Multiple Trajectory Analysis in Finite Mixture Modeling

Jang SCHILTZ (University of Luxembourg)

joint work with

Cédric NOEL(University of Lorraine & University of Luxembourg)

ASMDA 2023
June 8, 2023

Outline

1. Finite Mixture Models

Outline

Outline

Outline

Outline

Outline

General description of Finite Mixture models

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous sub-populations and to estimate, at the same time, a typical trajectory for each sub-population. (Nagin 2005, Schiltz 2015)

This model can be interpreted as functional fuzzy cluster analysis.

The basic model (Nagin 2005)

Consider a population of size N and a variable of interest Y.
Let $Y_i = y_{i_1}, y_{i_2}, ..., y_{i_T}$ be T measures of the variable, taken at times $t_1, ...t_T$ for subject number i and π_k the probability of a given subject to belong to group number k

For a given group G_k, we suppose conditional independence for the sequential realizations of the elements y_{i_t} over the T periods of measurements.

The density f of Y is given by

$$f(y_i; \psi) = \sum_{k=1}^{K} \pi_k g^k(y_i; \Theta_k), \qquad (1)$$

where $g^k(\cdot)$ denotes the distribution of y_{it} conditional on membership in group k and the role of the parameters Θ_k is to describe the shape of the trajectories in group k.

Predictors of trajectory group membership

x : vector of variables potentially associated with group membership (measured before t_1).

Multinomial logit model:

$$\pi_k(x_i) = \frac{e^{x_i\theta_k}}{\sum_{k=1}^{K} e^{x_i\theta_k}}, \qquad (2)$$

where θ_k denotes the effect of x_i on the probability of group membership for group k.

$$L = \prod_{i=1}^{N} \sum_{k=1}^{K} \frac{e^{x_i\theta_k}}{\sum_{k=1}^{K} e^{x_i\theta_k}} \prod_{t=1}^{T} p^k(y_{it}), \qquad (3)$$

where $p^k(\cdot)$ denotes the distribution of y_{it} conditional on membership in group k.

Adding covariates to the trajectories

Let W be a vector of covariates potentially influencing Y.

The likelihood then becomes

$$L = \prod_{i=1}^{N} \sum_{k=1}^{K} \frac{e^{x_i \theta_k}}{\sum_{k=1}^{K} e^{x_i \theta_k}} \prod_{t=1}^{T} p^k(y_{it} | A_i, W_i, \Theta_k).$$

Possible data distributions

- Poisson distribution
- Binary logit distribution
- Censored normal distribution
- Beta distribution

Outline

Basic Idea

We conjointly analysis the trajectories of J variables $Y^1, ..., Y^J$.

We suppose that the time series $(Y^j)_{1 \le j \le J}$ are independently distributed, conditional on the group memberships of the different time series, so that $P^{k_1 \dots k_J}(Y_i^1, \dots, Y_i^J) = \prod_{j=1}^J P^{k_j}(Y_i^j)$.

The likelihood function of the data can then we written as

$$P(Y_i^1, \dots, Y_i^J | A_i, W_i, X_i) = \sum_{(k_1, \dots, k_J) \in K_1 \times \dots \times K_J} \pi_{k_1 \dots k_J} \prod_{j=1}^{J} P^{k_j}(Y_i^j | A_i, W_i, \Theta_k^j)$$

Nagin's multitrajectory model

$$P(Y_i^1, \ldots, Y_i^J | A_i, W_i) = \sum_{(k_1,\ldots,k_J) \in K_1 \times \cdots \times K_J} \pi_{k_J | k_1 \ldots k_{J-1}} \times \cdots \times \pi_{k_2 | k_1} \times \pi_{k_1}$$
$$\prod_{j=1}^{J} \prod_{t=1}^{T} p^{k_j}(y_{it}^j | A_i, W_i, \Theta_k^j),$$

where $\pi_{k_j | k_1 \ldots k_{j-1}}$ is the probability of belonging to group j conditional on the membership to groups 1 to $j-1$.

Membership Probability

$$\pi_{k_1} = \frac{e^{\theta_{k_1} x_i}}{\sum_{k_1=1}^{K_1} e^{\theta_{k_1} x_i}}, \quad \pi_{k_2|k_1} = \frac{e^{\theta_{k_2}^{k_1} w_i^{k_2}}}{\sum_{k_2=1}^{K_2} e^{\theta_{k_2}^{k_1} w_i^{k_2}}}, \quad \cdots \quad ,$$

$$\pi_{k_J|k_1 \ldots k_{J-1}} = \frac{e_{k_J}^{\theta^{k_1 \cdots k_{J-1}} w_i^{k_J}}}{\sum_{k_J=1}^{K_J} e^{\theta_{k_J}^{k_1 \cdots k_{J-1}} w_i^{k_J}}}.$$

Drawbacks of this method is the absence of symmetry in the treatment of the variables and the numerical complications to compute the parameters in the case of more than two time series.

Outline

Group Membership Vector (Bel & Paap 2014)

Denote by $Z_i = (Z_{i1}, \ldots, Z_{iJ})$ the vector containing the group membership of individual i for the variables $Y^1, \ldots, Y^J$. $Z_i \in [\![1; K_1]\!] \times \cdots \times [\![1; K_J]\!]$.

Then,

$$P\left(Z_{ij} = k | z_{ih} \text{ for } h \neq j, X_i^j\right) = \frac{e^{B_{ij,k}}}{\sum_{h=1}^{K_j} e^{B_{ij,h}}},$$

where $B_{ij,k} = \alpha_{j,k} + \beta_{j,k} X_i^j + \sum_{h \neq j} \psi_{jh,kz_{ih}}$.

- $\alpha_{j,k}$ is a choice specific intercept ;
- $\beta_{j,k}$ is a vector corresponding to the covariate X_i^j ;
- z_{ih} the group membership of the individual i for Y^h ;
- $\psi_{jh,kl}$ is an association parameter between belonging to group k for Y^j and belonging to the group l for Y^h.

Model likelihood

$$L(Y_i^1, \ldots, Y_i^J | A_i, W_i, X_i)$$
$$= \sum_{(k_1,\ldots,k_J) \in K_1 \times \cdots \times K_J} \pi_{k_1 \ldots k_J} \prod_{j=1}^{J} \prod_{t=1}^{T^j} g^{k_j}(y_{it}^j | A_i, W_i, \Theta_k^j),$$

with

$$\pi_{k_1 \ldots k_J} = \frac{e^{\mu_{\boldsymbol{k}}}}{\sum_{\tilde{\boldsymbol{k}} \in S} e^{\mu_{\tilde{\boldsymbol{k}}}}},$$

where $\mu_{\boldsymbol{k}} = \sum_{j=1}^{J} \left(\theta_{k_j}^j X_i^j + \sum_{h<j} \psi_{hj,k_h k_j} \right).$

Parameter estimation: direct differentiation

For any parameter ζ,

$$\frac{\partial l}{\partial \zeta} = \sum_{i=1}^{n} \frac{\sum_{(k_1,\ldots,k_J)\in K_1\times\cdots\times K_J} \frac{\partial \pi_{k_1\ldots k_J}}{\partial \zeta} d_{\boldsymbol{k}} + \sum_{(k_1,\ldots,k_J)\in K_1\times\cdots\times K_J} \pi_{k_1\ldots k_J} \frac{\partial d_{\boldsymbol{k}}}{\partial \zeta}}{\sum_{(k_1,\ldots,k_J)\in K_1\times\cdots\times K_J} \pi_{k_1\ldots k_J} \prod_{j=1}^{J}\prod_{t=1}^{T^j} g^{k_j}(y_{it}^j | A_i, W_i, \Theta_k^j)},$$

where

$$d_{\boldsymbol{k}} = \prod_{j=1}^{J}\prod_{t=1}^{T^j} g^{k_j}(y_{it}^j | A_i, W_i, \Theta_k^j). \tag{4}$$

Parameter estimation: direct differentiation

$$\frac{\partial \pi_{\boldsymbol{k}}}{\partial \theta^j_{k_j}} d_{\boldsymbol{k}} = \sum_{i=1}^{n} X_i^j \left(\sum_{k_j \in \boldsymbol{k}} \pi_{\boldsymbol{k}} d_{\boldsymbol{k}} - \left(\sum_{k_j \in \boldsymbol{k}} \pi_{\boldsymbol{k}} \right) \sum_{\boldsymbol{k}} \pi_{\boldsymbol{k}} d_{\boldsymbol{k}} \right)$$

and

$$\frac{\partial \pi_{\boldsymbol{k}}}{\partial \psi_{jh,k_j k_h}} d_{\boldsymbol{k}} = \sum_{i=1}^{n} \left(\sum_{k_j, k_h \in \boldsymbol{k}} \pi_{\boldsymbol{k}} d_{\boldsymbol{k}} - \left(\sum_{k_j, k_h \in \boldsymbol{k}} \pi_{\boldsymbol{k}} \right) \sum_{\boldsymbol{k}} \pi_{\boldsymbol{k}} d_{\boldsymbol{k}} \right).$$

Here $\pi_{\boldsymbol{k}} = \pi_{k_1 \ldots k_J}$.

Parameter estimation: EM algorithm

We have for each step in the algorithm EM

$$E(z_{i\boldsymbol{k}}|\boldsymbol{Y}_i) = \tau_{i\boldsymbol{k}} = \frac{\pi_{\boldsymbol{k}} \prod_{j=1}^{J} \prod_{t=1}^{T^j} g^{k_j}(y_{it}^j|A_i, W_i, \Theta_k^j)}{\sum_{\boldsymbol{k}\in\boldsymbol{K}} \pi_{\boldsymbol{k}} \prod_{j=1}^{J} \prod_{t=1}^{T^j} g^{k_j}(y_{it}^j|A_i, W_i, \Theta_k^j)}$$

and

$$Q = \sum_{i=1}^{n} \sum_{\boldsymbol{k}\in\boldsymbol{K}} \tau_{i\boldsymbol{k}} \left[\log(\pi_{\boldsymbol{k}}) + \sum_{j=1}^{J} \sum_{t=1}^{T^j} \log\left(g^{k_j}(y_{it}^j|A_i, W_i, \Theta_k^j) \right) \right].$$

EM algorithm: problem to solve

$$argmin_{\Theta_k^j} \sum_{i=1}^{n} \sum_{\boldsymbol{k} \in \boldsymbol{K}} \sum_{t=1}^{T^j} \tau_{i\boldsymbol{k}} \log\left(g^{k_j}(y_{it}^j | A_i, W_i, \Theta_k^j)\right)$$

$$= argmin_{\Theta_k^j} \sum_{i=1}^{n} \left[\sum_{t=1}^{T^j} \log\left(g^{k_j}(y_{it}^j | A_i, W_i, \Theta_k^j)\right)\right] \left[\sum_{\substack{\boldsymbol{k} \in \boldsymbol{K} \\ k_j \in \boldsymbol{k}}} \tau_{i\boldsymbol{k}}\right]$$

Parameter estimation: EM algorithm

Then,

$$\theta_k^j \text{ is a root of } \theta_k^j :\mapsto \sum_{i=1}^{n} \sum_{\substack{\boldsymbol{k}\in\boldsymbol{K} \\ k\in\boldsymbol{k}}} x_{il}^j \left(\tau_{i\boldsymbol{k}} - \pi_{i\boldsymbol{k}}\right)$$

$$\psi_{jh,k_j k_l} \text{ is a root of } \psi_{jh,k_j k_l} :\mapsto \sum_{i=1}^{n} \sum_{\substack{\boldsymbol{k}\in\boldsymbol{K} \\ k_j,k_l\in\boldsymbol{k}}} \left(\tau_{i\boldsymbol{k}} - \pi_{i\boldsymbol{k}}\right).$$

Outline

Function signature

```
R> multitrajeR(Y, degre.Y, A ,RISK = NULL, TCOV = NULL, degre, degre.phi=0,
+         Model = "BETA", Method = "L",
+         ssigma = FALSE, itermax = 100, ymax = max(Y) + 1, ymin = min(Y) - 1,
+         hessian = TRUE, itermax = 100, paraminit = NULL,
+         control = list(fnscale=-1,  trace=1, REPORT=1,maxit = 100),
+         ProbIRLS = TRUE, refgr = 1, + fct = NULL, diffct = NULL, nbvar = NULL, nls.limite
```

Remark:

trajeR only works, if Y is a matrix!

`In case of doubt, you should start your code with`

```
Y<-as.matrix(Y)
```

Outline

Montreal Longintudinal Study

Example from D. Nagin. Compares the link between hyperactivity and opposition score. The hyperactivity is measured on a scale between 0 and 4 and the opposition behavior on a scale between 0 and 10.

Results

	Variable 1	
	TrajeR	traj (SAS)
β_{11}	-2.55343	-2.55348
β_{21}	0.48071	0.48074
β_{22}	-6.24514	-6.24425
β_{23}	-3.53665	-3.53676
β_{24}	14.90876	14.90347
β_{31}	2.66416	2.66411
β_{32}	-1.98842	-1.98840
β_{33}	-4.14192	-4.14164
σ	2.31949	2.3195

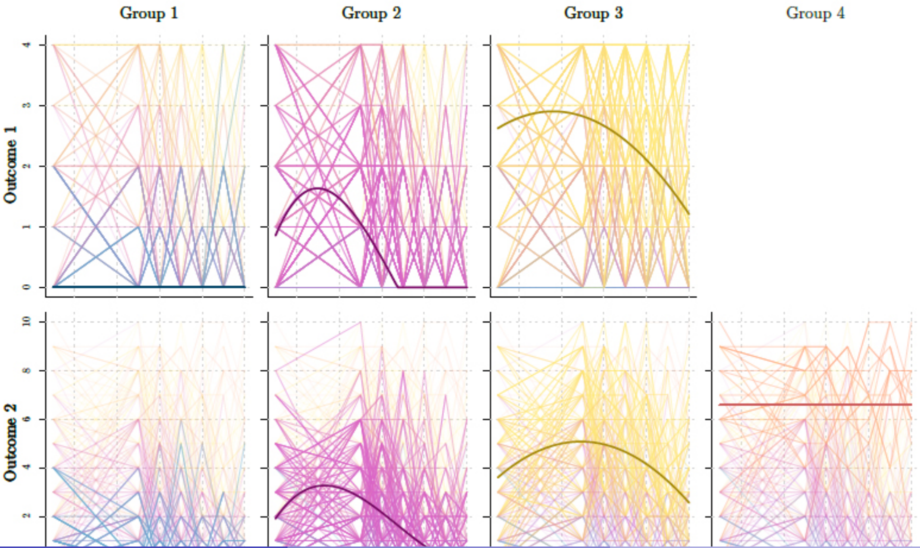
	Variable 2	
	TrajeR	traj (SAS)
β_{11}	-1.67252	-1.67258
β_{12}	-1.48300	-1.48308
β_{21}	2.15443	2.15440
β_{22}	-6.79570	-6.79186
β_{23}	-4.29020	-4.29356
β_{24}	20.58775	20.56512
β_{31}	4.96622	4.96615
β_{32}	-2.12531	-2.12503
β_{33}	-9.72094	-9.71914
β_{41}	6.59242	6.59246
σ	2.54359	2.5436

With `trajeR` we find the following 6 linking parameters.

Parameters	$\psi_{12,22}$	$\psi_{12,23}$	$\psi_{12,24}$	$\psi_{12,32}$	$\psi_{12,33}$	$\psi_{12,34}$
Theoretical	19.97949	10.51569	8.94481	28.44746	31.69029	40.0845

Results

–



Group 1
Group 2
Group 3
Group 4
Outcome 1
Outcome 2
0
1
2
3
4
2
4
6
8
10

Bibliography

- Nagin, D.S. 2005: *Group-based Modeling of Development*. Cambridge, MA.: Harvard University Press.
- Schiltz, J. 2015: A generalization of Nagin's finite mixture model. In: Dependent data in social sciences research: Forms, issues, and methods of analysis' Mark Stemmler, Alexander von Eye & Wolfgang Wiedermann (Eds.). Springer 2015.
- Nagin, D.S., Jones, B.L., Lima Passos, V. & Tremblay, R.E. 2018: Group-based multi-trajectory modeling. *Statistical Methods in Medical Research*, 27-7.
- Noel, C & Schiltz, J. 2022: TrajeR - an R package for finite mixture models. SSRN paper 4054519.
- Noel, C & Schiltz, J. 2023: Multitrajectory Analysis in Finite Mixture Models.