A new R package for Finite Mixture Models with an application to pension systems

Jang SCHILTZ (University of Luxembourg)

joint work with

Jean-Daniel GUIGOU (University of Luxembourg),
Bruno LOVAT (University of Lorraine)
and
Cédric NOEL(University of Lorraine & University of Luxembourg)

MAF 2022
April 20, 2022

Outline

Outline

1. Finite Mixture Models
2. The R package trajeR

Outline

1. Finite Mixture Models
2. The R package trajeR
3. Application to the Luxembourg pension system

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)

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

Aim of the Analysis

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:

$$y_{i_t} = \beta_0^k + \beta_1^k t + \beta_2^k t^2 + \beta_3^k t^3 + \beta_4^k t^4 + \varepsilon_{i_t}^k, \tag{1}$$

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

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

Predictors of trajectory group membership

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

Multinomial logit model:

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

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

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

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

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

Outline

1. Finite Mixture Models
2. The R package trajeR
3. Application to the Luxembourg pension system

Execution example

```
R> # Likelihood different sigma
R> solL = trajeR(Y = data[,2:11], A = data[,12:21],
+                degre = c(0,3,4),
+                Model = "CNORM", Method = "L",
+                hessian = TRUE, ssigma = FALSE)
```

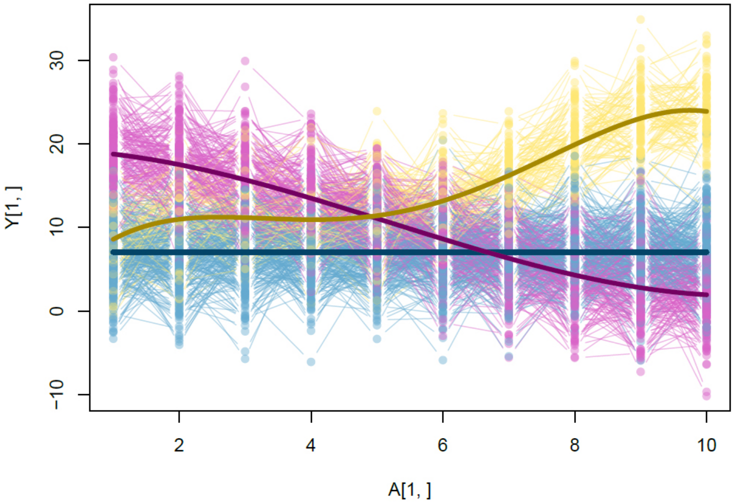
Output of result

group	Parameter	Estimate	Std. Error	T for H0: param.=0	Prob>|T|
1	Intercept	7.0494	0.08442	83.50741	0
2	Intercept	19.30454	0.6537	29.53102	0
	Linear	-0.09315	0.48451	-0.19225	0.84755
	Quadratic	-0.45614	0.09932	-4.5927	0
	Cubic	0.02919	0.00593	4.92296	0
3	Intercept	1.6695	1.53075	1.09064	0.27548
	Linear	10.11827	1.73041	5.84733	0
	Quadratic	-3.70726	0.59886	-6.19052	0
	Cubic	0.53764	0.07968	6.74723	0
	Quartic	-0.02459	0.00358	-6.85989	0
1	sigma1	3.95795	0.05912	66.94911	0
2	sigma2	4.11085	0.07232	56.84354	0
3	sigma3	4.00173	0.10076	39.71375	0
1	pi1	0.45891	0.02837	0	0
2	pi2	0.34901	0.0219	-12.49729	0
3	pi3	0.19208	0.01802	-48.32612	0

Likelihood : -14564.35

Graphical illustration of result

plotrajeR(solEM, Y = data[,2:11], A = data[,12:21], col = vcol)



Outline

The Luxembourg pension system

- Pay-as-you-go system + creation of a reserve (1.5 times the amount of the annual expenses).
- Very high replacement rate (over 90 %).

- Due too several reasons (longevity risk and labor market explosion in the 1990s) the system is not sustainable at all!
- Reform possibilities :
  - Parameter adjustement in the Pay-as-you-go system
  - et/ou Developp complementary systems (mix of funded and unfunded system)

The data

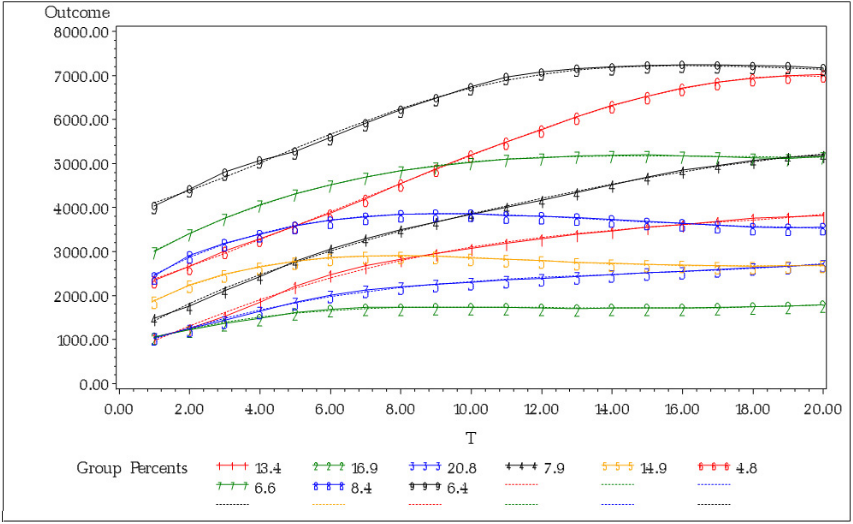
Salaries of workers in the private sector in Luxembourg from 1940 to 2006.

About 7 million salary lines corresponding to 718 054 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship (luxemburgish residents, foreign residents, foreign non residents)
- working status (white collar worker, blue collar worker)
- year of birth
- age in the first year of professional activity

Result of finite mixture model



Sustainability coefficient of the PAYG system

τ_1 = sum of all salaries earned by active workers / sum of all pensions paid to retirees at time t

$$\tau_1 = \frac{S_0 + ... + \frac{S_T}{(1+d)^T}}{\frac{k}{(1+d)^{T+1}} P_{T+1} + ... + \frac{k}{(1+d)^{T+T^*}} P_{T+T^*}}.$$

Sustainability coefficient of the funded system

$\tau_2 =$ total sum earned by the individual during his period of activity / sum of all the pensions that are paid to him thanks to the savings that he has accumulated

$$\tau_2 = \frac{S_j}{a_j(i-\lambda_j)} i \frac{(1+i)^T - (1+\lambda_j)^T}{(1+i)^T - 1}.$$

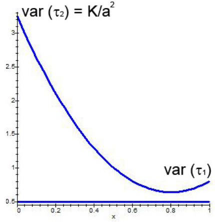
Global sustainability coefficient

$$\tau = x\tau_1 + (1 - x)\tau_2$$

is the number of euros necessary to pay 1 euro for the pension.

Here x euros come from the PAYG system and $1 - x$ euros from capitalisation.

We want to limit the risk of the hybrid system without reducing the pension and in the same time minimise the capitalisation effort.



Gain of sustainability and optimal saving amount

$$G(x) = \frac{var(\tau_1) - var[\tau(x)]}{var(\tau_1)}$$

measures the gain of sustainability of the mixed system with respect of the PAYG system.

We suppose that the utility function $U = U(a)$ of an active worker is decreasing in a.

Gain of sustainability and optimal saving amount

Theorem. The value $x = x^*$ for which the utility function U is maximal under the sustainability constraint

$$G(x) \leq G^*$$

is given by $x^* = 1 - G^*$.

Moreover the individual needs a constant annual saving amount

$$a^* = \sqrt{\frac{G^* K}{var(\tau 1)(1 - G^*)}},$$

where $K = Var[\frac{S_j}{a_j(i-\lambda_j)} i \frac{(1+i)^T - (1+\lambda_j)^T}{(1+i)^T - 1}]$ depends on the salary trajectory.

Example

An individual worker wants to divide by 2 the variability of his PAYG sustainability constraint needs to save annually at least the following amount (depending on his salary evolution subgroup):

Groupe	G1	G2	G3	G4	G5	G6	G7	G8	G9
Annuité	4466 €	713 €	1448 €	5231 €	220 €	6364 €	2809 €	743 €	3140 €

Papers presented

- 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. p. 107-126.
- Noel, C & Schiltz, J. 2022: TrajeR - an R package for finite mixture models. SSRN paper 4054519.
- Guigou, J.D., Lovat, B. & Schiltz, J. 2012: Optimal mix of funded and unfunded pension systems: the case of Luxembourg. Pensions 17-4, p. 208-222.