Description length and dimensionality reduction in functional data analysis

Poskitt, D.S.[a] and Sengarapillai, A.[ab]

[a] Department of Econometrics and Business Statistics, Monash University, Clayton, VIC 3800, Australia
[b] Department of Mathematics and Statistics, University of Jaffna, Sri Lanka

Abstract

The use of description length principles to select an appropriate number of basis functions for functional data is investigated. A flexible definition of the dimension of a random function that is constructed directly from the Karhunen-Loève expansion of the observed process or data generating mechanism is provided. The results obtained show that although the classical, principle component variance decomposition technique will behave in a coherent manner, in general, the dimension chosen by this technique will not be consistent in the conventional sense. Two description length criteria are described. Both of these criteria are proved to be consistent and it is shown that in low noise settings they will identify the true finite dimension of a signal that is embedded in noise. Two examples, one from mass spectroscopy and the other from climatology, are used to illustrate the basic ideas. The application of different forms of the bootstrap for functional data is also explored and used to demonstrate the workings of the theoretical results.

Author keywords

Bootstrap; Consistency; Dimension determination; Karhunen-Loève expansion; Signal-to-noise ratio; Variance decomposition

Indexed keywords

Basis functions; Bootstrap; Consistency; Conventional sense; Dimension determination; Dimensionality reduction; Finite dimensions; Flexible definition; Functional data analysis; Functional datas; Generating mechanism; Low noise; Mass spectroscopy; Principle component; Random functions; Theoretical result; Variance decomposition

Engineering controlled terms: Climatology

Engineering main heading: Mass spectrometry