New results in finite mixture modeling

Jang SCHILTZ (University of Luxembourg)

joint work with

Cédric NOEL(University of Lorraine & University of Luxembourg)

Luxembourg-Waseda Conference
January, 25 2023

Outline

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)

This model can be interpreted as functional fuzzy logic 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

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

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.

Possible data distributions

- Poisson distribution
- Binary logit distribution
- (Censored) normal distribution
- Beta distribution (Noel & S. 2023)

Likelihood function for a normal distribution

Notations:

- $\beta^j t = \beta_0^j + \beta_1^j t + \beta_2^j t^2 + \beta_3^j t^3 + \beta_4^j t^4$.
- ϕ: density of standard centered normal law.

Then,

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_j \prod_{t=1}^{T} \phi\left(\frac{y_{i_t} - \beta^j t}{\sigma}\right). \quad (2)$$

It is too complicated to get closed-forms equations.

SAS procedure proc Traj (Nagin & Jones 2008).

Predictors of trajectory group membership

X : vector of variables potentially associated with group membership.

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

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

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

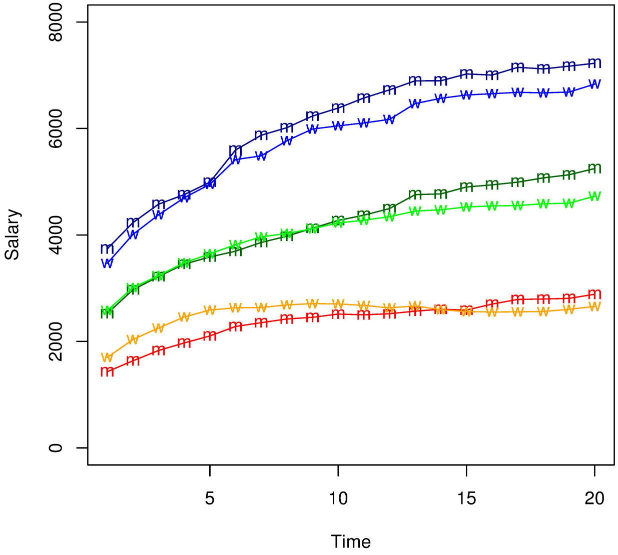
Adding covariates to the trajectories

Let W be a vector of covariates potentially influencing Y.

The likelihood then becomes

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

Salary groups: Men versus women (S. 2016)



Generalized finite mixture model (S. 2015)

Let $x_1 ... x_M$ and z_t be covariates potentially influencing Y.

We propose the following model:

$$y_{i_t} = \left(\beta_0^j + \sum_{l=1}^{M} \alpha_{0l}^j x_{i_l} + \gamma_0^j w_{i_t}\right) + \left(\beta_1^j + \sum_{l=1}^{M} \alpha_{1l}^j x_{i_l} + \gamma_1^j w_{i_t}\right) t + \left(\beta_2^j + \sum_{l=1}^{M} \alpha_{2l}^j x_{i_l} + \gamma_2^j w_{i_t}\right) t^2 + \left(\beta_3^j + \sum_{l=1}^{M} \alpha_{3l}^j x_{i_l} + \gamma_3^j w_{i_t}\right) t^3 + \left(\beta_4^j + \sum_{l=1}^{M} \alpha_{4l}^j x_{i_l} + \gamma_4^j w_{i_t}\right) t^4 + \varepsilon_{i_t}^j,$$

where $\varepsilon_{i_t} \sim \mathcal{N}(0, \sigma^j)$, σ^j being the standard deviation, constant in group j.

Statistical Properties (S. 2015)

The model's estimated parameters are the result of maximum likelihood estimation. As such, they are consistent and asymptotically normally distributed.

Confidence intervals of level α for the parameters β_k^j:

$$CI_\alpha(\beta_k^j) = \left[\hat{\beta}_k^j - t_{1-\alpha/2;N-(2+M)s} ASE(\hat{\beta}_k^j); \hat{\beta}_k^j + t_{1-\alpha/2;N-(2+M)s} ASE(\hat{\beta}_k^j)\right]. \quad (5)$$

Confidence intervals of level α for the disturbance factor σ_j:

$$CI_\alpha(\sigma_j) = \left[\sqrt{\frac{(N-(2+M)s-1)\hat{\sigma_j}^2}{\chi^2_{1-\alpha/2;N-(2+M)s-1}}}; \sqrt{\frac{(N-(2+M)s-1)\hat{\sigma_j}^2}{\chi^2_{\alpha/2;N-(2+M)s-1}}}\right]. \quad (6)$$

Outline

trajeR (Noel & S. 2022): Function signature

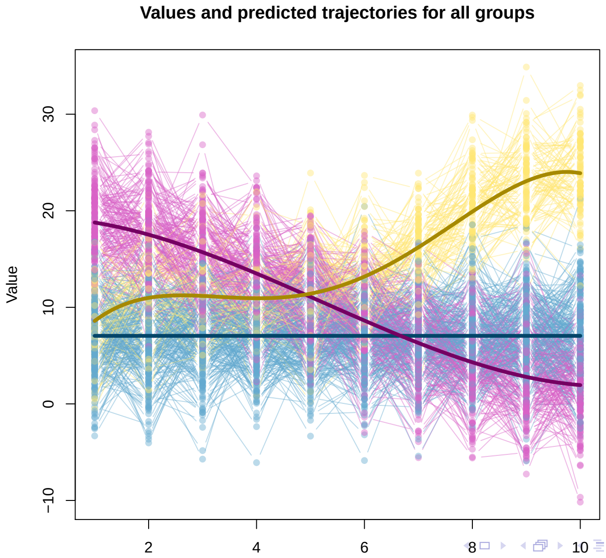
```
R> trajeR(Y, A, Risk = NULL, TCOV = NULL, degre, degre.phi = 0,
+         Model, Method = "L",
+         ssigma = FALSE, ymax = max(Y) + 1, ymin = min(Y) - 1,
+         hessian = TRUE, itermax = 100, paraminit = NULL,
+         ProbIRLS = TRUE, refgr = 1, + fct = NULL, diffct = NULL,nbvar = NULL, nls.lmiter = 50)
```

Numerical output of result

```
## Model : Beta
## Method : Likelihood
##
##   group   Parameter   Estimate   Std. Error   T for H0:   Prob>|T|
##                                                param.=0
## -----------------------------------------------------------------
##    mean
##       1   Intercept   -5.95316      0.1281   -46.4734          0
##              Linear    3.66558     0.07649    47.92297          0
##           Quadratic   -0.49316     0.01027   -48.04232          0
##    zeta
##       1   Intercept    2.26533      0.0993   22.81197           0
##              Linear   -0.00558     0.02466    -0.22636    0.82094
##
##    mean
##       2   Intercept    3.73504     0.04525    82.53444          0
##              Linear   -0.98061     0.01144   -85.70519          0
##    zeta
##       2   Intercept    2.35458     0.07128    33.03302          0
##              Linear   -0.00144     0.01771    -0.08113    0.93534
## -----------------------------------------------------------------
##       1         pi1      0.344     0.02069           0          0
##       2         pi2      0.656     0.02069    31.19708          0
## -----------------------------------------------------------------
## Likelihood : 2516.737
```

trajeR - Censored Normal Model

```
plot(solL, Y = data[,2:11], A = data[,12:21], col = vcol)
```

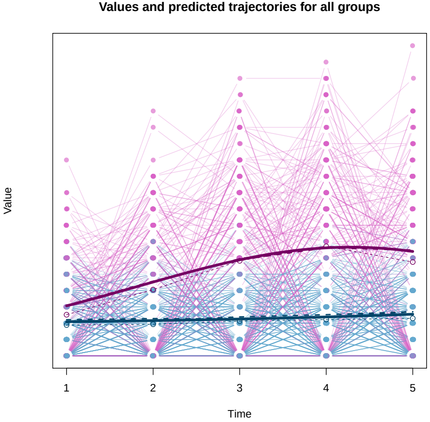


trajeR - Zero Inflated Poisson

```
Call TrajeR with 2 groups and a 2,2 degrees of polynomial shape of trajec
tory.
Model : Zero Inflated Poisson
Method : Likelihood

  group   Parameter   Estimate   Std. Error   T for H0:   Prob>|T|
                                               param.=0
-------------------------------------------------------------------
      1   Intercept    1.04873      0.09415    11.13945          0
             Linear    0.61506      0.06423     9.57663          0
          Quadratic   -0.07773      0.00998    -7.79183          0
               Nu11    0.03161      0.16749     0.18871    0.85034
               Nu12   -0.16776      0.05183    -3.23661    0.00123

      2   Intercept    0.90787      0.09589      9.4679          0
             Linear    0.01869      0.07032      0.2658    0.79042
          Quadratic    0.00619      0.01131     0.54724    0.58426
               Nu21   -1.08139      0.16018    -6.75111          0
               Nu22    0.00452      0.04689     0.09632    0.92327
-------------------------------------------------------------------
      1         pi1    0.34869      0.04936    13.93138          0
      2         pi2    0.65131      0.04936    26.59044          0
-------------------------------------------------------------------
Likelihood : -5162.009
```



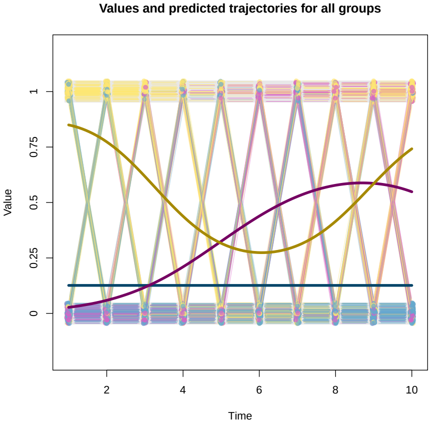
trajeR - Logit

```
Call TrajeR with 3 groups and a 0,3,4 degrees of polynomial shape of traj
ectory.
Model : Logit
Method : Likelihood

  group   Parameter   Estimate   Std. Error   T for H0:    Prob>|T|
                                               param.=0
-------------------------------------------------------------------
      1   Intercept   -1.94482      0.10059   -19.33387          0

      2   Intercept   -1.77021      0.03535   -50.07501          0
             Linear   -0.44043          NaN         NaN        NaN
          Quadratic    0.18687          NaN         NaN        NaN
              Cubic   -0.01232          NaN         NaN        NaN

      3   Intercept    1.74626          NaN         NaN        NaN
             Linear     0.6443      0.07247     8.89028          0
          Quadratic    -0.5554      0.02475   -22.44466          0
              Cubic    0.08204      0.00195    42.07317          0
          Quadratic   -0.00336          NaN         NaN        NaN
-------------------------------------------------------------------
      1         pi1    0.41816      0.10875     2.41645    0.01571
      2         pi2    0.21915      0.11423    -3.35532      8e-04
      3         pi3    0.36269       0.0623     1.93388    0.05318
-------------------------------------------------------------------
Likelihood : -2852.749
```



trajeR - Non Linear

We suppose that the variable Y_{it} is defined by

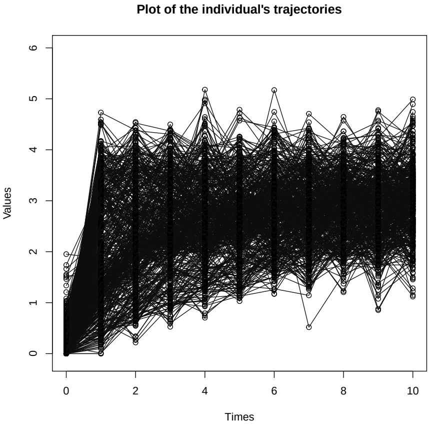
$$y_{it} = f(a_{it}; \beta_k) + \epsilon_{it} \tag{7}$$

where $\epsilon_{it} \sim \mathcal{N}(0;\ \sigma_k)$.

trajeR - Non Linear

Example with

$$f(t;\beta_k) = \frac{\beta_{k1} t}{\beta_{k2} + t}$$



By analysing the graph above, we fit initial parmaters as paraminit=**c**(0.25,0.25,0.25,0.25,2, 0.1,2.4,0.1,2.8,0.1,3,0.1,0.2,0.2,0.2,0.2)

trajeR - Non Linear

```
fct <- function(t, betak, TCOV){ return( (betak[1]*t)/(betak[2]+t) )
diffct <- function(t, betak, TCOV){ return(c( t/(betak[2]+t),
-(betak[1]*t)/(betak[2]+t)**2 ))

solEM = trajeR(Y = data[,2:12], A = data[,13:23], ng = 4, nbvar = 2,
Method = "EM", Model = "NL", hessian = TRUE, fct = fct, diffct =
diffct, paraminit = paraminit)
```

trajeR - Non Linear

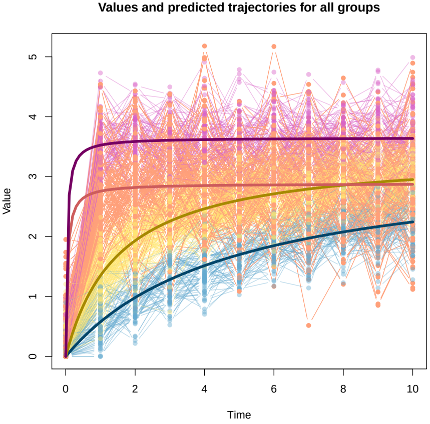
```
Call TrajeR with 4 groups and a 1,1,1,1 degrees of polynomial shape of tr
ajectory.
Model : Non Linear
Method : Expectation-maximization

  group   Parameter   Estimate   Std. Error   T for H0:    Prob>|T|
                                               param.=0
-----------------------------------------------------------------------
      1   Intercept    3.30542     0.18506     17.86117           0
             Linear    4.71859     0.60508      7.79825           0

      2   Intercept    3.17397     0.06338     50.07605           0
             Linear    1.29635     0.12496     10.37439           0

      3   Intercept    3.07645     0.06425     47.88329           0
             Linear    0.60537     0.11135      5.43645           0

      4   Intercept    3.34255     0.02515    132.91026           0
             Linear    0.04503      0.0236      1.90772     0.05648
-----------------------------------------------------------------------
      1      sigma1    0.63465         NaN          NaN         NaN
      2      sigma2    0.56821      0.0036    157.93961           0
      3      sigma3    0.60047     0.02216      27.0945           0
      4      sigma4    0.71747     0.01667      43.0461           0
-----------------------------------------------------------------------
      1         pi1    0.15427     0.01646       9.3702           0
      2         pi2    0.23712     0.02962      8.00592           0
      3         pi3    0.25955         NaN          NaN         NaN
      4         pi4    0.34906         NaN          NaN         NaN
-----------------------------------------------------------------------
Likelihood : -4891.536
```



trajeR - Some functions

- Membership's probability – `GroupProb(...)`
- Profiles of group – `GroupProfiles(...)`

	Gr 1	Gr 2	Gr 3	Gr 4
X1	1.8467	3.2709	6.0562	6.5621
X2	7.7916	-0.9212	-4.9915	5.3515
X3	56.1620	74.0841	41.0704	32.8957
X4	0.3636	0.2895	0.4972	0.3333
X5	0.2545	0.3289	0.4525	0.3421

trajeR - Model performance

- Average Posterior Probability – `AvePP(...)`
- Odds of Correct Classification – `OCC(...)`
- Estimated group probabilities versus proportion of the sample assigned to the group – `propAssign(...)`
- Confidence interval – `ConfIntT(...)`
- Summary – `adequacy(...)`

	1	2	3	4
Prob. est.	0.10227620	0.3118053	0.2247159	0.3612026
CI inf.	0.08165381	0.2720132	0.1916078	0.3161885
CI sup.	0.12689314	0.3542225	0.2625223	0.4086869
Prop.	0.09800000	0.3180000	0.2220000	0.3620000
AvePP	0.89837432	0.9395957	0.9843333	0.9973927
OCC	77.59454723	34.3317995	216.7625509	676.5346954

trajeR - Model selection

- AIC – `trajeRAIC(...)`
- BIC – `trajeRBIC(...)`
- Slope Heuristics – `trajeRSH(...)`

Outline

The Beta distribution

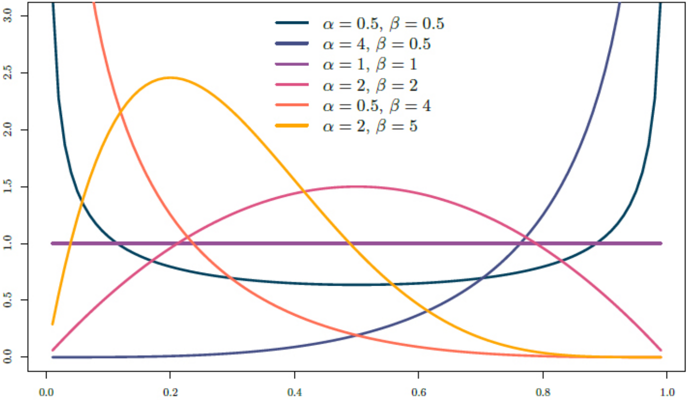


Figure 1 – *Example of different shapes of the Beta density for some parameters.*

Underlying Beta distribution

Density of y_{it} conditional to membership in group C_k :

$$g_k(y_{it};\mu_{kit},\phi_{kit}) = \frac{\Gamma(\phi_{kit})}{\Gamma(\mu_{kit}\phi_{kit})\Gamma((1-\mu_{kit})\phi_{kit})} y_{it}^{\mu_{kit}\phi_{kit}-1}(1-y_{it})^{(1-\mu_{kit})\phi_{kit}-1},$$

with

$$\mu_{kit} = \frac{e^{\beta_k A_{it}+\delta_k W_{it}}}{1+e^{\beta_k A_{it}+\delta_k W_{it}}} \text{ and } \phi_{kit} = \zeta_k A_{it}. \quad (8)$$

Likelihood of the data:

$$L = e^{\prod_{i=1}^{n}\left(\sum_{k=1}^{K}\pi_k \prod_{t=1}^{T} \frac{\Gamma(\phi_{kit})}{\Gamma(\mu_{kit}\phi_{kit})\Gamma((1-\mu_{kit})\phi_{kit})} y_{it}^{\mu_{kit}\phi_{kit}-1}(1-y_{it})^{(1-\mu_{kit})\phi_{kit}-1}\right)}. \quad (9)$$

Data

Data from 190 countries from "Our World In Data".

Main variable of interest: **contamination rate**. We create a panel with monthly data from January 2020 till April 2021.

Covariates: new cases, population size (in million inhabitants), total cases per million people, median age of the population, population density, number of inhabitants over 65 (in million inhabitants), government response stringency index, GDP per capita, extreme poverty index, cardiovascular death rate, diabetes prevalence rate, index of handwashing facilities, rate of hospital beds per thousand inhabitants, life expectancy, index of human development and stringency index.

The nine metrics used to calculate the **stringency index** are: school closures; workplace closures; cancellation of public events; restrictions on public gatherings; closures of public transport; stay-at-home requirements; public information campaigns; restrictions on internal movements; and international travel controls.

Individual trajectories

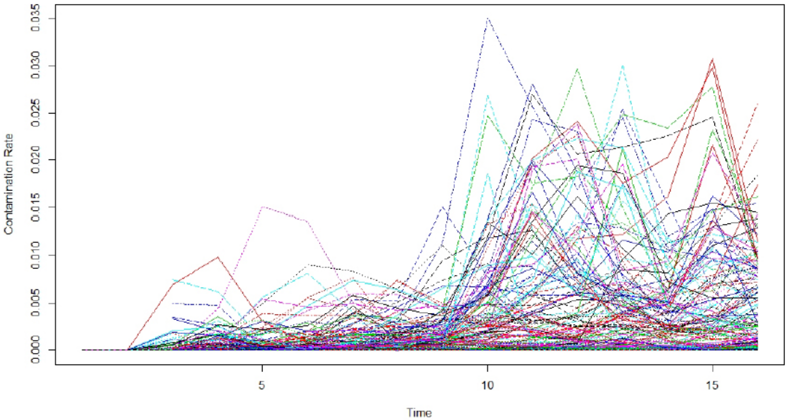


Figure 2 – *Contamination rates for all countries.*

Model selection

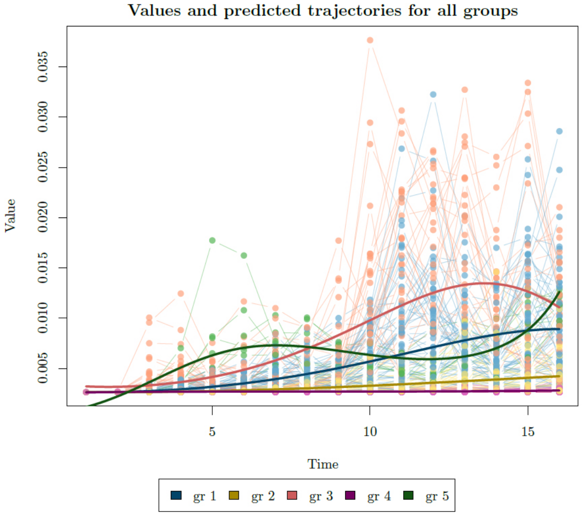
Kass and Wasserman's crierion: Let p_k be the probability that a model with k groups is the correct model. They show that p_k can be approximated by

$$p_k \approx \frac{e^{BIC_k - BIC_{max}}}{\sum_k e^{BIC_k - BIC_{max}}}.$$

Number of groups	AIC	BIC	Prob
2	29851.99	14902.64	0.00000
3	30341.00	15142.28	0.00000
3	29945.96	14936.64	0.00000
3	30777.14	15352.23	0.00000
4	30839.69	15370.52	0.00000
4	31192.78	15547.06	0.00001
5	31241.46	15558.41	0.99999

Table 1 – *Model selection criteria*

Typical trajectories



World Map with the five clusters

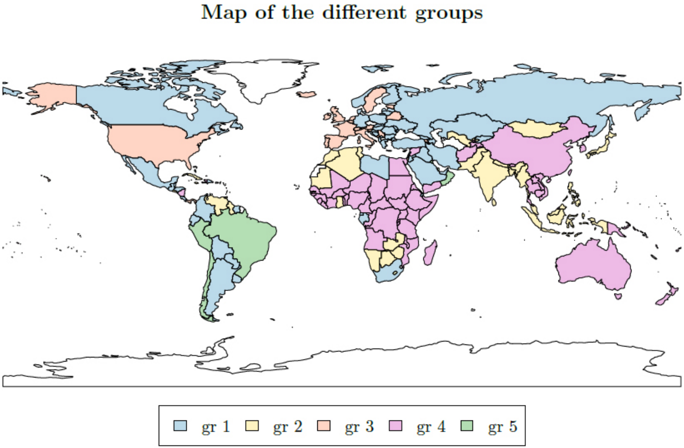


Figure 5 – *World map with the geographic distribution of the five groups*

Predictors of group membership

	Group 1			Group 2		
	Estimate	Std. Error	Prob>|T|	Estimate	Std. Error	Prob>|T|
intercept	-16.812	4.681	0	-4.805	3.422	0.16
median age	0.193	0.086	**0.024**	0.172	0.101	**0.088**
population density	-0.003	0.002	0.093	0.000	0.001	0.869
aged 65 older	-0.021	0.132	0.871	-0.060	0.126	0.631
life expectancy	0.073	0.080	0.364	-0.073	0.071	0.304
mean of stringency	0.112	0.023	**0**	0.092	0.023	**0**
	Group 3			**Group 5**		
	Estimate	Std. Error	Prob>|T|	Estimate	Std. Error	Prob>|T|
intercept	-67.733	19.400	0	-73.689	23.469	0.002
median age	0.129	0.158	0.412	0.418	0.205	**0.041**
population density	0.000	0.001	0.784	0.000	0.001	0.926
aged 65 older	0.109	0.178	0.542	-0.640	0.206	**0.002**
life expectancy	0.646	0.223	**0.004**	0.646	0.283	**0.023**
mean of stringency	0.185	0.054	**0.001**	0.228	0.075	**0.002**

Table 4 – *Predictors of group membership.*

Stringency index as time dependent covariate

Param.	sd	Test
Beta 1		
-5.843	0.026	0.000
-0.120	0.024	0.000
0.029	0.004	0.000
-0.001	0.000	0.000
Beta 2		
-5.927	0.003	0.000
-0.014	0.004	0.000
0.005	0.001	0.000
0.000	0.000	0.001
Beta 3		
-5.602	0.117	0.000
-0.421	0.070	0.000
0.076	0.009	0.000
-0.003	0.000	0.000
Beta 4		
-5.972	0.012	0.000
0.012	0.005	0.018
-0.001	0.001	0.043
0.000	0.000	0.027
Beta 5		
-7.304	0.366	0.000
0.701	0.147	0.000
-0.078	0.017	0.000
0.003	0.001	0.000

Param.	sd	Test
Phi 1		
14.337	0.317	0.000
-1.164	0.076	0.000
0.040	0.004	0.000
Phi 2		
19.866	0.570	0.000
-1.710	0.125	0.000
0.061	0.006	0.000
Phi 3		
9.624	0.369	0.000
-0.521	0.097	0.000
0.016	0.005	0.003
Phi 4		
12.887	0.372	0.000
0.148	0.085	0.082
-0.015	0.004	0.000
Phi 5		
7.384	0.137	0.000

Param.	sd	Test
Delta 1		
0.001	0.000	0.001
Delta 2		
0.000	0.000	0.955
Delta 3		
0.010	0.001	0.000
Delta 4		
0.000	0.000	0.000
Delta 5		
0.004	0.001	0.004

Param.	sd	Test
Prob. 1		
0.328	0.039	0.00
Prob. 2		
0.175	0.030	0.00
Prob. 3		
0.156	0.030	0.00
Prob. 4		
0.301	0.035	0.00
Prob. 5		
0.040	0.016	0.01

Table 5 – *parameters of the final model with time dependent covariates.*

Outline

Basic Idea

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

The 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,

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

where $\pi_{k_j | k_1 \dots k_{j-1}}$ 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.

A new approach

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

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

A simulation example

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

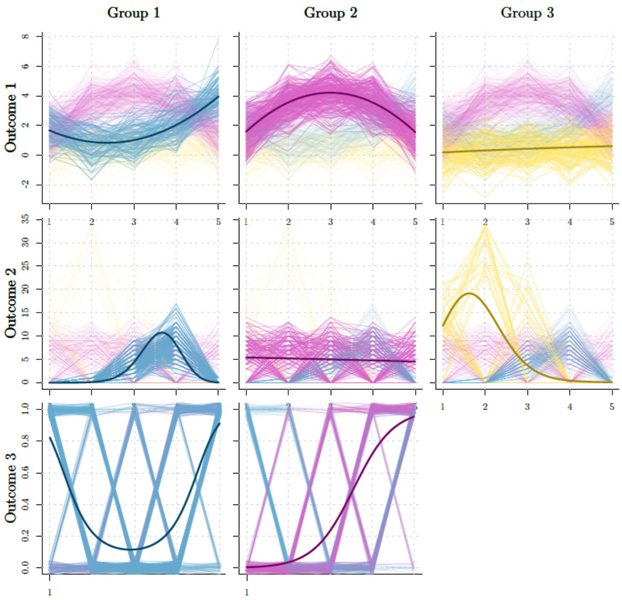
	Variable 1		
	Th.	EM	L
β_{11}	3.53	3.37162	3.37154
β_{12}	-2.25	-2.13530	-2.13523
β_{13}	0.47	0.45090	0.45089
β_{21}	-1.620	-1.71513	-1.71513
β_{22}	3.900	3.96422	3.96422
β_{23}	-0.650	-0.66183	-0.66183
β_{31}	0.263	0.04619	0.04624
β_{32}	0.036	0.15730	0.15725
β_{33}	0.010	-0.00882	-0.00881
σ	1	2.69943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



A real data example: 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

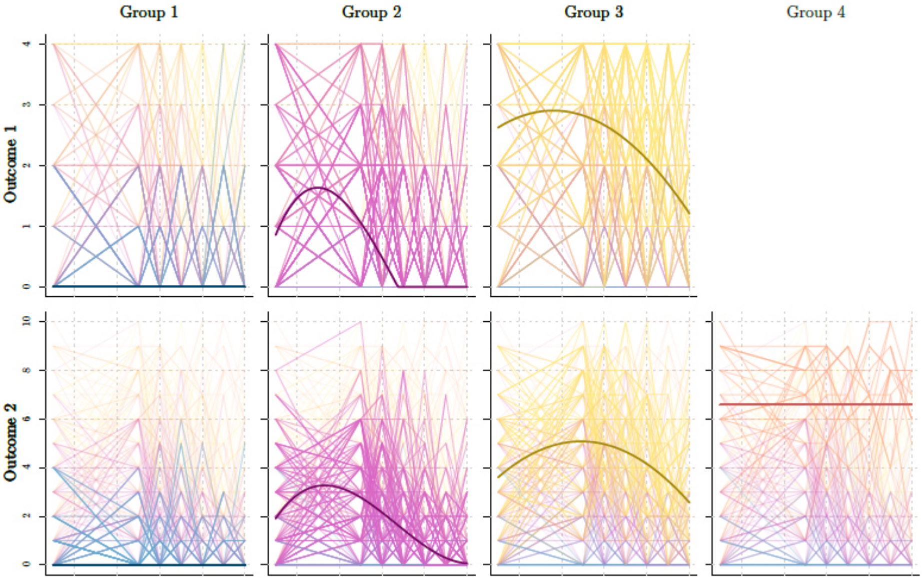
	Variable 1	
	TrajeR	traj (SAS)
β_{11}	-2.55343	-2.55348
β_{21}	0.48071	0.48074
β_{22}	-6.24514	-6.24425
β_{23}	-3.53665	-3.53676
β_{24}	14.90876	14.90347
β_{31}	2.66416	2.66411
β_{32}	-1.98842	-1.98840
β_{33}	-4.14192	-4.14164
σ	2.31949	2.3195

	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



Outline

Identifiability

Definition:

A finite mixture model is identifiable if a given data set leads to a uniquely determined set of model parameter estimations up to a permutation of the clusters.

Identifiability of the parameters is a necessary condition for the existence of consistent estimators for any statistical model.

Without identifiability, there might be several solutions for the parameter estimation problem.

Notations

Distribution f of a finite mixture model:

$$f(y_i;\Omega) = \sum_{k=1}^{K} \pi_k g_k(y_i;\theta^k).$$

Cumulative distribution function F of a finite mixture model:

$$F(y_i;\Omega) = \sum_{k=1}^{K} \pi_k G_k(y_i;\theta^k).$$

Mixtures and mixing distributions

Let $\mathcal{F} = \left\{F(y;\omega),\ y \in \mathbb{R}^T,\ \omega \in \mathbb{R}_K^{s+2}\right\}$ be a family of T-dimensional cdf's indexed by a parameter set ω, such that $F(y;\omega)$ is measurable in $\mathbb{R}^T \times \mathbb{R}_K^{s+2}$.

The the $s+2$-dimensional cdf $H(x) = \int_{\mathbb{R}_K^{s+2}} F(y;\omega)dG(\omega)$ is the image of the above mapping, of the $s+2$-dimensional cdf G.

The distribution H is called the mixture of $\mathcal{F}$ and G its mixing distribution.

Let $\mathcal{G}$ denote the class of all $s+2$-dimensional cdf's G and $\mathcal{H}$ the induced class of mixtures H.

Then $\mathcal{H}$ is identifiable if Q is a one-to-one map from $\mathcal{G}$ onto $\mathcal{H}$.

Characterization of identifiability

The set $\mathcal{H}$ of all finite mixtures of class $\mathcal{F}$ of distributions is the convex hull of $\mathcal{F}$.

$$\mathcal{H} = \left\{ H(y) : H(y) = \sum_i c_i F(y, \omega_i),\ c_i > 0, \sum_i c_i = 1,\ F(y, \omega_i) \in \mathcal{F} \right\}. \tag{10}$$

Theorem

A necessary and sufficient condition for the class $\mathcal{H}$ of all finite mixtures of the family $\mathcal{F}$ to be identifiable is that $\mathcal{F}$ is a linearly independent family over the field of real numbers.

The Model

$$Y_{it} = f(a_{it}; \beta^k, \delta^k) + \varepsilon_{it}^k = \beta^k A_{it} + \delta^k W_{it} + \varepsilon_{it}^k. \tag{11}$$

We can write

$$\mathcal{L}\left((Y_i)_{i \in I}\right) = \bigotimes_{i \in I} F_{A_i, W_i, J}. \tag{12}$$

Identifiability of a model means that knowing the data distribution $\mathcal{L}(Y_i), i \in I$, one can uniquely identify the mixing distribution J.

That is, no two distinct sets of parameters lead to the same data distribution.

Nagin's base model

$$\mathcal{C}_1 = \left(F_{A,J} \ : \ F_{A,J} = \bigotimes_{i \in I} F_{A_i,J} \right)_{J \in \Omega_1}$$

Theorem

Let $h_j = \min \left\{ q \ : \ \{A_{ij}, i \in I\} \subseteq \cup_{i=1}^{q} H_i \ \ H_i \in \mathcal{H}_{n-1} \right\}$.
If there exist j *such that* $|S(J)| < h_j, \ \forall J$ *then* $\mathcal{C}_1$ *is identifiable.*

Adding covariates independent of cluster membership

$$\mathcal{C}_2 = \left(F_{A,J} \ : \ F_{A,J} = \bigotimes_{i \in I} F_{A_i, W_i, J} \right)_{J \in \Omega_1} , \tag{13}$$

$$\mathcal{C}_{2A} = \left(F_{A,J} \ : \ F_{A,J} = \bigotimes_{i \in I} F_{A_i, J} \right)_{J \in \Omega_1} , \tag{14}$$

$$\mathcal{C}_{2W} = \left(F_{A,J} \ : \ F_{A,J} = \bigotimes_{i \in I} F_{W_i, J} \right)_{J \in \Omega_1} . \tag{15}$$

Theorem

If $\mathcal{C}_{2A}$ and $\mathcal{C}_{2W}$ are identifiable and W_{ij} is not a multiple of A_{ij}, for all i, j, then $\mathcal{C}_2$ is identifiable.

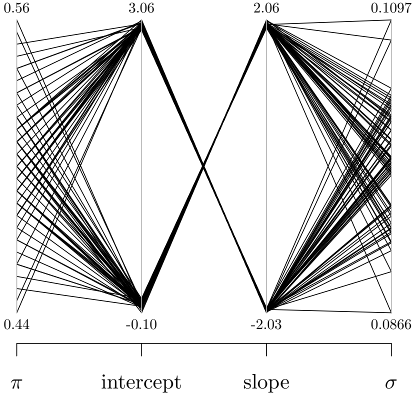
Numerical Example

- Two clusters with sizes $\pi_1 = \pi_2 = \frac{1}{2}$.
- Two time-points 1 and 2.
- Same variability in both clusters $\sigma = 0.1$

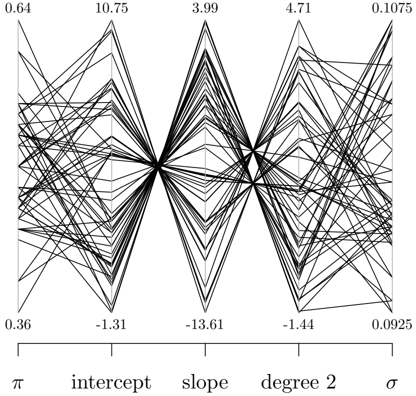
We simulate 50 samples of 100 trajectories with parameters

- $\beta^1 = (3, -2)$ and $\beta^2 = (0, 2)$ (linear model)
- $\beta^1 = (10, -12.5, 3.5)$ and $\beta^2 = (-2, 5, -1)$ (polynomial model).

Parallel coordinate plots of the estimated parameter



Linear Model



Parabolic Model

The generalized model

Theorem

The model is identifiable if

- $d_k < T$ *for all* $1 \leq k \leq K$ *and all* a_{it} *are distinct, for all* i_t.
- W_k *has full rank for all* $1 \leq k \leq K$.
- $rk(A_k, W_k) = rk(A_k) + rk(W_k)$, *for all* $1 \leq k \leq K$.

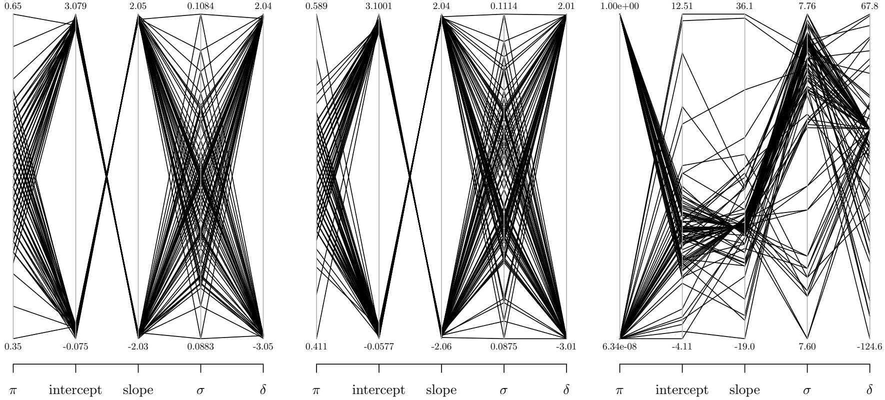
Numerical Example

- Two clusters with sizes $\pi_1 = \pi_2 = \frac{1}{2}$.
- Two time-points 1 and 2.
- Same variability in both clusters $\sigma = 0.1$
- Shape description parameters $\beta_1 = (3, -2)$, $\beta_2 = (0, 2)$, $\delta_1 = 2$ and $\delta_2 = -3$.

We simulate 50 samples of 100 trajectories for 3 types of models:

- The covariate is independent of time and only takes values 0 or 1
- The covariate is time dependent but in a nonlinear way
- The covariate is time dependent in a linear way

Parallel coordinate plots of the estimated parameter



Model 1 Model 2 Model 3

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.
- Noel, C & Schiltz, J. 2022: TrajeR - an R package for finite mixture models. SSRN paper 4054519.
- Noel, C & Schiltz, J. 2023: Finite Mixture Models for an underlying Beta distribution with an application to COVID-19 data.
- Noel, C & Schiltz, J. 2024: Multitrajectory Analysis in Finite Mixture Models.
- Noel, C & Schiltz, J. 2024: Identifiability for Finite Mixture Models.