Dynamic functional principal components

Siegfried Hörmann
Mathematics Department
ULB, Brussels

Data in many fields of science are sampled from processes that can most naturally be described as functional. Examples include growth curves, temperature curves, curves of financial transaction data and patterns of pollution data. Functional data analysis (FDA) is concerned with the statistical analysis of such data. Since these are intrinsically infinite dimensional objects, tools for dimension reduction are desirable. The functional principal analysis (FPCA) takes here a leading role. It is a key tool in many important empirical and theoretical problems.

A problem with classical FPCA is that it operates in a static way and doesn't take into account any possible serial dependence of the functional observations. Such dependence occurs quite frequently, e.g. if the data consist of a continous time process which has been cut into segments (e.g. days). Though cross-sectionally uncorrelated for a fixed observation, the classical FPC-score vectors have non-diagonal cross-correlations. This means that we cannot analyze them componentwise (like in the i.i.d. case), but need to consider them as vector time series which are less easy to handle and interpret. In particular, a functional principal component with small eigenvalue, hence negligible instantaneous impact on some observation, may have a major impact on the lagged values. Classical static FPCs, thus, in a time series context, will not lead to an adequate dimension reduction technique, as they do in the i.i.d. case. This motivates the development of *dynamic functional principal components*. The idea is to transform the (possibly infinite dimensional) functional time series, into a vector time series (of low dimension 3 or 4, say), where the individual component processes are mutually uncorrelated, and explain a bigger part of the dynamics and variability of the original process.

In this talk will propose such a dynamic version of FPCA for general data structures (Hilbertian data) and study its properties. An empirical analysis and a real data example will be given.

This talk is based on joint work with Łukasz Kidziński and Marc Hallin.