Modèles mixtes linéaires à classes latentes : Package LCMM

Chaymae YOUSFI[1,4] Alexandre BUREAU[2,4]
Michel MAZIADE[3,4]

[1]Département de mathématiques et de statistique, Université Laval,
[2]Département de médecine sociale et préventive, Université Laval,
[3]Département de psychiatrie et neurosciences, Université Laval,
[4]Centre de Recherche CERVO, Québec, Canada

15 Mai 2019

Plan de la présentation

Cadre général du modèle

- Usage des Modèles Mixtes Linéaires (LMM) : études longitudinales.

hypothèses du modèle

- La variable réponse longitudinale est continue.
- Les effets aléatoires et les erreurs sont gaussiennes.
- La linéarité des relations avec la variable réponse.
- L'homogénéité de la population.
- Les données manquantes sont manquantes aléatoirement .

hypothèses du modèle

- La variable réponse longitudinale est continue.
- Les effets aléatoires et les erreurs sont gaussiennes.
- La linéarité des relations avec la variable réponse.
- L'homogénéité de la population.
- Les données manquantes sont manquantes aléatoirement .

hypothèses du modèle

- La variable réponse longitudinale est continue.
- Les effets aléatoires et les erreurs sont gaussiennes.
- La linéarité des relations avec la variable réponse.
- L'homogénéité de la population.
- Les données manquantes sont manquantes aléatoirement .

hypothèses du modèle

- La variable réponse longitudinale est continue.
- Les effets aléatoires et les erreurs sont gaussiennes.
- La linéarité des relations avec la variable réponse.
- L'homogénéité de la population.
- Les données manquantes sont manquantes aléatoirement .

hypothèses du modèle

- La variable réponse longitudinale est continue.
- Les effets aléatoires et les erreurs sont gaussiennes.
- La linéarité des relations avec la variable réponse.
- L'homogénéité de la population.
- Les données manquantes sont manquantes aléatoirement .

Package sous R : LCMM

Le package lcmm :

- Proust-Lima
- https://cran.r-project.org/package=lcmm
- modèles mixtes linéaires,
- modèles mixtes linéaires à classes latentes,

Package sous R : LCMM

Le package lcmm :

- Proust-Lima
- https://cran.r-project.org/package=lcmm
- modèles mixtes linéaires,
- modèles mixtes linéaires à classes latentes,

Package sous R : LCMM

Le package lcmm :

- Proust-Lima
- https://cran.r-project.org/package=lcmm
- modèles mixtes linéaires,
- modèles mixtes linéaires à classes latentes,

Package sous R : LCMM

Le package lcmm :

- Proust-Lima
- https://cran.r-project.org/package=lcmm
- modèles mixtes linéaires,
- modèles mixtes linéaires à classes latentes,

Package sous R : LCMM

Le package lcmm :

- Proust-Lima
- https://cran.r-project.org/package=lcmm
- modèles mixtes linéaires,
- modèles mixtes linéaires à classes latentes,

Notations et définitions

- N sujets et G classes latentes.
- On considère une variable latente discrète, c_i avec (i, i=1,..., N), modélisant l'appartenance à la classe latente :

Ainsi la probabilité d'appartenance du sujet i,(i, i=1,..., N) à la classe g, (g,g=1,...,G) est :

$$\pi_{ig} = P(c_i = g/X_{1i}) = \frac{e^{\xi_{0g} + X_{1i}^T * \xi_{1g}}}{\sum_{l=1}^{G} e^{\xi_{0l} + X_{1i}^T * \xi_{1l}}}$$

Notations et définitions

- N sujets et G classes latentes.
- On considère une variable latente discrète, c_i avec (i, i=1,..., N), modélisant l'appartenance à la classe latente :

Ainsi la probabilité d'appartenance du sujet i,(i, i=1,..., N) à la classe g, (g,g=1,...,G) est :

$$\pi_{ig} = P(c_i = g/X_{1i}) = \frac{e^{\xi_{0g} + X_{1i}^T * \xi_{1g}}}{\sum_{l=1}^{G} e^{\xi_{0l} + X_{1i}^T * \xi_{1l}}}$$

Notations et définitions

- N sujets et G classes latentes.
- On considère une variable latente discrète, c_i avec (i, i=1,..., N), modélisant l'appartenance à la classe latente :

Ainsi la probabilité d'appartenance du sujet i,(i, i=1,..., N) à la classe g, (g,g=1,...,G) est :

$$\pi_{ig} = P(c_i = g/X_{1i}) = \frac{e^{\xi_{0g} + X_{1i}^T * \xi_{1g}}}{\sum_{l=1}^{G} e^{\xi_{0l} + X_{1i}^T * \xi_{1l}}}$$

Notations et définitions

- N sujets et G classes latentes.
- On considère une variable latente discrète, c_i avec (i, i=1,..., N), modélisant l'appartenance à la classe latente :

Ainsi la probabilité d'appartenance du sujet i,(i, i=1,..., N) à la classe g, (g,g=1,...,G) est :

$$\pi_{ig} = P(c_i = g/X_{1i}) = \frac{e^{\xi_{0g} + X_{1i}^T * \xi_{1g}}}{\sum_{l=1}^{G} e^{\xi_{0l} + X_{1i}^T * \xi_{1l}}}$$

Fonction hlme : Fondement théorique

Les mesures répétées du marqueur longitudinal $Y_{ij}(j, j = 1, ..., n_i)$ sont :

$$Y_{ij}/_{c_i=g} = Z_{ij}^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g + \epsilon_{ij}$$

tel que :

- Z_{ij}, X_{2ij} et X_{3ij} sont les vecteurs des covariables
- $u_{ig}(\mu_j, \omega_g^2 B)$
- $\epsilon_{ij}(0, \sigma^2)$

Fonction hlme : Fondement théorique

Les mesures répétées du marqueur longitudinal $Y_{ij}(j, j = 1, ..., n_i)$ sont :

$$Y_{ij}/_{c_i=g} = Z_{ij}^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g + \epsilon_{ij}$$

tel que :

- Z_{ij}, X_{2ij} et X_{3ij} sont les vecteurs des covariables
- $u_{ig}(\mu_j, \omega_g^2 B)$
- $\epsilon_{ij}(0, \sigma^2)$

Fonction hlme : Fondement théorique

Les mesures répétées du marqueur longitudinal $Y_{ij}(j, j = 1, ..., n_i)$ sont :

$$Y_{ij}/_{c_i=g} = Z_{ij}^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g + \epsilon_{ij}$$

tel que :

- Z_{ij}, X_{2ij} et X_{3ij} sont les vecteurs des covariables
- $u_{ig}(\mu_j, \omega_g^2 B)$
- $\epsilon_{ij}(0, \sigma^2)$

Fonction hlme : Fondement théorique

Les mesures répétées du marqueur longitudinal $Y_{ij}(j, j = 1, ..., n_i)$ sont :

$$Y_{ij}/_{c_i=g} = Z_{ij}^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g + \epsilon_{ij}$$

tel que :

- Z_{ij}, X_{2ij} et X_{3ij} sont les vecteurs des covariables
- $u_{ig}(\mu_j, \omega_g^2 B)$
- $\epsilon_{ij}(0, \sigma^2)$

Fonction hlme : Fondement théorique

Les mesures répétées du marqueur longitudinal $Y_{ij}(j, j = 1, ..., n_i)$ sont :

$$Y_{ij}/_{c_i=g} = Z_{ij}^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g + \epsilon_{ij}$$

tel que :

- Z_{ij}, X_{2ij} et X_{3ij} sont les vecteurs des covariables
- $u_{ig}(\mu_j, \omega_g^2 B)$
- $\epsilon_{ij}(0, \sigma^2)$

Fonction hlme : Fondement théorique

Les mesures répétées du marqueur longitudinal $Y_{ij}(j, j = 1, ..., n_i)$ sont :

$$Y_{ij}/_{c_i=g} = Z_{ij}^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g + \epsilon_{ij}$$

tel que :

- Z_{ij}, X_{2ij} et X_{3ij} sont les vecteurs des covariables
- $u_{ig}(\mu_j, \omega_g^2 B)$
- $\epsilon_{ij}(0, \sigma^2)$

Fonction hlme : Implémentation sur R

hlme(fixed=Y$\sim$ *Time*+X_2 + X_3+Time :X_2+Time :X_3+X_2 :
X_3,random=$\sim$ Z,
subject="Identifiant",mixture=$\sim$ *Time*+X_3,
classmb=$\sim X_1$,ng=G,
data=*ourdata*)

Fonction lcmm : Fondement théorique

- Soit la classe latente g, g(1,...,G).
- Le processus latent est :

$$\Delta_i(t)/_{c_i=g} = Z_i^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g$$

- Relation entre la variable réponse et le processus latent :

$$H(Y_i(t)/_{c_i=g}, \eta) = \Delta_i(t)/_{c_i=g} + \epsilon_i(t)$$

où :
$H(,\eta)$=linéaire, Bêta CDF, etc

Fonction lcmm : Fondement théorique

- Soit la classe latente g, g(1,...,G).
- Le processus latent est :

$$\Delta_i(t)/_{c_i=g} = Z_i^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g$$

- Relation entre la variable réponse et le processus latent :

$$H(Y_i(t)/_{c_i=g}, \eta) = \Delta_i(t)/_{c_i=g} + \epsilon_i(t)$$

où :
$H(, \eta)$=linéaire, Bêta CDF, etc

Fonction lcmm : Fondement théorique

- Soit la classe latente g, g(1,...,G).
- Le processus latent est :

$$\Delta_i(t)/_{c_i=g} = Z_i^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g$$

- Relation entre la variable réponse et le processus latent :

$$H(Y_i(t)/_{c_i=g}, \eta) = \Delta_i(t)/_{c_i=g} + \epsilon_i(t)$$

où :
$H(,\eta)$=linéaire, Bêta CDF, etc

Fonction lcmm : Fondement théorique

- Soit la classe latente g, g(1,...,G).
- Le processus latent est :

$$\Delta_i(t)/_{c_i=g} = Z_i^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g$$

- Relation entre la variable réponse et le processus latent :

$$H(Y_i(t)/_{c_i=g}, \eta) = \Delta_i(t)/_{c_i=g} + \epsilon_i(t)$$

où :
$H(,\eta)$=linéaire, Bêta CDF, etc

Fonction lcmm : Fondement théorique

- Soit la classe latente g, g(1,...,G).
- Le processus latent est :

$$\Delta_i(t)/_{c_i=g} = Z_i^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g$$

- Relation entre la variable réponse et le processus latent :

$$H(Y_i(t)/_{c_i=g}, \eta) = \Delta_i(t)/_{c_i=g} + \epsilon_i(t)$$

où :
$H(,\eta)$=linéaire, Bêta CDF, etc

Fonction lcmm : Fondement théorique

- Soit la classe latente g, g(1,...,G).
- Le processus latent est :

$$\Delta_i(t)/_{c_i=g} = Z_i^T * u_{ig} + X_{2ij}^T * \beta + X_{3ij}^T * \gamma_g$$

- Relation entre la variable réponse et le processus latent :

$$H(Y_i(t)/_{c_i=g}, \eta) = \Delta_i(t)/_{c_i=g} + \epsilon_i(t)$$

où :
$H(,\eta)$=linéaire, Bêta CDF, etc

Fonction lcmm : Implémentation sur R

$$\text{lcmm(fixed=Y}\sim \mathit{Time} + X1 + \mathit{Time} : X1, \mathit{random} =\sim \mathit{Time},$$
$$\text{subject="Identifiant",mixture=}\sim \mathit{Time},$$
$$\text{classmb=}\sim X_2 + X_3\text{,ng=G,}$$
$$\text{data=}\mathit{ourdata})$$

Fonction jointlcmm : Fondement théorique

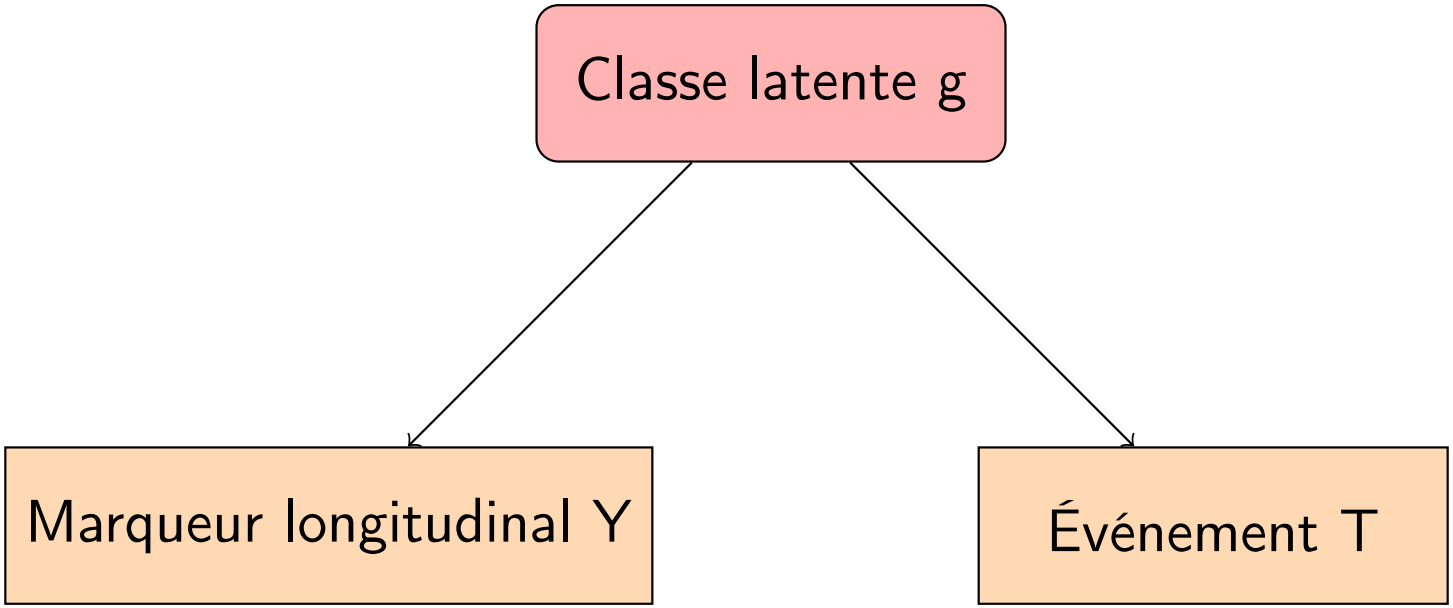

Figure – Modélisation conjointe d'un temps d'événement et d'un marqueur longitudinal

Fonction jointlcmm : Implémentation sur R

jointlcmm(fixed=Y$\sim$ *Time* + $X1$ + *Time* : $X1$, *random* =$\sim$ *Time*,
subject="Identifiant",mixture=$\sim$ *Time*,
classmb=$\sim X_2 + X_3$,ng=G,data=*ourdata*,
survival=Surv(survie,evenement)$\sim X_1$+mixture(X_4),
hazard='Weibull-Specific')

Application : Présentation des données

- Objectif :

  On cherche à stratifier les 67 sujets de notre étude en deux sous-populations selon le risque de développer des troubles psychiatriques.

- N=67 sujets et G =2 classes latentes,
- Les covariables : le score polygénique SZ, la variable sexe, la variable trauma.

Application : Présentation des données

- Objectif :
  On cherche à stratifier les 67 sujets de notre étude en deux sous-populations selon le risque de développer des troubles psychiatriques.
- N=67 sujets et G =2 classes latentes,
- Les covariables : le score polygénique SZ, la variable sexe, la variable trauma.

Application : Présentation des données

- Objectif :
  On cherche à stratifier les 67 sujets de notre étude en deux sous-populations selon le risque de développer des troubles psychiatriques.
- N=67 sujets et G =2 classes latentes,
- Les covariables : le score polygénique SZ, la variable sexe, la variable trauma.

Application : Présentation des données

- Objectif :
  On cherche à stratifier les 67 sujets de notre étude en deux sous-populations selon le risque de développer des troubles psychiatriques.
- N=67 sujets et G =2 classes latentes,
- Les covariables : le score polygénique SZ, la variable sexe, la variable trauma.

Fonction hlme

```
Maximum Likelihood Estimates:

Fixed effects in the class-membership model:
(the class of reference is the last class)

                              coef      Se   Wald p-value
intercept class1           4.68327 3.26358  1.435 0.15128
score_sz_imp_10 class1     0.30781 0.16502  1.865 0.06214

Fixed effects in the longitudinal model:

                        coef      Se   Wald p-value
intercept class1     0.56644 1.32236  0.428 0.66839
intercept class2     1.01795 0.71059  1.433 0.15199
t class1            -0.01197 0.05062 -0.236 0.81306
t class2             0.03919 0.01591  2.463 0.01379
score_sz_imp_10      0.05327 0.02575  2.069 0.03856


Variance-covariance matrix of the random-effects:
          intercept       t
intercept   1.75459
t          -0.05518 0.00189
```

Fonction jointlcmm : problème de convergence

```
Iteration process:
     Maximum number of iteration reached without convergence
     Number of iterations:  100
     Convergence criteria: parameters= 0
                         : likelihood= 1e+09
                         : second derivatives= 1

Goodness-of-fit statistics:
     maximum log-likelihood: -215.73
     AIC: 463.46
     BIC: 498.73


Maximum Likelihood Estimates:

Fixed effects in the class-membership model:
(the class of reference is the last class)

                      coef Se Wald p-value
intercept class1   0.00000
sexe_bin class1    0.00000
```

Fonction jointlcmm : résolution des problèmes de convergence

On peut résoudre les problèmes de convergence des modèles ajustés en modifiant :

- Les paramètres de convergence : convB,convL, convG (par défaut 10^{-4}),
- Le nombre d'itérations : maxiter (par défaut 100),
- Le vecteur des valeurs initiales des paramètres : B.

Fonction jointlcmm : résolution des problèmes de convergence

On peut résoudre les problèmes de convergence des modèles ajustés en modifiant :

- Les paramètres de convergence : convB,convL, convG (par défaut 10^{-4}),
- Le nombre d'itérations : maxiter (par défaut 100),
- Le vecteur des valeurs initiales des paramètres : B.

Fonction jointlcmm : résolution des problèmes de convergence

On peut résoudre les problèmes de convergence des modèles ajustés en modifiant :

- Les paramètres de convergence : convB,convL, convG (par défaut 10^{-4}),
- Le nombre d'itérations : maxiter (par défaut 100),
- Le vecteur des valeurs initiales des paramètres : B.

Fonction jointlcmm : résolution des problèmes de convergence

On peut résoudre les problèmes de convergence des modèles ajustés en modifiant :

- Les paramètres de convergence : convB,convL, convG (par défaut 10^{-4}),
- Le nombre d'itérations : maxiter (par défaut 100),
- Le vecteur des valeurs initiales des paramètres : B.

Fonction jointlcmm : Résolution du problème de convergence

```
Maximum Likelihood Estimates:

Fixed effects in the class-membership model:
(the class of reference is the last class)

                      coef      Se   Wald p-value
intercept class1  -1.24278 0.96522 -1.288 0.19790
sexe_bin class1    1.90447 1.09110  1.745 0.08090

Parameters in the proportional hazard model:

                                        coef      Se   Wald p-value
event1 +/-sqrt(Weibull1) class 1    0.13858 0.02354  5.886 0.00000
event1 +/-sqrt(Weibull2) class 1    1.81689 0.41934  4.333 0.00001
event1 +/-sqrt(Weibull1) class 2    0.14946 0.01661  9.000 0.00000
event1 +/-sqrt(Weibull2) class 2    2.12150 0.45736  4.639 0.00000
```

Fonction jointlcmm : Suite de la résolution du problème de convergence

```
Fixed effects in the longitudinal model:

                              coef       Se    Wald p-value
intercept class1          -3.55825  1.00411  -3.544 0.00039
intercept class2           1.72433  0.98757   1.746 0.08081
t class1                   0.01080  0.02816   0.384 0.70129
t class2                   0.05093  0.02513   2.027 0.04269
score_sz_imp_10 class1    -0.14427  0.04039  -3.572 0.00035
score_sz_imp_10 class2     0.09065  0.03170   2.860 0.00424


Variance-covariance matrix of the random-effects:
          intercept       t
intercept   1.55586
t          -0.05637 0.00216

                             coef       Se
Residual standard error   0.71775  0.08503
```

multlcmm et mpjlcmm

- multlcmm
- mpjlcmm
  https://github.com/CecileProust-Lima/lcmm/tree/mpj

multlcmm et mpjlcmm

- multlcmm
- mpjlcmm
  https://github.com/CecileProust-Lima/lcmm/tree/mpj

multlcmm et mpjlcmm

- multlcmm
- mpjlcmm
  https://github.com/CecileProust-Lima/lcmm/tree/mpj

La classification aposteriori et les probabilités d'appartenance aux classes latentes

La fonction pprob permet d'avoir :

- La classification aposteriori selon les données longitudinales et le temps d'événement,
- Les probabilités d'appartenance individuelle à chaque classe latente.

La classification aposteriori et les probabilités d'appartenance aux classes latentes

La fonction pprob permet d'avoir :

- La classification aposteriori selon les données longitudinales et le temps d'événement,
- Les probabilités d'appartenance individuelle à chaque classe latente.

La classification aposteriori et les probabilités d'appartenance aux classes latentes

La fonction pprob permet d'avoir :

- La classification aposteriori selon les données longitudinales et le temps d'événement,
- Les probabilités d'appartenance individuelle à chaque classe latente.

Exemple

	identifiant	class	probYT1	probYT2
1	2962	1	1.000000e+00	8.699235e-25
2	2967	1	9.928730e-01	7.126986e-03
3	2968	2	5.726906e-12	1.000000e+00
4	2969	2	1.641901e-11	1.000000e+00
5	2970	1	1.000000e+00	1.812887e-12
6	2973	1	1.000000e+00	1.532902e-14
7	2975	2	3.761364e-06	9.999962e-01
8	2978	1	9.981270e-01	1.873018e-03
9	2990	2	2.644583e-01	7.355417e-01
10	4703	2	4.793673e-12	1.000000e+00
11	4705	1	1.000000e+00	6.279319e-14
12	4707	1	9.969627e-01	3.037259e-03
13	4709	1	1.000000e+00	9.691614e-09
14	4710	1	1.000000e+00	8.897081e-30

Risque cumulatif individuel à un certain t

- Soit un sujet donnée i, i(=1,...,N),
  On peut calculer le risque cumulatif pour le sujet i à un certain âge donné t :

$$F_i(t/covariables) = \sum_{l=1}^{G} proba(c_i = l) * G(t/l)$$

Risque cumulatif individuel à un certain t

- Soit un sujet donnée i, i(=1,...,N),
  On peut calculer le risque cumulatif pour le sujet i à un certain âge donné t :

$$F_i(t/covariables) = \sum_{l=1}^{G} proba(c_i = l) * G(t/l)$$

Risque cumulatif individuel à un certain t

- Soit un sujet donnée i, i(=1,...,N),
  On peut calculer le risque cumulatif pour le sujet i à un certain âge donné t :

$$F_i(t/covariables) = \sum_{l=1}^{G} proba(c_i = l) * G(t/l)$$

Exemple

- On choisit t=36 ans, calculons le risque cumulatif de développer la maladie à cet âge pour un échantillon de nos sujets étudiés :

	identifiant	probabilite aposteriori a 36	classe
1	2962	4.635042e-09	1
2	2967	5.630017e-03	1
3	2968	7.899570e-01	2
4	2969	7.899570e-01	2
5	2970	4.636474e-09	1
6	2973	4.635054e-09	1
7	2975	7.899541e-01	2
8	2978	1.479609e-03	1
9	2990	5.810463e-01	2
10	4703	7.899570e-01	2
11	4705	4.635091e-09	1
12	4707	2.399308e-03	1
13	4709	1.229100e-08	1
14	4710	4.635042e-09	1

Références

- Joint modelling of multivariate longitudinal outcomes and a time-to-event : a nonlinear latent class approach
  Cécile Proust-Lima, Pierre Joly, Jean-François Dartigues, Hélène Jacqmin-Gadda
- Joint latent class models for longitudinal and time-to-event data : A review
  Cécile Proust-Lima, Mbéry Séne, Jeremy MG Taylor and Hélène Jacqmin-Gadda
- https://cran.r-project.org/web/packages/lcmm/lcmm.pdf
- https://arxiv.org/pdf/1503.00890.pdf