Principal Component Analysis of Dependent Data

Paulo Canas Rodrigues, paulocanas@fct.unl.pt

Department of Mathematics and CMA. Faculty of Science and Technology, Nova University of Lisbon, Monte de Caparica, 2829-516 Caparica, Portugal.

Keywords: Principal Component Analysis. Time Series. SSA. MSSA.

AMS: 62H25, 62M15

Abstract

In many of the statistical methods, and in particular in Principal Component Analysis (PCA), it is assumed that all n observations $x_1, x_2, \ldots, x_n$ are independent. The aim of this work is to analyze the consequences of the application of PCA to observations that do not verify this hypothesis.

Skinner (1986) and Konishi (1992), among others, state that the type of dependence in the observations is varied. Being time series the most frequent situation in practice, we present the Singular Spectrum Analysis (SSA), Golyandina (2001), and the Multichannel Singular Spectrum Analysis (MSSA), Oliveira (2003) or Kim (1998) as methods based in PCA.

This work is divided in two parts. In the first we present a description of SSA, which uses the auto-covariance matrix of a time series with only one variable and one application. In the second part we present the multivariate case, the MSSA, together with another application.

References

[1] Golyandina, N., V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Séries Structure: SSA and Related Thechniques*, Chapman & hall, 2001.

[2] Konishi, S., AND C. R. Rau, "Principal component analysis for multivariate familial data," *79. Biometrika*, pp. 631–641, 1992.

[3] Kim, K. Y., AND Q. Wu, "A Comparison Study of EOF Techniques: Analysis of Nonstationary Data with Periodic Statistics," *12. Journal of Climate*, pp. 185–199, 1998.

[4] Oliveira, I., *Correlated Data in Multivariate Analysis*, University of Aberdeen, 2003.

[5] Skinner, C. J., D. J. Holmes, AND T. M. F. Smith, "The effect of sample design on principal component analysis," *81. J. Amer. Statist. Assoc.*, pp. 789–798, 1986.