

multlcm : fonction d'estimation de modèles mixtes à processus latent pour données longitudinales multivariées

Viviane Philipps, Cécile Proust-Lima

INSERM U897, Epidémiologie et Biostatistique, Bordeaux, France
Université Bordeaux Segalen, ISPED, Bordeaux, France

28 juin 2013

Lyon

Deuxièmes Rencontres R

= mesures d'une **quantité sous-jacente** d'intérêt (latente, non observée)

Ex : bien-être, qualité de vie, douleur, ...

cognition évaluée par un ensemble de **tests psychométriques**

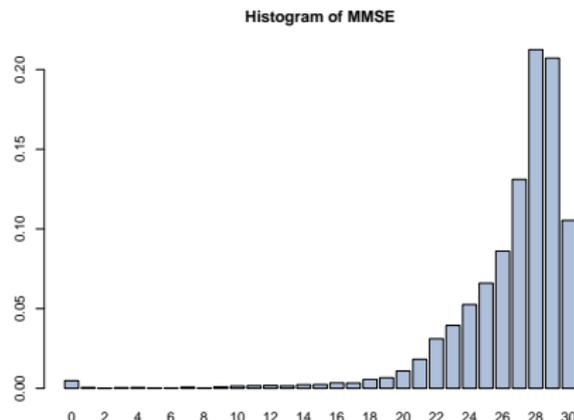
Exemples de tests psychométriques :

- Mini-Mental State Examination
- Grober-Buschke
- Trail Making Test
- Isaacs Test Set
- ...

→ obtention de plusieurs **scores**

Propriétés métrologiques des échelles de mesure :

- variable souvent bornée
- distribution souvent asymétrique
- effets plancher/plafond : beaucoup d'observations au min/max
- curvilinéarité : une différence d'un point n'a pas la même signification biologique selon le score initial



Modèle linéaire mixte :

- variable d'intérêt continue et gaussienne
- effet constant des prédicteurs sur la variable d'intérêt
→ pas adapté aux tests psychométriques

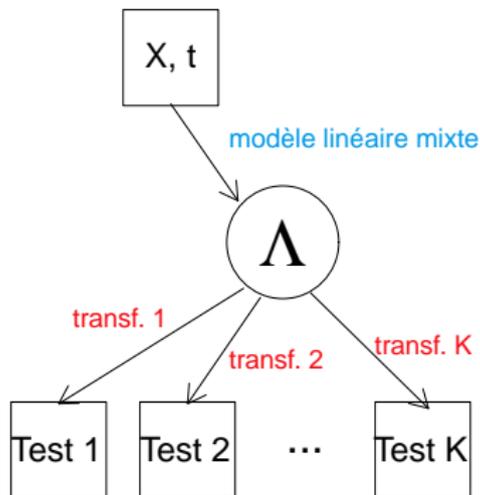
Modèle linéaire mixte :

- variable d'intérêt continue et gaussienne
- effet constant des prédicteurs sur la variable d'intérêt
→ pas adapté aux tests psychométriques

Modèle mixte à processus latent :

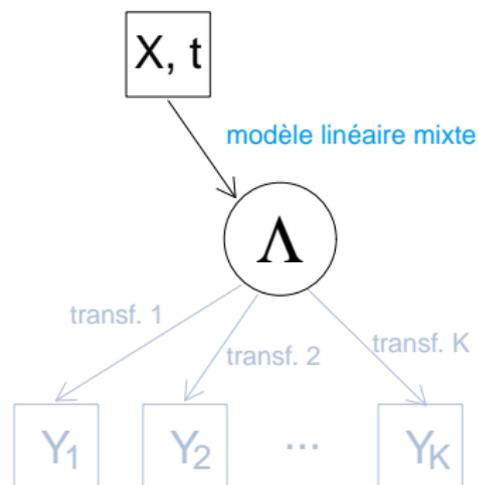
(Proust et al, 2006 - Proust-Lima et al, 2012)

- définition d'un processus latent commun à tous les tests
- modèle linéaire mixte pour expliquer la trajectoire de ce processus latent
- fonction de lien (transformation) pour relier un test au processus latent et prendre en compte les propriétés métrologiques



Modèle mixte à processus latent

Modèle structurel



Pour le sujet i ($i = 1, \dots, N$) et pour tout temps $t \in \mathbb{R}$:

$$\Lambda_i(t) = X_i(t)\beta + Z_i(t)U_i + \omega_i(t)$$

avec $\beta_0 = 0$ et $\text{Var}(u_0) = 1$

avec X_i, Z_i : vecteur de covariables

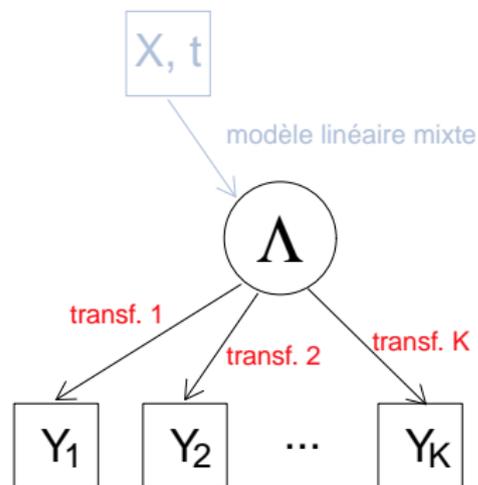
β : vecteur des effets fixes

U_i : effets aléatoires de loi $\mathcal{N}(0, \Sigma)$

ω_i : mouvement brownien de variance σ_ω^2 ou processus autorégressif de variance σ_ω^2 et de coefficient de corrélation ρ

Modèle mixte à processus latent

Equations d'observation



Pour le test k ($k = 1, \dots, K$), le sujet i ($i = 1, \dots, N$) et la répétition j ($j = 1, \dots, n_{ik}$) :

$$H_k(Y_{ijk}, \eta_k) = \Lambda_i(t_{ijk}) + \tilde{X}_i(t_{ijk})\gamma_k + \alpha_{ik} + \varepsilon_{ijk}$$

avec H_k : transformations paramétrées (linéaire, beta standardisée, splines)

Y_{ijk} : score du sujet i à la j^{e} répétition du test k

\tilde{X}_i : sous-matrice de X_i

γ_k : vecteur des contrastes associés au test k

α_{ik} : intercept aléatoire spécifique au test et de loi $\mathcal{N}(0, \sigma_{\alpha_k}^2)$

ε_{ijk} : erreurs de mesures de loi $\mathcal{N}(0, \sigma_{\varepsilon_k}^2)$

- Linéaires :

$$H(y, \eta) = \frac{y - \eta_1}{\eta_2}$$

→ modèle linéaire mixte multivarié

- Beta standardisées :

$$H(y, \eta) = \frac{B(y, \eta_1, \eta_2) - \eta_3}{\eta_4}$$

avec $B(y, \eta_1, \eta_2)$ fonction de répartition d'une loi beta de paramètres η_1, η_2

- Splines :

$$H(y, \eta) = \eta_0 + \sum_{l=1}^L \eta_l I_l(y)$$

avec $(I_l)_{l=1, \dots, L}$ base de I-splines quadratiques

Tous les paramètres $\theta = (\beta, \Sigma, \sigma_\omega, \eta, \gamma, \sigma_\alpha, \sigma_\varepsilon)$ sont estimés simultanément.

D'après les hypothèses du modèle,

$$H(Y_i) = (H(Y_{i11}, \eta_1), \dots, H(Y_{in_{iK}K}, \eta_K))' \sim \mathcal{N}(E_i, V_i)$$

Log-vraisemblance du modèle :

$$l(y, \theta) = l(H(y), \theta) + \ln(J_H(y, \theta))$$

- estimation par maximum de vraisemblance :
 - algorithme de type Marquardt
 - critères de convergence stricts
- fonction `multlcmm` du package `lcmm`

Etude PAQUID (*Letenneur et al, 1994*) :

- cohorte prospective pour étudier le vieillissement cognitif
- personnes âgées de plus de 65 ans
- suivi tous les 2 ou 3 ans
- batterie de tests psychométriques à chaque visite

Echantillon :

- 2252 sujets
- suivis 1, 3, 5, 7, 10 et 13 ans après l'inclusion

Décrire le déclin cognitif dans PAQUID à partir de 4 tests psychométriques :

- **MMSE** (Mini-Mental State Examination) (*Folstein et al, 1975*)
→ différents champs de la cognition
- **IST** (Isaacs Test Set) (*Isaacs et al, 1973*)
→ nommer successivement des noms de villes/couleurs/animaux/fruits en 15 secondes
- **BVRT** (Benton Visual Retention Test) (*Benton, 1965*)
→ reconnaître successivement 15 figures parmi 4 proposées
- **DSSTW** (Digit Symbol Substitution Test) (*Wechsler, 1981*)
→ substituer un maximum de symboles par le chiffre correspondant en 90 secondes

Hypothèse : ces 4 tests mesurent la cognition globale

- évolution quadratique du processus latent en fonction de l'âge
- effets aléatoires sur les sujets
- relation non-linéaire entre le processus et les tests
- intercept aléatoire spécifique aux tests

- évolution quadratique du processus latent en fonction de l'âge
- effets aléatoires sur les sujets
- relation non-linéaire entre le processus et les tests
- intercept aléatoire spécifique aux tests

```
m <- multlcm (MMSE+IST+BVRT+DSSTW~1+temps+I(temps^2),  
             # effets fixes du modèle linéaire mixte
```

- évolution quadratique du processus latent en fonction de l'âge
- effets aléatoires sur les sujets
- relation non-linéaire entre le processus et les tests
- intercept aléatoire spécifique aux tests

```
m <- multlcm (MMSE+IST+BVRT+DSSTW~1+temps+I(temps^2),  
  random=~1+temps+I(temps^2), subject="numero",  
  # effets aléatoires
```

- évolution quadratique du processus latent en fonction de l'âge
- effets aléatoires sur les sujets
- relation non-linéaire entre le processus et les tests
- intercept aléatoire spécifique aux tests

```
m <- multlcmm(MMSE+IST+BVRT+DSSTW~1+temps+I(temps^2),  
  random=~1+temps+I(temps^2), subject="numero",  
  idiag=FALSE,  
  # effets aléatoires corrélés
```

- évolution quadratique du processus latent en fonction de l'âge
- effets aléatoires sur les sujets
- relation non-linéaire entre le processus et les tests
- intercept aléatoire spécifique aux tests

```
m <- multlcm (MMSE+IST+BVRT+DSSTW~1+temps+I(temps^2),  
  random=~1+temps+I(temps^2), subject="numero",  
  idiag=FALSE,  
  cor=BM(temps),  
  # mouvement brownien
```

- évolution quadratique du processus latent en fonction de l'âge
- effets aléatoires sur les sujets
- relation non-linéaire entre le processus et les tests
- intercept aléatoire spécifique aux tests

```
m <- multlcm (MMSE+IST+BVRT+DSSTW~1+temps+I(temps^2),  
  random=~1+temps+I(temps^2), subject="numero",  
  idiag=FALSE,  
  cor=BM(temps),  
  link=c("beta", "beta", "beta", "beta"),  
  # transformation beta pour chaque test
```

Spécification du modèle

- évolution quadratique du processus latent en fonction de l'âge
- effets aléatoires sur les sujets
- relation non-linéaire entre le processus et les tests
- intercept aléatoire spécifique aux tests

```
m <- multlcm (MMSE+IST+BVRT+DSSTW~1+temps+I(temps^2),  
  random=~1+temps+I(temps^2), subject="numero",  
  idiag=FALSE,  
  cor=BM(temps),  
  link="beta",  
  # transformation beta pour chaque test
```

- évolution quadratique du processus latent en fonction de l'âge
- effets aléatoires sur les sujets
- relation non-linéaire entre le processus et les tests
- intercept aléatoire spécifique aux tests

```
m <- multlcmm(MMSE+IST+BVRT+DSSTW~1+temps+I(temps^2),  
  random=~1+temps+I(temps^2), subject="numero",  
  idiag=FALSE,  
  cor=BM(temps),  
  link="beta",  
  randomY=TRUE,  
  # intercept aléatoire spécifique
```

- évolution quadratique du processus latent en fonction de l'âge
- effets aléatoires sur les sujets
- relation non-linéaire entre le processus et les tests
- intercept aléatoire spécifique aux tests

```
m <- multlcmm(MMSE+IST+BVRT+DSSTW~1+temps+I(temps^2),  
  random=~1+temps+I(temps^2), subject="numero",  
  idiag=FALSE,  
  cor=BM(temps),  
  link="beta",  
  randomY=TRUE,  
  data=paquid)
```

Statistical Model:

Dataset: paquid

Number of subjects: 2252

Number of observations: 29018

Number of latent classes: 1

Number of parameters: 32

Link functions: Standardised Beta CdF for mmse

Standardised Beta CdF for ist

Standardised Beta CdF for bvrt

Standardised Beta CdF for dsstw

Iteration process:

Convergence criteria satisfied

Number of iterations: 17

Convergence criteria: parameters= 1.3e-05

likelihood= 9.3e-05

second derivatives= 1.1e-07

Goodness-of-fit statistics:

maximum log-likelihood: -70935.24

AIC: 141934.49

BIC: 142117.51

Fixed effects in the longitudinal model:

	coef	Se	Wald	p-value
intercept (not estimated)	0.00000			
temps	-0.48390	0.09697	-4.990	0
I(temps^2)	-0.28391	0.03364	-8.438	0

Variance-covariance matrix of the random-effects:
(the variance of the first random effect is not estimated)

	intercept	temps	I(temps^2)
intercept	1.00000		
temps	-0.07842	0.56842	
I(temps^2)	-0.01956	-0.25295	0.14628

summary(m) 3/3

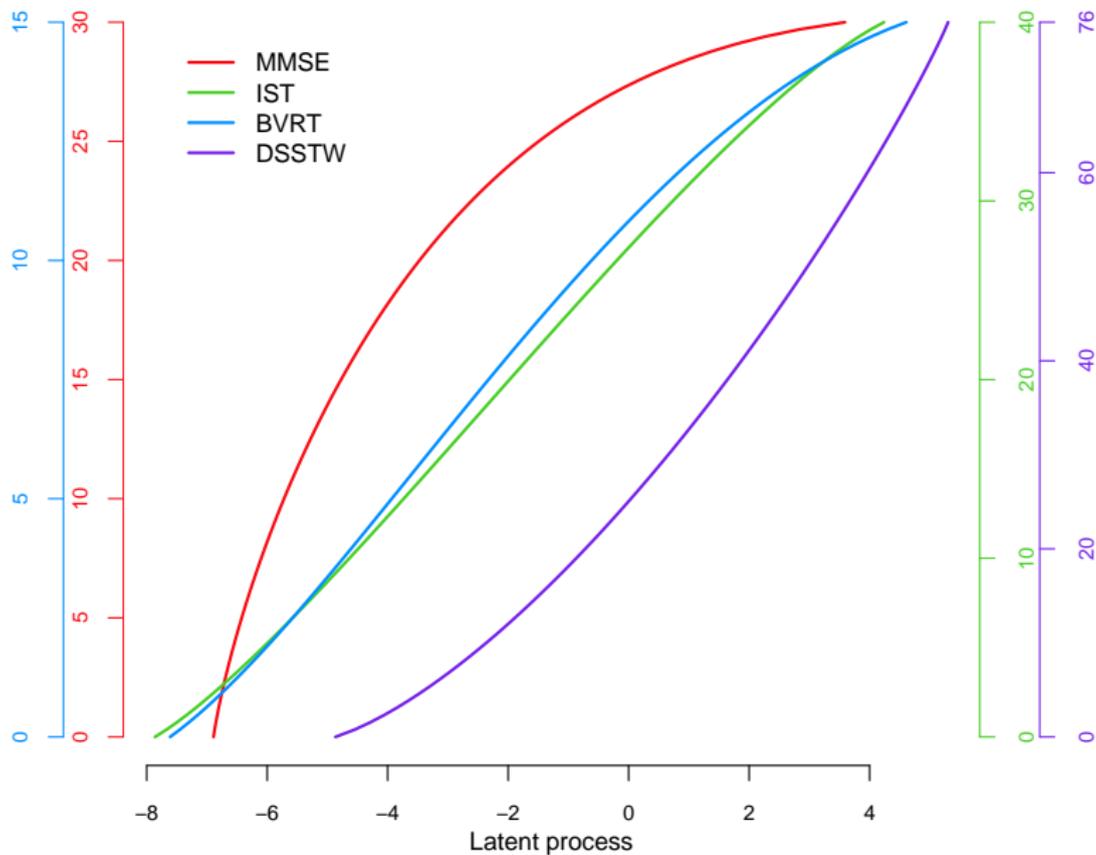
```
                coef      Se  
BM standard error: 0.68348 0.07203
```

```
                mmse   ist   bvrt dsstw  
Residual standard error:          1.124 1.021 1.285 0.477  
Standard error of the random effect: 0.743 0.834 0.656 0.738
```

Parameters of the link functions:

```
                coef      Se   Wald p-value  
mmse-Beta1      1.48003 0.03393  43.622 0.00000  
mmse-Beta2     -0.42961 0.04241 -10.131 0.00000  
mmse-Beta3      0.51856 0.01374  37.754 0.00000  
mmse-Beta4      0.06605 0.00581  11.363 0.00000  
ist-Beta1       0.06672 0.04101   1.627 0.10378  
...  
dsstw-Beta4     0.09538 0.00821  11.624 0.00000
```

plot.linkfunction(m)



Prédictions

Données de prédiction :

```
newdata <- data.frame(temps=seq(0,3,length.out=100))  
newdata$age <- newdata$temps*10+65
```

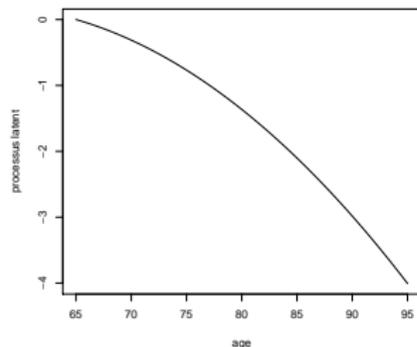
Prédictions

Données de prédiction :

```
newdata <- data.frame(temps=seq(0,3,length.out=100))  
newdata$age <- newdata$temps*10+65
```

Prédiction du processus latent :

```
predictL(m,newdata)
```



Prédictions

Données de prédiction :

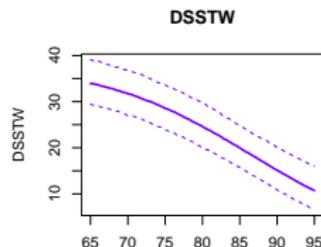
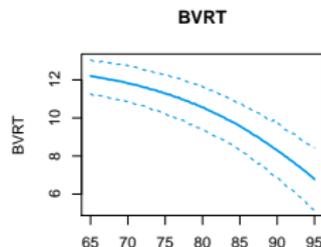
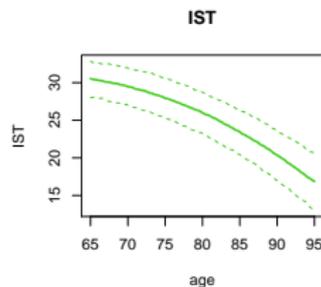
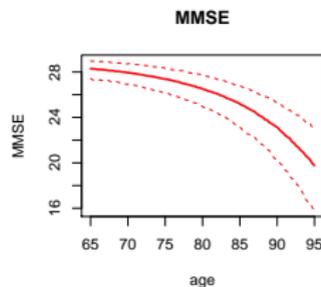
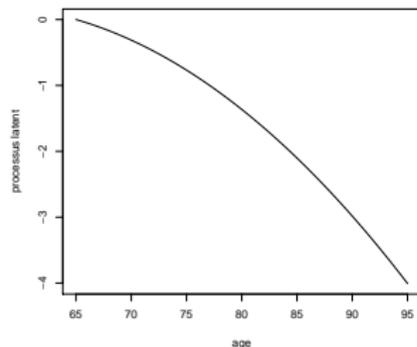
```
newdata <- data.frame(temps=seq(0,3,length.out=100))  
newdata$age <- newdata$temps*10+65
```

Prédictions des scores :

```
predictY(m,newdata,draws=TRUE)
```

Prédiction du processus latent :

```
predictL(m,newdata)
```



Particularités de `multlcm` :

- modèle mixte multivarié
- s'adapte à différents types de données
- possibilité d'analyser des profils de trajectoires (classes latentes)

Financement : Projet ANR Mobidyq

Références :

Proust, Jacqmin-Gadda, Taylor et al. (2006). *Biometrics*, 62, 1014-24
Proust-Lima, Amieva, Jacqmin-Gadda (2012). *British J Math Stat Psychol* - sous presse