Functional Linear Mixed Models

for Sparsely or Irregularly Sampled Data

Jona Cederbaum[1] and Sonja Greven[1]

[1] *Department of Statistics, LMU Munich, Germany*

`Jona.Cederbaum@stat.uni-muenchen.de`

Abstract

WE propose a non-parametric statistical method to analyze sparsely or irregularly sampled functional data that involve an additional correlation structure. Sources of correlation may be very general, such as repeated measurements, grouping in the data, or crossed designs.

We extend the Functional Linear Mixed Model for longitudinal functional data of Greven, Crainiceanu, Caffo, Reich, Electronic Journal of Statistics, 2010 to more general correlated functional data which are not sampled on a fine grid and for which only a small number of measurements may be available per curve.

Estimation is based on dimension reduction via Functional Principal Component Analysis (FPCA). Our procedure allows the decomposition of the variability of the data as well as the estimation of main effects of interest. We apply our methods in simulations (not presented here) and an application from speech production research.

The Functional Linear Mixed Model

The Functional Linear Mixed Model can be seen as a functional analogue of the scalar Linear Mixed Model (Laird and Ware, Biometrics, 1982) in such a way that random effects are replaced by random processes. The unit of observation is a curve.

The General Functional Linear Mixed Model (GFLMM)

$$Y_i(d) = \mu(\boldsymbol{x}_i, d) + \boldsymbol{z}_i^T B(d) + E_i(d) + \varepsilon_{id}, \quad i = 1, \ldots, n, \tag{1}$$

- $Y_i(\cdot)$: random function observed at arguments d in some set $\mathcal{D}$
- $\mu(\boldsymbol{x}_i, d)$: fixed main effects surface with known covariates $\boldsymbol{x}_i$
- $B(d)$: random functions with known covariates $\boldsymbol{z}_i$
- $E_i(d)$: curve-specific deviations in the form of a smooth residual curves
- ε_{id}: white noise measurement error with variance $\sigma^2(d)$
- n: number of observed curves

Assumption: $B(d)$, $E_i(d)$, and ε_{id} are assumed independent for all i

Estimation

Mean-, auto-covariance-, and eigenfunctions are assumed to be smooth. Dimension reduction is mandatory for estimation of functional data. We use dimension reduction via FPCA whereby the dominant modes of variation are extracted.

Challenges for irregularly or sparsely sampled data

We face some challenges (theoretical and computational) when dealing with irregularly or even sparsely sampled data:

- the PC scores cannot be estimated via numerical integration as usually done in FPCA
- smoothing of single curves may be impossible due to few measurement points
- smoothing in general is less accurate in the sparse case than for dense grid-data
- computational problems arise due to large a number of unique sampling points across curves (no Kronecker products can be used)
- implementation is more challenging with different measurement points

Yao, Müller, and Wang, JASA, 2005 propose a method to perform FPCA for sparse independent functional data. We extend this to the case of correlated functional data.

We propose an **estimation algorithm** consisting of four main steps which is exemplarily described for model (2):

1. Estimation of the fixed main effects function under working independence, i.e.

$$Y_{ijh}(t) = \mu(\boldsymbol{x}_{ij}, t) + \varepsilon_{ijt}$$

- subsequent centering of the data: $\tilde{Y}_{ijh}(t) = Y_{ijh}(t) - \hat{\mu}(\boldsymbol{x}_{ij}, t)$

2. Estimation of the auto-covariance functions and $\sigma^2(t)$ with variance decomposition

$$\begin{aligned} Cov\{\tilde{Y}_{ijh}(t), \tilde{Y}_{i'j'h'}(t')\} &= Cov\{B_i(t), B_i(t')\}\delta_{i,i'} + Cov\{(C_j(t), C_j(t'))\}\delta_{j,j'} \\ &+ \left[Cov\{E_{ijh}(t), E_{ijh}(t')\} + \sigma^2(t)\delta_{t,t'}\right]\delta_{i,i'}\delta_{j,j'}\delta_{h,h'} \end{aligned}$$

with $\delta_{ii'} = \begin{cases} 1, & \text{if } i = i' \\ 0, & \text{otherwise} \end{cases}$

by bivariate penalized splines.

- subsequent evaluation on a fine grid for eigen decomposition
- strength is borrowed across curves (particularly important in sparse setting)

Data Application

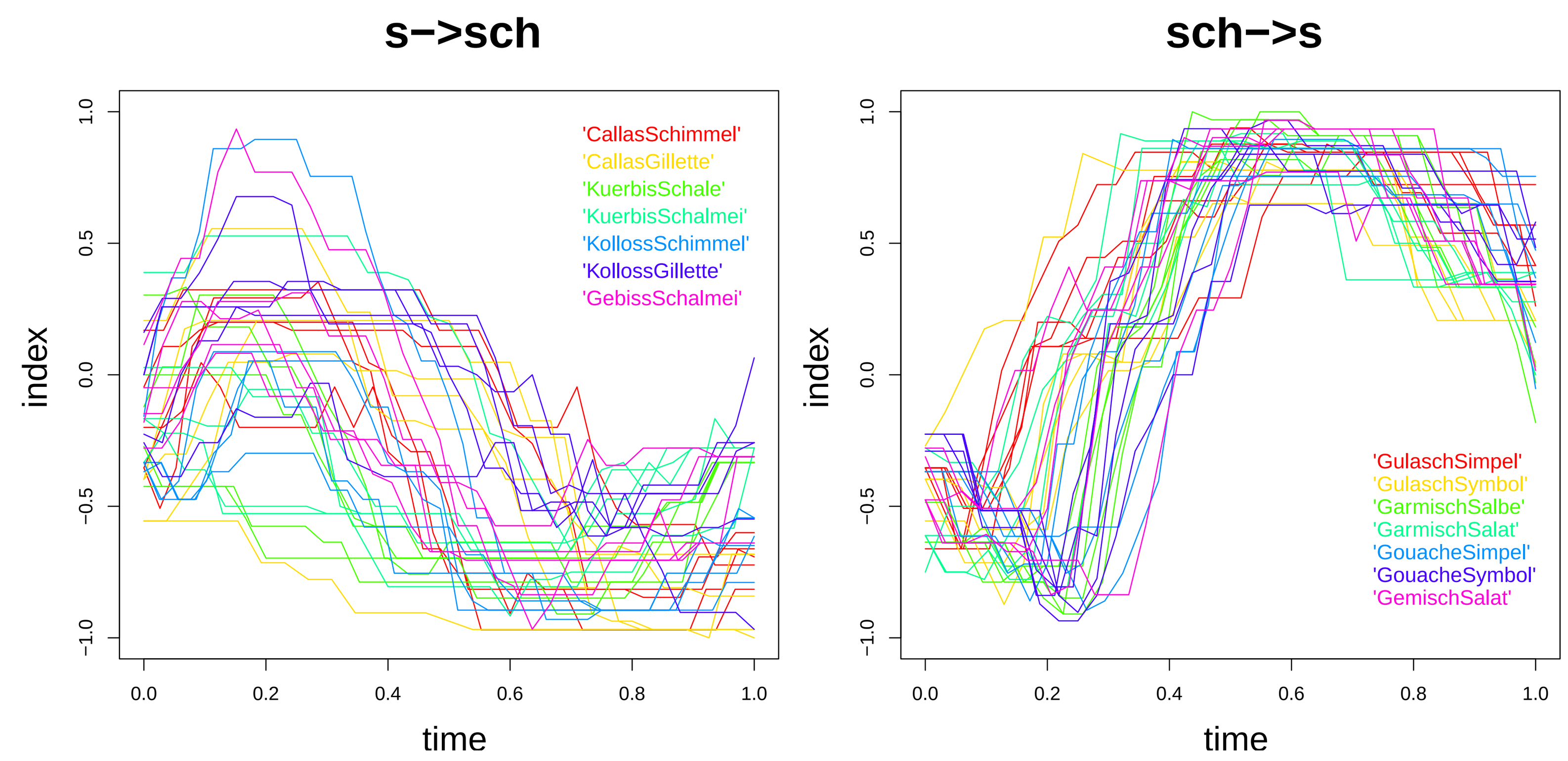

Figure 1: *Index development for one subject. Index values vary between and 1 and -1, where index value 1 stands for the sound "s" and value -1 for the sound "sch". Each word combination is read out five times. The curves belonging to one word combination are the same color.*

Speech production researchers are interested in the change of articulation when certain consonants follow each other.

- 140 different word combinations are read out loud by 9 subjects while their tongue movement is summarized in an one-dimensional index over time (Y)
- standardization of the different reading durations results in irregularly spaced measurements of the index between curves
- each word combination is read out up to five times by each subject
  → correlated measurements both for each word combination and for each subject

We propose to fit a Functional Linear Mixed Model with crossed design. This is a special case of model (1).

$$Y_{ijh}(t) = \mu(\boldsymbol{x}_{ij}, t) + B_i(t) + C_j(t) + E_{ijh}(t) + \varepsilon_{ijht}, \tag{2}$$

- $Y_{ijh}(t)$: index over time for subject i, word combination j and repetition h
- $\mu(\boldsymbol{x}_{ij}, t)$: main fixed effect with known covariates, e.g.
  - order of consonants
  - syllables stressed or not
  - which vowels enclose the sounds
- $B_i(t)$ and $C_j(t)$: random functional intercepts for subjects and word combinations
- $E_{ijh}(t)$: word-, speaker-, and repetition-specific smooth random deviation
- ε_{ijht}: white noise measurement error

3. Expansions of $B_i(t)$, $C_j(t)$, and $E_{ijh}(t)$ in truncated bases of eigenfunctions of the auto-covariance functions which can be estimated from the data

▷ results in Linear Mixed Model

$$Y_{ijh}(t) = \hat{\mu}(\boldsymbol{x}_{ij}, t) + \underbrace{\sum_{k=1}^{N_B} \xi_{ik}^B \phi_k^B(t)}_{\hat{B}_i(t)} + \underbrace{\sum_{k=1}^{N_C} \xi_{jk}^C \phi_k^C(t)}_{\hat{C}_j(t)} + \underbrace{\sum_{k=1}^{N_E} \xi_{ijhk}^E \phi_k^E(t)}_{\hat{E}_{ijh}(t)} + \varepsilon_{ijht}$$

- ξ_{ik}^B, ξ_{jk}^C, and ξ_{ijhk}^E: uncorrelated random variables with zero mean and variances corresponding to the ordered eigenvalues in the decomposition
- $\phi_k^B(t)$, $\phi_k^C(t)$, and $\phi_k^E(t)$: corresponding eigenfunctions

4. Estimation of the PC scores as BLUPs of the Linear Mixed Model

▷ Linear Mixed Model does not need to be fitted

▷ computational highly efficient