Seminar on Functional Data Analysis
A short summary

Hansruedi Künsch, ETH Zurich

Spring Semester 2014

FDA as infinite dimensional statistics

- ▶ Functional data analysis is multivariate statistics in infinite dimensions.
- ▶ Observations are always finite dimensional, so we need to convert them to functions.
- ▶ If observation points are dense, the relevant dimension may be much smaller than number of observations.
- ▶ If observation points are sparse and different between subjects, need functions for comparison, alignment or use as explanatory variables in a regression model.
- ▶ Principal components analysis shows the directions where most variability in a sample of functions occurs.
- ▶ Kernel estimators for regression function and densities exist also for functional data. Semimetrics are a tool to avoid the curse of (infinite) dimensionality.

Regularization

- Underlying principle for FDA is regularization based on basis expansion and smoothness assumptions.
- Regularization by penalization is preferred over truncation in a basis expansion.
- Lack of smoothness of a function x is usually quantified by

$$\int (Lx(t))^2 dt$$

  where L is a differential operator. Standard choices are $Lx(t) = D^2 x(t)$ for splines and $Lx(t) = (D + (\frac{2\pi}{T})^2 D^3)x(t)$ for T-periodic functions.
- To estimate m-th derivatives, the penalty should involve derivatives of order $m + 2$.
- (Generalized) crossvalidation is the preferred method to choose the amount of regularization.

Statistical methods generalized to FDA

- To generalize a standard statistical method to functional data, turn subscripts j, k into function arguments s, t, replace sums by integrals and add a penalty term.

- Example: Linear regression with scalar response

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i \rightarrow Y_i = \beta_0 + \int \beta(t) x_i(t) dt + \varepsilon_i.$$

Fitting by penalized least squares

$$\arg\min \left(\sum_i (Y_i - \beta_0 - \int \beta(t) x_i(t) dt)^2 + \lambda \int (L\beta(t))^2 dt \right).$$

- Basis expansions of β and x_i allow to compute the integrals and lead to linear equations for the unknown coefficients.

- Extensions to linear regression with functional response exist.

Distinctive features of FDA

- Functional data analysis is multivariate statistics with variables ordered in time or space.
- Important information is contained in derivatives of curves.
- Principal differential analysis allows to study linear relations between functions and their derivatives.
- Registration (alignment) of curves is a tool to study variation between subjects other than shifts and amplitude variation.

General remarks

- ▶ The main goal of seminars is not to learn a new topic, but to learn how to read a book chapter or a scientific paper and how to present the material in an understandable way.
- ▶ The book by Ramsay and Silverman has its emphasis on intuitive introduction of concepts and practical advice. At some places I would prefer more clarity and precision, using mathematical language.
- ▶ Talks in last two weeks gave a flavor of asymptotic results based on limit theorems.