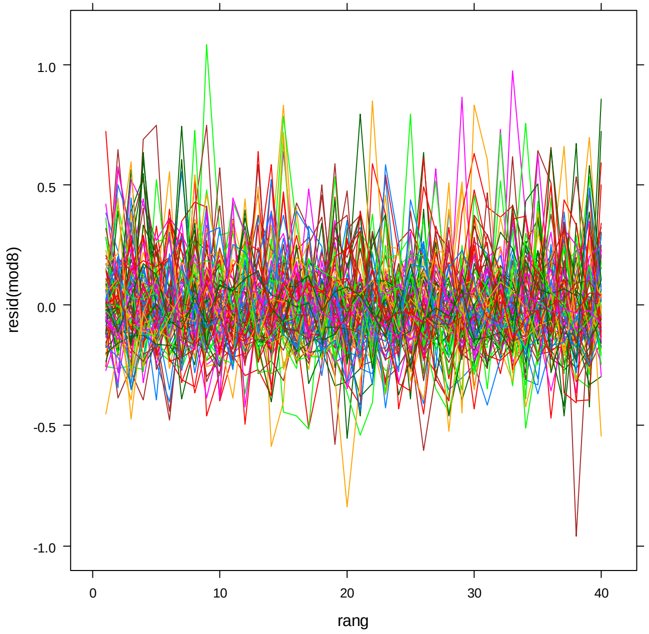
Sortie R

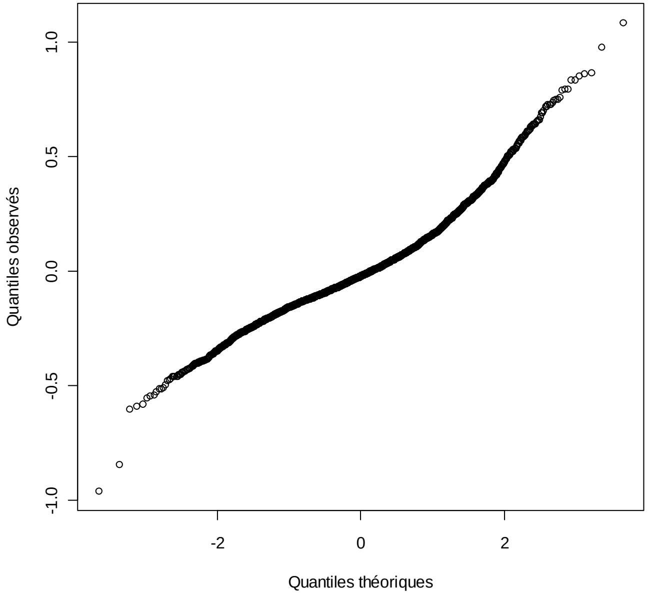
```
Linear mixed model fit by REML
Formula: tr ~ age + rang + rang:mat  + (rang | sujet) + (mat |
sujet) + (rang:mat | sujet)
   Data: mydata
   AIC  BIC logLik deviance REMLdev
 913.2 1021 -439.6    837.8   879.2
Random effects:
 Groups   Name        Variance   Std.Dev.   Corr
 sujet    (Intercept) 1.1303e-02 0.10631669
          rang:mat0   1.8585e-05 0.00431098 -0.914
          rang:mat1   1.7146e-05 0.00414073 -0.472  0.790
 sujet    (Intercept) 4.8664e-02 0.22059976
          mat1        1.1576e-02 0.10759102 -0.470
 sujet    (Intercept) 4.1264e-02 0.20313439
          rang        7.8824e-07 0.00088783 -0.983
 Residual             6.1029e-02 0.24704022
Number of obs: 4193, groups: sujet, 110

Fixed effects:
              Estimate Std. Error t value
(Intercept)  5.4955137  0.2652198  20.721
age          0.0248131  0.0035407   7.008
rang        -0.0030179  0.0005941  -5.080
rang:mat1   -0.0026426  0.0005258  -5.026

Correlation of Fixed Effects:
          (Intr) age    rang
age       -0.994
rang      -0.048  0.003
rang:mat1  0.010 -0.004 -0.499
```

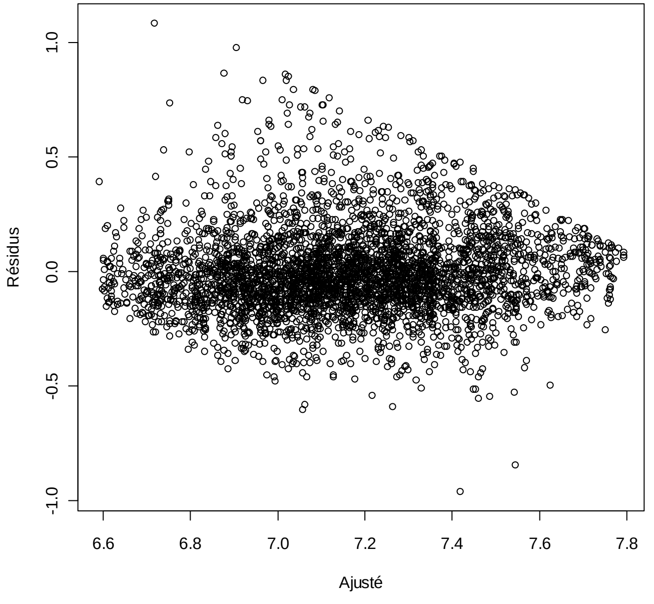
LMM appliqué aux TR

Diagnostics

Données: RG_lntr_cleaned







LMM appliqué aux TR

Interprétation :

Diminution du TR au cours du temps.

TR plus élevé pour les items positifs que pour les items négatifs.

Augmentation du TR avec l'âge.

Diminution du TR au cours du temps plus importante pour les items positifs que négatifs.

DI dans le niveau initial de réponse, dans l'évolution des TR au cours du temps et dans la sensibilité à la manipulation expérimentale.

2
Modélisation de données répétées binaires

Rappel : modèles pour données binaires

Deux fonctions de répartition classiquement employées dans la modélisation d'une variable binaire :

Loi logistique $F(x) = \dfrac{e^x}{1+e^x}$, moy.=0, var.=$\dfrac{\pi^2}{3}$,

fonction inverse: $F^{-1}(t) = \log(\dfrac{t}{1-t})$ (fonction logit)

Loi normale centrée réduite $F(x) = \Phi(x) = \int_{-\infty}^{x} \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$

fonction inverse: $F^{-1}(t) = \Phi^{-1}(t)$ (fonction probit)

Rappel : modèles pour données binaires

Modèle de régression logistique

$$\pi_i = \Pr(\boldsymbol{Y}_i = 1) = \frac{e^{\boldsymbol{X}_i \beta}}{1 + e^{\boldsymbol{X}_i \beta}},$$

$$\pi_i = \frac{1}{1 + e^{-\boldsymbol{X}_i \beta}} = \Psi(\boldsymbol{X}_i \beta)$$

$\Psi(.)$ fonction de distribution cumulative logistique,

$$\text{logit} = \log\left[\frac{\pi_i}{1 - \pi_i}\right] = \boldsymbol{X}_i \beta.$$

Modèle de régression probit

$$\pi_i = \Pr(\boldsymbol{Y}_i = 1) = \Phi(\boldsymbol{X}_i \beta),$$

$\Phi(.)$ fonction de distribution cumulative normale.

$$\Phi^{-1}(\pi_i) = \boldsymbol{X}_i \beta$$

Rappel : modèles pour données binaires

<u>Formulation en modèle à variable de réponse latente</u> :

Seuil τ sur une variable continue y^* telle que :

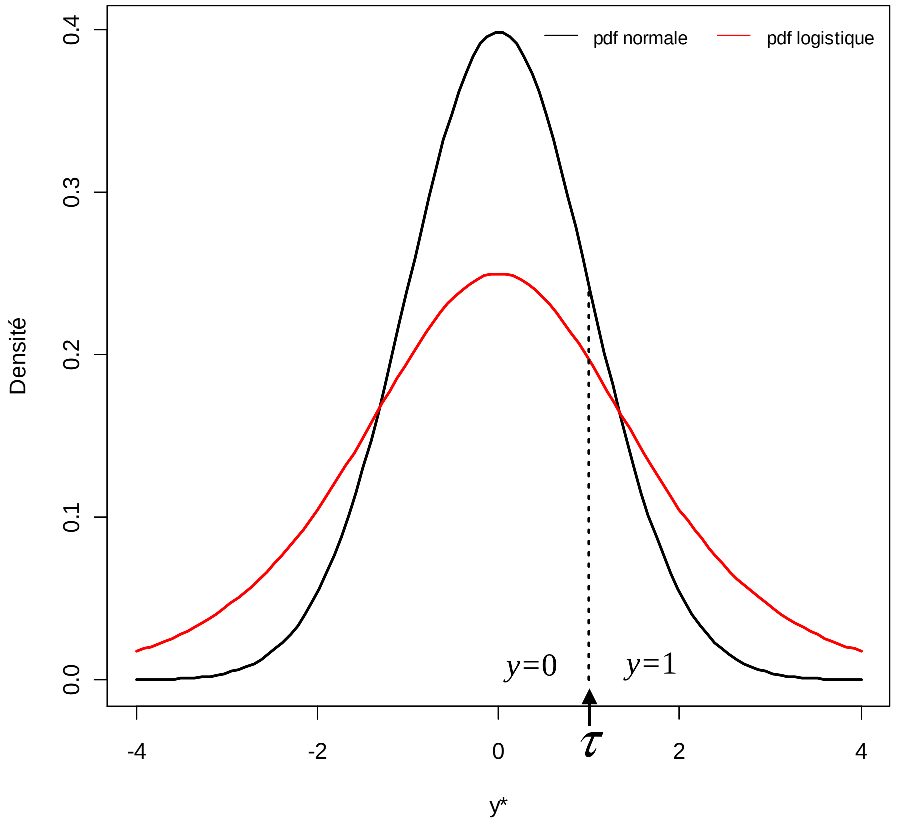
$$\begin{cases} y = 1 \text{ si } y^* \geq \tau \\ y = 0 \text{ si } y^* < \tau \end{cases}$$

$$y_i^* = \boldsymbol{X}_i \beta + \varepsilon_i$$

Logistique : $\varepsilon_i \sim \mathcal{L}(0, \frac{\pi^2}{3})$,

Probit : $\varepsilon_i \sim \mathcal{N}(0,1)$.

y_i^* n'étant pas observée, l'échelle des erreurs est fixée (i.e., non estimée).



Modèle mixte linéaire généralisé (GLMM) pour données binaires

Introduction des effets aléatoires dans l'expression du prédicteur linéaire des modèles de régression logistique ou probit.

$$\eta_i = \boldsymbol{X}_{ij}\beta + \boldsymbol{Z}_{ij}\boldsymbol{v}_i, \ \boldsymbol{v}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma}_v)$$

GLMM logistique

$$\pi_{ij} = \Pr(\boldsymbol{Y}_{ij} = 1 | \boldsymbol{v}_i) = \frac{1}{1 + e^{-(\boldsymbol{X}_{ij}\beta + \boldsymbol{Z}_{ij}\boldsymbol{v}_i)}}$$

GLMM probit

$$\pi_{ij} = \Pr(\boldsymbol{Y}_{ij} = 1 | \boldsymbol{v}_i) = \Phi(\boldsymbol{X}_{ij}\beta + \boldsymbol{Z}_{ij}\boldsymbol{v}_i)$$

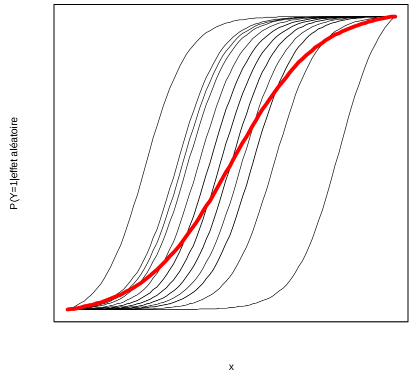
Modèle mixte linéaire généralisé (GLMM) pour données binaires

Interprétation des paramètres

Le vecteur β de paramètres inconnus des effets fixes doit être interprété conditionnellement aux effets aléatoires (interprétation subject-specific) L'interprétation marginale des paramètres n'est en général pas possible.

Moyenne conditionnée par les effets aléatoires :

$$E(\boldsymbol{Y}_{ij} | \boldsymbol{v}_i) = P(\boldsymbol{Y}_{ij} = 1 | \boldsymbol{X}_i, \boldsymbol{Z}_i \beta, \boldsymbol{v}_i) = \frac{e^{\boldsymbol{X}_{ij}\beta + \boldsymbol{Z}_{ij}\boldsymbol{v}_i}}{1 + e^{\boldsymbol{X}_{ij}\beta + \boldsymbol{Z}_{ij}\boldsymbol{v}_i}}$$



Moyenne au niveau de la population déduite du modèle mixte :

$$E\left[\boldsymbol{Y}_{ij}\right] = E\left[P(\boldsymbol{Y}_{ij} = 1 | \boldsymbol{X}_i, \boldsymbol{Z}_i \beta, \boldsymbol{v}_i)\right] = E\left[\frac{e^{\boldsymbol{X}_{ij}\beta + \boldsymbol{Z}_{ij}\boldsymbol{v}_i}}{1 + e^{\boldsymbol{X}_{ij}\beta + \boldsymbol{Z}_{ij}\boldsymbol{v}_i}}\right] \neq \frac{e^{\boldsymbol{X}_{ij}\beta}}{1 + e^{\boldsymbol{X}_{ij}\beta}}$$

=> modèles GEE pour l'évolution marginale.

Modèle mixte linéaire généralisé (GLMM) pour données binaires

Estimation des paramètres

L'inférence est basée sur le modèle marginalisé de $\boldsymbol{Y}_i$ et s'obtient par *intégration numérique* car la distribution marginale ne peut pas être obtenue analytiquement.

$$L(\beta, \mathbf{G}) = \prod_{i=1}^{n} f(y_{ij} \,|\, \beta, \mathbf{G})$$

Différents types d'approximation de la vraisemblance marginale :

SAS : NLMIXED (calcul approché de l'intégrale, individu par individu, par quadrature gaussienne adaptative avec Q points de quadrature).

R: nlme; lme4; glmmML (ajuste un GLMM avec ordonnée à l'origine aléatoire par ML) ; glmmPQL (équivalent GLIMMIX: *Penalized quasi-likelihood*); MCMCglmm (méthode bayésienne).

Mplus: quadrature gaussienne adaptative , MCMC, etc.

GLMM appliqué à la précision

Illustration : vide - GLMM logistique avec intercept aléatoire

Code SAS

```
PROC NLMIXED DATA=RGmultivar QPOINTS=25;
PARMS beta0=11 varu=1;
z = beta0 + u;
IF (pr=1) THEN
  p = 1 / (1 + EXP(-z));
ELSE
  p = 1 - (1 / (1 + EXP(-z)));
ll = LOG(p);
MODEL pr ~ GENERAL(ll);
RANDOM u ~ NORMAL(0,varu) SUBJECT=id;
ESTIMATE 'icc' varu/(varu+(((ATAN(1)*4)**2)/3));
RUN;
```

BIC = 2066

ICC=0.307

GLMM appliqué à la précision

Illustration : M1 - GLMM logistique avec intercept aléatoire: effets fixes âge, mat et âge*mat

Code SAS

```
PROC NLMIXED DATA=RGmultivar QPOINTS=25;
PARMS beta0=11 beta1=-.10 beta2=-3.5 beta3=.03 varu=1;
z = beta0 + beta1*age + beta2*mat + beta3*age*mat + u;
IF (pr=1) THEN
  p = 1 / (1 + EXP(-z));
ELSE
  p = 1 - (1 / (1 + EXP(-z)));
ll = LOG(p);
MODEL pr ~ GENERAL(ll);
RANDOM u ~ NORMAL(0,varu) SUBJECT=id;
ESTIMATE 'icc' varu/( (((ATAN(1)*4)**2)/3) + varu);
RUN;
```

BIC = 1964

GLMM appliqué à la précision

Illustration : M2 - GLMM logistique avec intercept aléatoire: effets fixes âge et mat

Code SAS

```
PROC NLMIXED DATA=RGmultivar QPOINTS=25;
PARMS beta0=11 beta1=-.10 beta2=-3.5 varu=1;
z = beta0 + beta1*age + beta2*mat + u;
IF (pr=1) THEN
  p = 1 / (1 + EXP(-z));
ELSE
  p = 1 - (1 / (1 + EXP(-z)));
ll = LOG(p);
MODEL pr ~ GENERAL(ll);
RANDOM u ~ NORMAL(0,varu) SUBJECT=id;
ESTIMATE 'icc' varu/( (((ATAN(1)*4)**2)/3) + varu);
RUN;
```

BIC = 1962

GLMM appliqué à la précision

Illustration : M3 - GLMM logistique avec intercept et pente mat aléatoires: effets fixes âge et mat

Code SAS

```
PROC NLMIXED DATA=RGmultivar QPOINTS=25;
PARMS beta0=10 beta1=-.10 beta2=-1.5 v0=1 c01=0.1
v1=0.1;
z = beta0 + beta1*age + beta2*mat + u0 + u1*mat;
IF (pr=1) THEN
  p = 1 / (1 + EXP(-z));
ELSE
  p = 1 - (1 / (1 + EXP(-z)));
ll = LOG(p);
MODEL pr ~ GENERAL(ll);
RANDOM u0 u1 ~ NORMAL([0,0], [v0,c01,v1]) SUBJECT=id
OUT=blup;
ESTIMATE 're corr' c01/SQRT(v0*v1);
RUN;
```

BIC = 1938

GLMM appliqué à la précision

Fit Statistics

-2 Log Likelihood	1909.7
AIC (smaller is better)	1921.7
AICC (smaller is better)	1921.8
BIC (smaller is better)	1937.9

Paramètre	Estimation	Erreur type	DDL	Valeur du test t	Pr > |t|
beta0	10.5333	1.3325	108	7.90	<.0001
beta1	-0.08632	0.01721	108	-5.02	<.0001
beta2	-1.3197	0.2791	108	-4.73	<.0001
v0	1.4853	0.5078	108	2.92	0.0042
c01	-0.9067	0.5118	108	-1.77	0.0793
v1	1.9257	0.6788	108	2.84	0.0054

M4 : covariance entre les 2 effets aléatoires c01 fixée à 0: BIC = 1944

Libellé	Estimation	Erreur type	DDL	Valeur du test t	Pr > |t|
re corr	-0.5361	0.1613	108	-3.32	0.0012

GLMM appliqué à la précision

Code R : m3 <- lmer(pr~1+mat+age+(mat|sujet),family="binomial",data=mydata, REML = FALSE)

```
Generalized linear mixed model fit by the Laplace
approximation
Formula: pr ~ 1 + mat + age + (mat | sujet)
   Data: mydata
  AIC  BIC logLik deviance
 1921 1959 -954.5     1909
Random effects:
 Groups Name        Variance Std.Dev. Corr
 sujet  (Intercept) 1.5585   1.2484
        mat1        2.0622   1.4360   -0.584
Number of obs: 4400, groups: sujet, 110

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 10.57142    1.31154   8.060 7.61e-16 ***
mat1        -1.36228    0.22660  -6.012 1.84e-09 ***
age         -0.08629    0.01690  -5.104 3.32e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

Correlation of Fixed Effects:
     (Intr) mat1
mat1 -0.201
age  -0.989  0.089
```

GLMM appliqué à la précision

<u>Interprétation</u>

Moindre niveau de précision aux items positifs par rapport aux items négatifs.

Diminution de la précision avec l’âge.

Absence d’évolution de la précision au cours du temps.

DI dans le niveau moyen de précision et de sensibilité à la manipulation expérimentale.

GLMM appliqué à la précision

M3 : solution pour les effets aléatoires

Obs	id	Effect	Estimate	StdErr Pred	DF	tValue	Probt
1	1	u0	-0.03721	1.04369	108	-0.03565	0.97163
2	1	u1	-0.74863	1.14850	108	-0.65184	0.51589
37	19	u0	-2.97850	0.54956	108	-5.41980	0.00000
38	19	u1	2.27909	0.77719	108	2.93249	0.00411

3
Modélisation jointe de données répétées continues et binaires

Modèles joints pour données longitudinales

Soient deux vecteurs aléatoires de mesures répétées $\boldsymbol{Y}_1$ et $\boldsymbol{Y}_2$, comment construire une densité de probabilité jointe $f(y_1,y_2)$ de $(\boldsymbol{Y}_1,\boldsymbol{Y}_2)$?

Modèles multivariés : spécification directe de la distribution jointe $f(y_1,y_2)$.

Avantages: inférences sur les caractéristiques marginales de $\boldsymbol{Y}_1$ et $\boldsymbol{Y}_2$ et de leurs associations, traitement symétrique de $\boldsymbol{Y}_1$ et $\boldsymbol{Y}_2$.

Difficultés quand $\boldsymbol{Y}_1$ et $\boldsymbol{Y}_2$ sont de type différent.

Modèles joints pour données longitudinales

Modèles conditionnels : factorisation de la densité jointe

$$f(y_1, y_2) = f(y_1 | y_2) f(y_2) = f(y_2 | y_1) f(y_1)$$

Avantages: spécification séparée du modèle marginal et du modèle conditionnel

Difficultés: spécification du modèle conditionnel $f(y_1 | y_2)$ (relation entre $\boldsymbol{Y}_1$ et la covariable $\boldsymbol{Y}_2$ variant avec le temps), factorisations différentes de la distribution jointe.

Modèles joints pour données longitudinales

<u>Modèles à paramètres partagés</u> : soit $\boldsymbol{\upsilon}$ un vecteur d'effets aléatoires commun aux modèles de $\boldsymbol{Y}_1$ et de $\boldsymbol{Y}_2$ avec $\boldsymbol{Y}_1$ et $\boldsymbol{Y}_2$ indépendantes conditionnellement à $\boldsymbol{\upsilon}$.

$$f(y_1, y_2) = \int f(y_1 | \upsilon) f(y_2 | \upsilon) f(\upsilon) d\upsilon$$

Avantages: $\boldsymbol{Y}_1$ et $\boldsymbol{Y}_2$ n'ont pas besoin d'être du même type. Même interprétation des paramètres des modèles de $f(y_1,y_2)$, de $f(y_1)$ et de $f(y_2)$.

Modèles joints pour données longitudinales

<u>Modèles à effets aléatoires</u> : variables latentes distinctes et corrélées pour $\boldsymbol{Y}_1$ et $\boldsymbol{Y}_2$. Soient υ_1 et υ_2 des vecteurs d'effets aléatoires de densité jointe $f(\upsilon_1, \upsilon_2)$

$$f(y_1, y_2) = \iint f(y_1 | \upsilon_1) f(y_2 | \upsilon_2) f(\upsilon_1, \upsilon_2) d\upsilon_1 d\upsilon_2$$

Mêmes avantages que les modèles à paramètres partagés avec des hypothèses moins strictes sur la relation entre $\boldsymbol{Y}_1$ et $\boldsymbol{Y}_2$.

Modèle linéaire généralisé mixte bivarié

<u>Références</u> : Faes *et al.* (2008, 2009)

Vecteur de réponse bivarié $\boldsymbol{Y}_i = \left(\boldsymbol{Y}_{i1}^{'}, \boldsymbol{Y}_{i2}^{'}\right)^{'}$

Avec les mêmes notations que précédemment:

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i(\boldsymbol{\eta}_i) + \boldsymbol{\varepsilon}_i = \boldsymbol{h}(\boldsymbol{X}_i\beta + \boldsymbol{Z}_i\upsilon_i) + \boldsymbol{\varepsilon}_i$$

La moyenne et le terme d'erreur diffèrent selon le type d'observations ainsi que la fonction de lien inverse **h**(.) : identité pour données continues, logit pour données binaires.

$$\mathrm{V}_i = \mathrm{Var}(\boldsymbol{Y}_i) \simeq \Delta_i \boldsymbol{Z}_i \boldsymbol{G} \boldsymbol{Z}_i^{'} \Delta_i^{'} + \Sigma_i$$

GLMM bivarié

<u>Modèle à intercepts aléatoires</u> (Faes, Geys & Catalano, 2009)

Soient $\boldsymbol{Y}_{i1j}$ continue et $\boldsymbol{Y}_{i2j}$ binaire au temps j pour l'individu i :

$$\begin{pmatrix} \boldsymbol{Y}_{i1j} \\ \boldsymbol{Y}_{i1j} \end{pmatrix} = \begin{pmatrix} \beta_{10} + \beta_1 \boldsymbol{X}_{ij} + \boldsymbol{\upsilon}_{i1} \\ \dfrac{e^{\beta_{20} + \beta_2 \boldsymbol{X}_{ij} + \boldsymbol{\upsilon}_{i2}}}{1 + e^{\beta_{20} + \beta_2 \boldsymbol{X}_{ij} + \boldsymbol{\upsilon}_{i2}}} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1j} \\ \varepsilon_{i2j} \end{pmatrix},$$

avec effets aléatoires normalement distribués et erreurs ε_{i1j} et ε_{i2j} indépendantes,

$$\begin{pmatrix} \upsilon_{i1} \\ \upsilon_{i2} \end{pmatrix} \sim \mathcal{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_2\tau_1 & \tau_2^2 \end{pmatrix} \right\},$$

et ρ corrélation entre effets aléatoires.

Approximation de $\rho_{\mathbf{Y}_1\mathbf{Y}_2}$:

$$\rho_{\mathbf{Y}_1\mathbf{Y}_2} = \frac{\rho\tau_1\tau_2}{\sqrt{\tau_1^2 + \sigma^2}\sqrt{\tau_2^2 + \dfrac{\pi^2}{3}}}.$$

GLMM bivarié avec NLMIXED

Code SAS (adapté de *Faes et al.*, 2008)

```
PROC NLMIXED data=RGunivar QPOINTS=5;
parms beta11=5.35 beta12=-0.002 beta13=0.02 beta14=0.04 beta15=-0.004
beta21=10.57 beta22=-1.36 beta23=-0.08
sigma=0.5
tau1=0.25 tau2=0.25;
if distvar = "Normal" then do;
mean = beta11 + beta12*rang + beta13*age + beta14*mat+ beta15*rang*mat + u1;
dens = -0.5*log(2*3.14159265358) - log(sigma) -0.5*(rep-mean)**2/(sigma**2);
ll = dens;
end;
if distvar = "Binary" then do;
eta = beta21 + beta22*mat + beta23*age + u2;
p = exp(eta)/(1+exp(eta));
ll = rep*log(p) + (1-rep)*log(1-p);
end;
model rep ~ general(ll);
random u1 u2 ~normal([0,0],[tau1**2,rho*tau1*tau2,tau2**2]) subject=id out=upsilon;
estimate "tau1^2" tau1*tau1;
estimate "tau2^2" tau2*tau2;
estimate "sigma^2" sigma*sigma;
estimate "rho_1_2" rho*tau1*tau2/(sqrt(tau1*tau1+sigma*sigma)*sqrt(tau2*tau2+(3.141592653**2/3)));
predict eta out=eta;
Run;
```

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

$$\Pr(x=1) = \binom{1}{x} p^x (1-p)^{1-x}$$

GLMM bivarié avec NLMIXED

Specifications

Data Set	WORK.RGUNIVAR
Dependent Variable	rep
Distribution for Dependent Variable	General
Random Effects	u1 u2
Distribution for Random Effects	Normal
Subject Variable	id
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions

Observations Used	8593
Observations Not Used	207
Total Observations	8800
Subjects	110
Max Obs Per Subject	80
Parameters	12
Quadrature Points	25

Fit Statistics

-2 Log Likelihood	2899.1
AIC (smaller is better)	2923.1
AICC (smaller is better)	2923.1
BIC (smaller is better)	2955.5

GLMM bivarié avec NLMIXED

Paramètres estimés

Paramètre	Estimation	Erreur type	DDL	Valeur du test t	Pr > |t|
beta11	5.3599	0.2666	108	20.11	<.0001
beta12	-0.00240	0.000507	108	-4.72	<.0001
beta13	0.02627	0.003567	108	7.37	<.0001
beta14	0.04704	0.01647	108	2.86	0.0051
beta15	-0.00385	0.000698	108	-5.51	<.0001
beta21	10.5400	1.2784	108	8.24	<.0001
beta22	-1.2740	0.1409	108	-9.04	<.0001
beta23	-0.08902	0.01663	108	-5.35	<.0001
sigma	0.2586	0.002862	108	90.36	<.0001
tau1	0.2802	0.01938	108	14.46	<.0001
tau2	-1.0380	0.1179	108	-8.81	<.0001
rho	0.4948	0.09477	108	5.22	<.0001

Additional Estimates

Libellé	Estimation	Erreur type	DDL	Valeur du test t	Pr > |t|
tau1^2	0.07853	0.01086	108	7.23	<.0001
tau2^2	1.0774	0.2447	108	4.40	<.0001
sigma^2	0.06686	0.001480	108	45.18	<.0001
rho_1_2	-0.1806	0.04049	108	-4.46	<.0001

GLMM bivarié avec Mplus

```
TITLE:  modèle de croissance à 2 niveaux

define: lntr=log(tr);
        rm=rang*mat;

DATA: FILE IS Rglob_cleaned.txt;

VARIABLE: NAMES ARE id mmse age rang temps mat pr tr;
usevar are id rang age mat pr lntr rm;
categorical is pr;
WITHIN =rang mat rm;
BETWEEN = age;
CLUSTER = id;
missing are .;

ANALYSIS: TYPE = TWOLEVEL RANDOM;
          estimator=bayes;

MODEL:
%WITHIN%
!modélisation lntr niveau 1
lntr ON rang rm;
s_lntr|lntr ON mat;
!modélisation pr niveau 1
pr ON rang rm;
s_pr| pr on mat;

%BETWEEN%
!covariance entre effets aléatoires
lntr with s_lntr s_pr pr;
s_lntr with s_pr pr;
s_pr with pr;
!régression des effets aléatoires sur l'âge
lntr s_lntr s_pr pr on age;
```

Estimation de la distribution jointe des *p* effets aléatoires : méthode d'intégration numérique très coûteuse quand p >3 ou 4.

GLMM bivarié avec Mplus

MODEL RESULTS

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	95% C.I. Upper 2.5%
Within Level					
LNTR ON					
RANG	0.001	0.001	0.110	-0.001	0.004
RM	-0.003	0.001	0.000	-0.005	-0.002
PR ON					
RANG	-0.021	0.008	0.003	-0.036	-0.004
RM	0.014	0.005	0.003	0.003	0.024
Residual Variances					
LNTR	0.053	0.001	0.000	0.051	0.055

GLMM bivarié avec Mplus

MODEL RESULTS

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	95% C.I. Upper 2.5%
Between Level					
S_LNTR ON					
AGE	0.000	0.002	0.463	-0.004	0.005
S_PR ON					
AGE	-0.015	0.012	0.120	-0.038	0.008
LNTR ON					
AGE	0.021	0.004	0.000	0.013	0.030
PR ON					
AGE	-0.017	0.018	0.157	-0.056	0.018
LNTR WITH					
S_LNTR	-0.020	0.006	0.000	-0.032	-0.009
S_PR	0.032	0.031	0.143	-0.026	0.097
S_LNTR WITH					
S_PR	-0.025	0.016	0.037	-0.062	0.003
PR	0.031	0.025	0.097	-0.013	0.085
S_PR WITH					
PR	-0.668	0.217	0.000	-1.170	-0.398
LNTR WITH					
PR	-0.091	0.051	0.027	-0.199	0.001
Intercepts					
LNTR	5.660	0.317	0.000	4.981	6.264
S_LNTR	0.029	0.172	0.440	-0.297	0.322
S_PR	0.253	0.876	0.357	-1.578	1.895
Thresholds					
PR$1	-4.410	1.378	0.000	-7.224	-1.577
Residual Variances					
PR	1.205	0.341	0.000	0.729	2.042
LNTR	0.093	0.015	0.000	0.070	0.125
S_LNTR	0.024	0.004	0.000	0.019	0.032
S_PR	0.500	0.150	0.000	0.296	0.872