Multiple Trajectory Analysis in Finite Mixture Modeling

Jang SCHILTZ (University of Luxembourg)

joint work with

Cédric NOEL(University of Lorraine & University of Luxembourg)

ASMDA 2021
June 2, 2021

Outline

Outline

Outline

Outline

Outline

Outline

General description of Nagin's model

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.

Hence, this model can be interpreted as functional fuzzy cluster analysis.

Finite mixture model (Daniel S. Nagin (Carnegie Mellon University))

- mixture : population composed of a mixture of unobserved groups
- finite : sums across a finite number of groups

The Likelihood Function (1)

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. $A_i = \{t_1, ...t_T\}$

π_k : probability of a given subject to belong to group number k

$$\Rightarrow \pi_k \text{ is the size of group } k.$$

$$\Rightarrow P(Y_i) = \sum_{k=1}^{K} \pi_k P^k(Y_i), \quad (1)$$

where $P^k(Y_i)$ is probability of Y_i if subject i belongs to group k.

The Likelihood Function (2)

Aim of the analysis: Find K groups of trajectories of a given kind, for instance polynomials of degree 4, $P(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$.

Statistical Model:

$$y_{i_t} = \beta_0^k + \beta_1^k t + \beta_2^k t^2 + \beta_3^k t^3 + \beta_4^k t^4 + \varepsilon_{i_t}^k, \tag{2}$$

where $\varepsilon_{i_t}^k \sim \mathcal{N}(0, \sigma_k)$, σ_k being the standard deviation, constant inside group k.

We try to estimate a set of parameters $\Omega = \left\{\beta_0^k, \beta_1^k, \beta_2^k, \beta_3^k, \beta_4^k, \pi_k, \sigma_k\right\}$ which allow to maximize the probability of the measured data.

Possible data distributions

- count data $\Rightarrow$ Poisson distribution
- binary data $\Rightarrow$ Binary logit distribution
- censored data $\Rightarrow$ Censored normal distribution

Predictors of trajectory group membership

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

Multinomial logit model:

$$\pi_k(x_i) = \frac{e^{x_i\theta_k}}{\sum_{k=1}^{K} e^{x_i\theta_k}}, \tag{3}$$

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

$$L = \prod_{i=1}^{N} \sum_{k=1}^{K} \frac{e^{x_i\theta_k}}{\sum_{k=1}^{K} e^{x_i\theta_k}} \prod_{t=1}^{T} p^k(y_{it}), \tag{4}$$

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

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

Outline

Basic Idea

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

The constrained model

We suppose the existence of K unique groups representing the combined development of Y^j, $1 \leq j \leq J$.

The conditional independence of the time measurements for the outcomes for a given group is implies that

$$P^{(}Y_i^1, \ldots, Y_i^J | A_i, W_i) = \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} \prod_{t=1}^{T} p^k(y_{it}^j | A_i, W_i, \Theta_k^j).$$

The unconstrained model

We suppose the trajectories for a variable Y^l can be linked to trajectories for all other variables $Y^j, j \neq l$. Then,

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

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

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

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

One drawback of this method is the great expansion of the number of parameters and the fact that the parameters are hardly interpretable.

Outline

Group Membership Vector

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

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.

Number of parameters

The Hammersley-Clifford Theorem allows to write the conditional probabilities as

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

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

Proposition

The numbers of parameters is

$$\sum_{j=1}^{J} (K_j - 1) \times (ncol(X_i^j) + 1) + \sum_{1 \leq j \neq j' \leq J} (K_j - 1)(K_{j'} - 1).$$

Outline

Setup

We simulate trajectories for 200 individuals and 3 variables.

- Y^1 (normal distribution) : $\beta_{1,1} = (3.53, -2.25, 0.47)$, $\beta_{1,2} = (-1.62, 3.9, -0.65)$, $\beta_{1,3} = (0.263, 0.036, 0.01)$, $\sigma_{1,1} = \sigma_{1,2} = \sigma_{1,3} = 1$;
- Y^2 (ZIP distribution) : $\beta_{2,1} = (1.2, 2.3, -1.2, 0.5, -0.1)$, $\beta_{2,2} = (2)$, $\beta_{2,3} = (-7.5, 0, 2.2, -.4)$, $\nu_1 = (-2, 1)$, $\nu_2 = (-1, 0.1)$, $\nu_3 = (0, -1)$;
- Y^3 (logit distribution) : $\beta_{3,1} = (6.32, -5.8, 1)$, $\beta_{3,2} = (-6.69, 1.92)$.

Furthermore, we choose all $\theta_{j,k} = 0$ and $\psi = (-3, 3, 4, 0, -2, 5, 1, 0)$. We launch trajeR for each variable separately and we use the results as as initial values for the multi-trajectory model.

Results

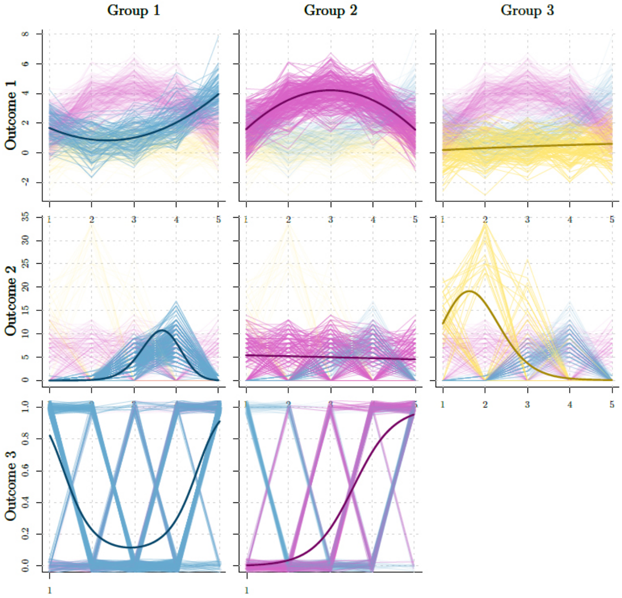
	Variable 1		
	Th.	EM	L
β_{11}	3.53	3.37162	3.37154
β_{12}	-2.25	-2.13530	-2.13523
β_{13}	0.47	0.45090	0.45089
β_{21}	-1.620	-1.71513	-1.71513
β_{22}	3.900	3.96422	3.96422
β_{23}	-0.650	-0.66183	-0.66183
β_{31}	0.263	0.04619	0.04624
β_{32}	0.036	0.15730	0.15725
β_{33}	0.010	-0.00882	-0.00881
σ	1	2.69943	2.69943
θ_1	0	0.00000	0.00000
θ_2	0	2.00645	2.00672
θ_3	0	-5.40659	-5.43048

	Variable 2		
	Th.	EM	L
β_{11}	1.2	-3.92980	-3.92950
β_{12}	2.3	-4.11672	-4.11653
β_{13}	-1.2	3.88076	3.88051
β_{14}	0.5	-0.68312	-0.68305
β_{15}	-0.1	0.01616	0.01615
β_{21}	2	2.01483	2.01483
β_{31}	-7.5	-0.39232	-0.39204
β_{32}	0	4.55945	4.55895
β_{33}	2.2	-1.69650	-1.69623
β_{34}	-0.4	0.17702	0.17698
ν_{11}	-2	-0.98558	-0.98296
ν_{12}	1	-0.7915	-0.79232
ν_{21}	-1	-1.08749	-1.08734
ν_{22}	0.1	0.13039	0.13035
ν_{31}	0	-3.38530	-3.38515
ν_{32}	-1	1.53658	1.53652
θ_1	0	0.00000	0.00000
θ_2	0	-0.07507	-0.07461
θ_3	0	0.14593	0.14628

	Variable 3		
	Th.	EM	L
β_{11}	6.32	6.28035	6.28024
β_{12}	-5.8	-5.73551	-5.73545
β_{13}	1	0.98963	0.98962
β_{21}	-6.69	-7.46613	-7.46609
β_{22}	1.92	2.09930	2.09930
θ_1	0	0.00000	0.00000
θ_2	0	-2.09538	-2.09295

Parameters	$\psi_{12,22}$	$\psi_{12,23}$	$\psi_{12,32}$	$\psi_{12,33}$	$\psi_{13,22}$	$\psi_{13,32}$	$\psi_{23,22}$	$\psi_{23,32}$
Theoretical	-3	3	4	0	-2	5	1	0
EM	-5.47292	4.33422	-4.57086	-4.46658	-2.12297	3.23218	1.18008	-7.96007
Likelihood	-5.46968	4.35722	-4.57081	-3.3298	-2.12013	3.23351	1.17703	-10.00915

Results



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

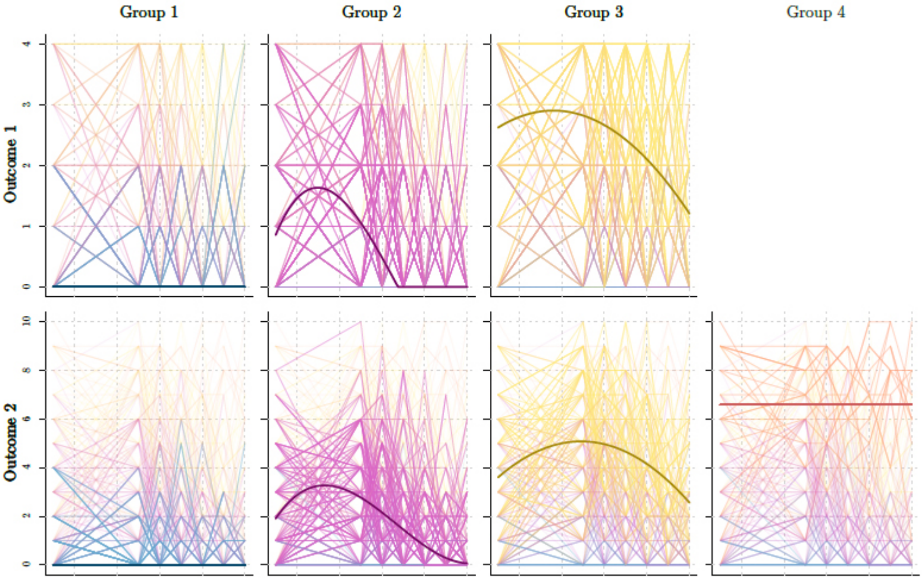
	Variable 1	
	TrajeR	traj (SAS)
β_{11}	-2.55343	-2.55348
β_{21}	0.48071	0.48074
β_{22}	-6.24514	-6.24425
β_{23}	-3.53665	-3.53676
β_{24}	14.90876	14.90347
β_{31}	2.66416	2.66411
β_{32}	-1.98842	-1.98840
β_{33}	-4.14192	-4.14164
σ	2.31949	2.3195

	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



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. 2021: TrajeR - an R package for finite mixture models.
- Noel, C & Schiltz, J. 2021: A new algorithm for group-based multi-trajectory modeling.