A latent class growth curve model for walking behaviour in an indoor mobility test

Leonardo Grilli and Carla Rampichini
University of Florence

Summary

Purpose
This work is motivated by a study on aging based on a representative population living in the Chianti geographic area (Tuscany, Italy). Multiple factors may influence the ability to walk and no standard criteria are currently available to establish whether these factors are functioning within the "normal" range.

Data
We consider the *InCHIANTI* study, a representative population-based study of the elderly living in the Chianti area (Tuscany, IT): http://inchiantistudy.net. Our work exploits data collected during the performance of an indoor mobility test to discriminate individuals at high-risk of mobility disability or falls. A large number of outcomes was collected during the test. Many individual characteristics were collected before the walk test: every subject answered a questionnaire and performed a medical examination and a session of mobility and mental performance.

Method
We specify a latent class growth curve model to detect walking impairment. The model explicitly considers non-ignorable missing data due to individuals not completing the walking test. To this end we specify a selection model (Muthén et al, 2011).
The effect of subject pre-test characteristics on the probability of belonging to each latent class is also investigated.

Results
Our findings suggest that subjects aggregate into distinct behavioural profiles, characterized by age and gender. Model results can be used to predict future health problems.

The 400mt walking test

We consider a subsample from the last follow up (2013-2014): 315 subjects aged 35-93 years that performed the 400 meter walk test. This test is well known in literature. It's predictive of cardiovascular, mental, cognitive, musculoskeletal and neurological problems and it recognize psychophysical difficulty that aren't shown by other tests (Chang et al., 2006; Tian et al., 2015; Vestergaard et al., 2009).

The total run of 400 meter was split into 44 straights of 9.09 meter (from now on **straight=lap**). For every completed lap, several variables are measured, including the walk time and the signals from the accelerometer of a smart-phone located next to the last vertebra. Missing values arise from individuals who did not complete the test. The first two and the last two laps are excluded from the analysis, so we consider **40 laps**. The variable of interest in the present work is the **straight walk speed** of each lap, defined as lap time over lap length (9.09 mt). This is a simple measure of walking problems.

Statistical model

We specify a latent class growth curve model with non-ignorable drop-out adapting the proposal of Bartolucci and Murphy (2015).

- i: individual (i = 1, …, 315)
- l: lap, namely a straight line of 9.09 mt (l = 1, …, 40)
- B_{il}: status of individual i at lap l (0 = walking, 1 = drop-out)
- Y_{il}: speed (km/h) of individual i at lap l
- U_i: latent class of individual i $(U_i = u,\ u\epsilon\{1,\dots,k\})$

$$Pr(B_{il} = 1|\, U_i = u) = \frac{\exp(\boldsymbol{x}_l{}'\boldsymbol{\gamma}_u)}{1+\exp(\boldsymbol{x}_l{}'\boldsymbol{\gamma}_u)} \quad (1)$$

$$Y_{il}|\, B_{il} = 0, U_i = u \sim N(\mu_{lu}, \sigma^2) \quad \mu_{lu} = \boldsymbol{x}_l{}'\boldsymbol{\beta}_u \quad (2)$$

$$\log\frac{Pr(U_i = u)}{Pr(U_i = 1)} = \boldsymbol{z}_i{}'\boldsymbol{\delta}_u \ ,\ u = 2,\dots,k \quad (3)$$

- $\boldsymbol{x}_l$ = column vector with the terms of an orthogonal polynomial of order r. Here r=3
- $\boldsymbol{z}_i$ = column vector with the covariates of individual i here including: the constant, age and gender.

Model fitting

ML estimation is performed using the EM algorithm as implemented in R by Bartolucci and Murphy (2015).
We fit a set of models varying the number of latent classes from 1 to 5. We choose the model with the minimum value of the Normalized Entropy Criterion (NEC) of Celeux and Soromenho (1996), i.e. the model with k=4 latent classes, as shown in Table 1.

Table 1. Model selection

k	N par	BIC	NEC
1	8	42850.5	1
2	19	31652.3	0.000097
3	30	25694.9	0.000192
4	41	21767.8	0.000045
5	52	19307.6	0.000152

Main results

Table 2 reports sizes and main characteristics of the k=4 latent classes. The individuals are classified on the basis of the maximum posterior probability.
The latent classes are ordered on the basis of the average predicted speed at the first lap.
There are no classes with extremely small or large size. Almost all dropouts are in latent class 1, where the individuals are older and slower. Latent class 4 is characterized by younger and faster individuals, with a prevalence of males.

Table 2. Model results: subjects classification and class characteristics

Latent class	N of subjects	% drop-out	Age (mean)	% females	1st lap average speed
1	73	39.7	82.7	56.2	3.87
2	91	1.1	78.0	58.2	5.48
3	94	0.0	63.5*	50.0	6.81
4	57	0.0	51.0*	35.1*	8.09
Overall	*315*	*9.5*	*69.9*	*51.1*	*5.98*

* the corresponding coefficient in eq. (3) is statistically significant at 5%

Figure 1. Sample (solid) and fitted (dashed) survival functions of latent class 1

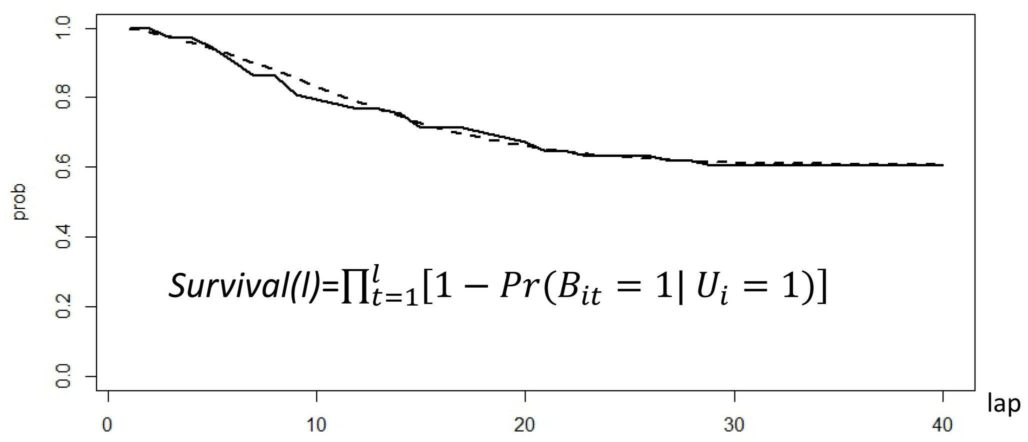
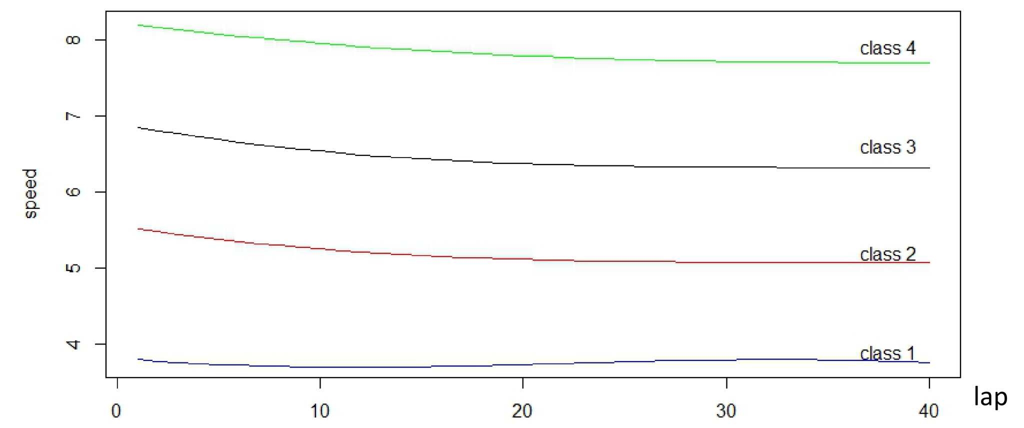


Figure 2. Predicted trajectories $\mu_{lu} = \boldsymbol{x}_l{}'\boldsymbol{\beta}_u$ by latent class



Predicted trajectories

Figure 1 above depicts the survival function for latent class 1: the dropouts (about 40%) are concentrated in the first half of the test.
Figure 2 above represents the fitted trajectories for the four classes: the trajectories are clearly ranked on the basis of the speed and do not cross.
Figures 3 and 4 compare fitted and sample trajectories showing a good model fitting. Except for latent class 1, that is affected by dropouts, the trajectories have a similar decreasing shape, with a stronger decline in the first half of the test.

Figure 3. Average sample (solid) and fitted (dashed) trajectories by latent class

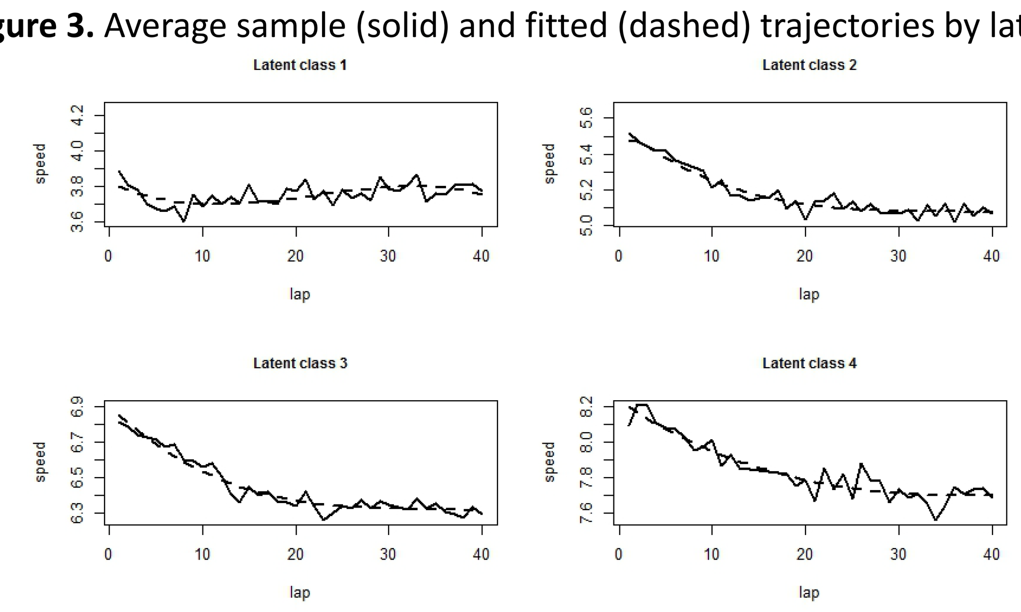
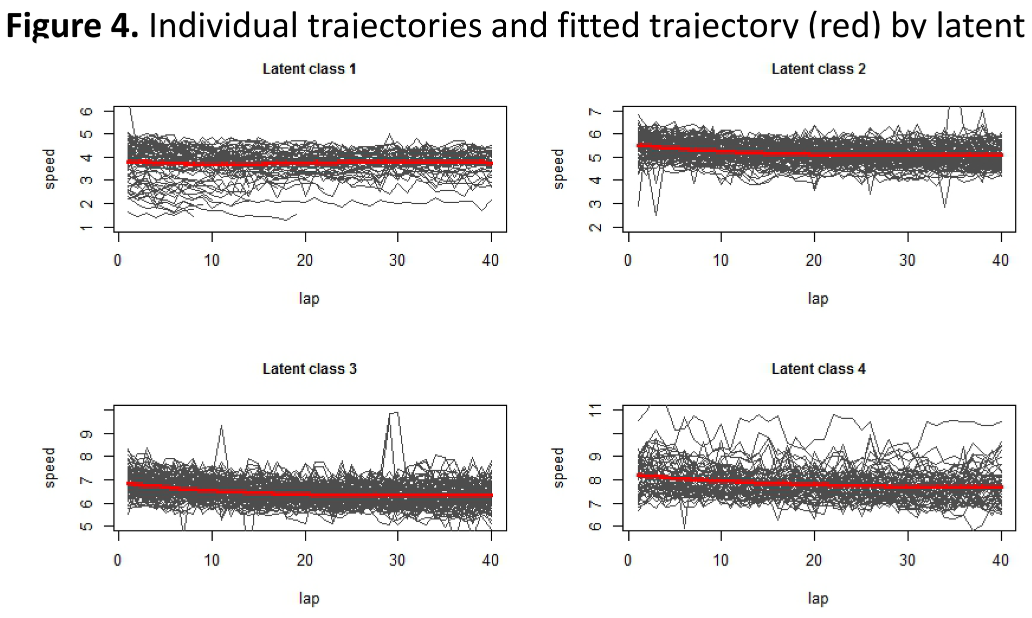


Figure 4. Individual trajectories and fitted trajectory (red) by latent class



Future work

The model can be extended to account for other features, such as heteroscedasticity among individuals and autocorrelation within latent classes. Moreover, it is worth to jointly analyze the measures taken from the accelerometer.

Contact

Leonardo Grilli, Carla Rampichini
Dipartimento di Statistica, Informatica, Applicazioni "G.Parenti"
Viale Morgagni, 59
50134 Firenze (Italy)
Email: grilli@disia.unifi.it, rampichini@disia.unifi.it

References

Bartolucci F. and Murphy T.B. (2015) A finite mixture latent trajectory model for modeling ultrarunners' behavior in a 24-hour race. *Journal of Quantitative Analysis in Sports*,11, 193–203.
Celeux G., Soromenho G. (1996) An Entropy Criterion for Assessing The Number of Clusters in a Mixture Model, *Journal of Classification*, 13, 195–212.
Chang M., Cohen-Mansfield J., Ferrucci L., Leveille S., Volpato S., de Rekeneire N., Guralnick J. M. (2004) Incidence of loss of ability to walk 400 meters in functionally limited older population, *J Am Geriatric Soc.*, 52(12), 2094-2098.
Muthén, B., Asparouhov, T., Hunter, A., & Leuchter, A. (2011). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. Psychological Methods, 16
Tian Q., Resnick S. M., Ferrucci L., Studenski S. A. (2015) Intra-individual lap time variation of the 400-m walk, an early mobility indicator of executive function decline in high-functioning older adults?, *AGE*, 37: 115.
Vestergaard S., Patel K. V., Bandinelli S., Ferrucci L. e Guralnik J. M. (2009) Characteristics of 400-Meter Walk Test Performance and Subsequent Mortality in Older Adults, *Rejuvenation Research*, 12, 177-184 .