A latent mixed-effects model for longitudinal categorical data in Parkinson disease

Pierre-Emmanuel Poulet[1,2], Jean-Christophe Corvol[2], Stanley Durrleman[1,2]

1. Aramis project Team, Inria, 2.Institut du Cerveau et de la Moelle épinière (ICM), Hôpital de la Pitié Salpétrière, Paris, France

https://gitlab.com/icm-institute/aramislab/leaspy

Introduction

We propose an extension of Leaspy [1,2,3] as a **generic method to analyze longitudinal categorical data** :

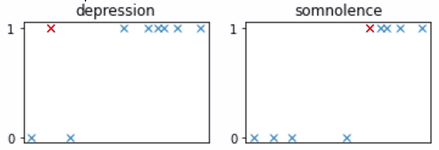
- a **non-linear mixed effect** model describes the **latent evolution** of the disease
- **personalization of individual parameters** allows to partially **predict** future data
- the **multivariate** model jointly estimates observations

We applied the model to the NSPARK database, focusing on the onset of disease-related symptoms for Parkinson disease patients. We show the potential of this method on data as noisy and coarse as presence/absence of symptoms at each visit.

Database

N	Initial age	Diagnosis
2821	66.6 ± 10.7	PD
# reported symptoms	# visits	Follow-up duration
23*	5.0 ± 1.5	2.6 ± 1.3

Example :



Due to the noise in the data and the treatment impact, we decided to look at the first occurrence of each symptom (onset).

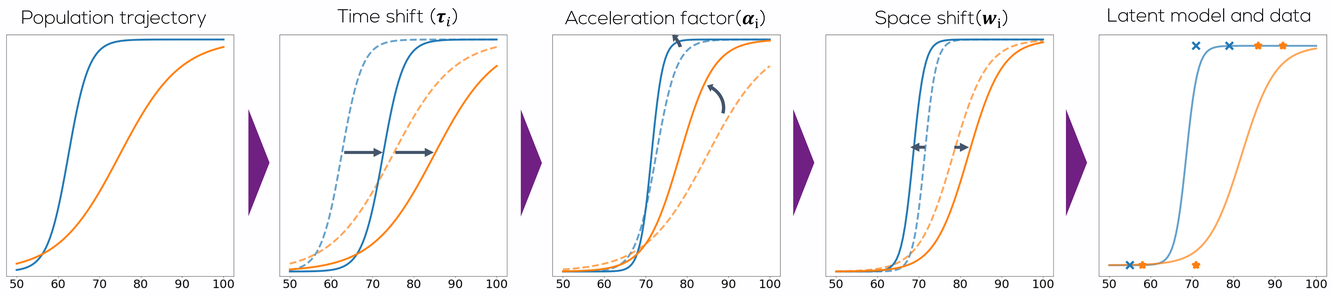
Method

- Observations : $\boldsymbol{y_{ij}} \in \{0,1\}^n$ of patient i at time $\boldsymbol{t_{ij}}$ (array of 0/1 for each of the n symptoms)
- Latent model : we model $\boldsymbol{y_{ij}}$ as a Bernoulli realization with parameter $\boldsymbol{\eta_{ij}} \in \mathcal{M}$ a Riemannian manifold; $\boldsymbol{\eta_{ij}} = \mathbb{P}(\boldsymbol{y_{ij}} = 1)$ is a function of the time describing the evolution of the Bernoulli parameter
- Population average trajectory : $\boldsymbol{\gamma_0(t)}$ is a geodesic in $\mathcal{M}$
- Individual effect :
  - Space-shift parameter $\boldsymbol{w_i}$ describes the shift of the individual trajectory in $\mathcal{M}$: $\boldsymbol{\gamma_i(t)} = Exp_{\boldsymbol{w_i}}(\boldsymbol{\gamma_0}(t))$ [exp-parallelization]
  - Time reparameterization : time-shift $\boldsymbol{\tau_i}$ and acceleration factor $\boldsymbol{\alpha_i}$ for individual disease timeline $\hat{\boldsymbol{t}}_{ij} = \boldsymbol{\alpha_i}(\boldsymbol{t_{ij}} - \boldsymbol{\tau_i})$
- Non-linear mixed effect model : $\boldsymbol{\eta_{ij}} = \boldsymbol{\gamma_i}(\hat{\boldsymbol{t}}_{ij})$
- Estimation of parameters with a Monte Carlo Markov Chain Stochastic Approximation of Expectation-Maximization algorithm (MCMC SAEM)

Comparing to previous Leaspy model [1,2,3], this model seeks to minimize the crossentropy between $\boldsymbol{\eta_{ij}}$ and $\boldsymbol{y_{ij}}$ and not the mean squared error.

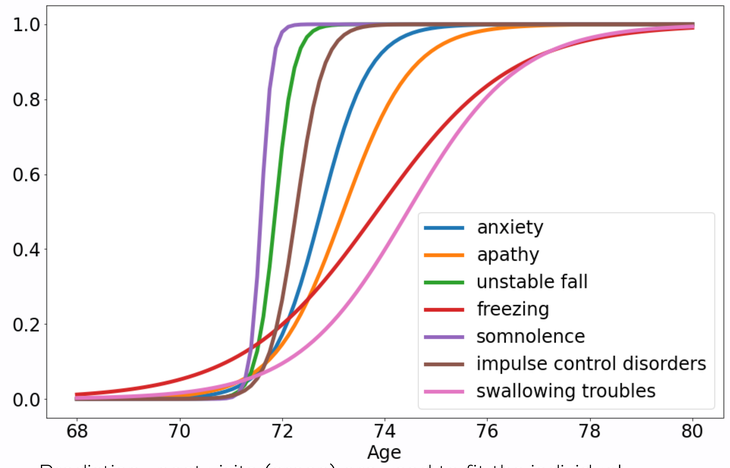
The code will be available in a future release in Leaspy library (see git link).

From population average trajectory to individual latent disease progression with two symptoms (blue and orange)



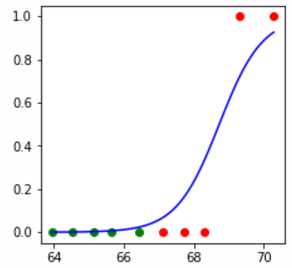
Results

- Population trajectory (figure on the right) shows the order in which symptoms usually appear
- The sharper the slope the more precise the onset of the symptom
- A slower progressing curve means the symptom is less predictable
- Performance is evaluated with ROC AUC for all symptoms : mean AUC is 0.83 on the fitted data
- Personalized trajectory with individual parameters allows to partially predict the onset of symptoms in future visits (not used in model fitting)
- Two groups appear :
  - 11 symptoms with AUC under 0.6, hard to predict
  - 12 symptoms with AUC around 0.65-0.7, showing prediction potential
- The inherent variance linked to disease's symptoms as well as treatment's effect explains the difficulties of the described prediction task

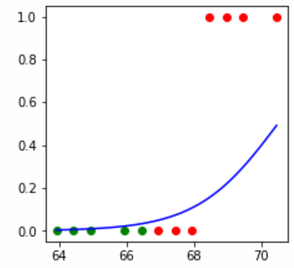


Prediction : past visits (green) are used to fit the individual trajectory (blue) which is then used to predict future visits (red)

Apathy (good predictability)



Freezing (harder to predict)



Conclusions

Our model is an attempt to link fine-grained continuous disease progression models and raw binary data, which is more noisy and provides less information. However we show that it is possible to exploit such data. The mixed-effect model allows to describe an average population trajectory while the individualization of the disease progression can be used to obtain partial prediction power. This can be a powerful tool to leverage categorical markers for disease understanding and eventually prognostic system.

References

[1] Schiratti J-B, Allassonnière S, Colliot O, Durrleman S. Learning spatiotemporal trajectories from manifold-valued longitudinal data. In Advances in Neural Information Processing Systems, pp. 2395– 2403, 2015.
[2] Schiratti J-B, Allassonnière S, Colliot O, Durrleman S. A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. In Journal of Machine Learning Research (JMLR) 18(1), pp. 4840-4872, 2017.
[3] Louis M., Couronné R., Koval I., Charlier B., Durrleman S. Riemannian geometry learning for disease progression modelling. In International Conference on Information Processing in Medical Imaging, pp. 542-553, 2019.