A note on the number of components retained for multivariate functional principal components analysis

Steven Golovkine

June 13, 2023

Abstract

Your abstract.

1 Introduction

We aim to show that the procedure proposed by [1] may leaad to inconsistency in the retained number of components, based on extensive simulation.

The simulation may varies has follow:

- Number of curves $N = 25, 50, 100, 200$
- Number of sampling points $M = 25, 50, 100, 200$
- Number of components $P = 2, 5, 10, 20, 50$
- No noise and we assume the curve are sampled on a common grid.
- Based on the Karhunen-Loève decomposition, make sure that the decreasing of the eigenvalues is coherent with KL assumptions. The data are defined with a large number of components and different decreasing of the eigenvalues scenarios.
- We do the same for the percentage of variance explained. We set the percentage of variance explained for the multivariate components to be $\alpha\%$ ($\alpha = 50, 75, 90, 95, 99$) and we change the percentage of variance explained by the univariate components.
- The quality of the estimation is based on different measures: the number of retained components (only if we set the percentage of variance explained), the estimation of the eigenvalues ($\log -$AE), the estimation of the multivariate eigenfunctions (ISE) and the reconstruction of the curves (MISE).
- Data are simulated with [1] setting and we can use ICHEC to run them.

2 Ideas

- This should be a quick paper on the selection of the number of components for MFPCA.
- Based on how the number of components is selected in MFPCA [1].
- Just considering the number of components, based on an univariate expansion, we speculate that we need for that say K univariate components to effectively estimate K multivariate components. Let K_p be the number of estimated components for the pth feature and K the number of multivariate components we want to estimate. Computationally speaking, we can estimate up to $\sum_{p=1}^{P} K_p$ multivariate components. We however claim that the number of accurately estimated components is only $\min_{p=1,\dots,P} K_p$.
- The same phenomenon appears with the percentage of variance explained, we can not retrieve $\alpha\%$ of the variance with the multivariate curves, if the univariate components also explained $\alpha\%$ of the univariate curves. This might be related to the "multivariate testing" phenomena. Mentioned in the Supplementary material of [1]. We rerun the sensitivity analysis and aim to show that the percentage of variance explained might be different of one expects.
- This is going to be a practical paper, no proof, except pratical proofs will be presented here. Only work on extensive simulations.

References

[1] C. Happ and S. Greven. Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains. 113(522):649–659.