13 April, 11:10 - 11:30, ESTGV Auditorium

Dynamic Principal Component Analysis

Isabel Silva[1], **Maria Eduarda Silva**[2],
[1] Faculdade de Engenharia, Universidade do Porto and CIDMA, ims@fe.up.pt
[2] Faculdade de Economia, Universidade do Porto and CIDMA, mesilva@fep.up.pt

Multidimensional time series are observed in the most varied fields of application. Principal Component Analysis (PCA) can be used to reduce dimensionality. However, formal inference procedures based on principal components rely on the independence (and multivariate normality) of the observations, a condition that is violated for time series data. In this work, we describe a frequency domain version of PCA proposed by Brillinger [1] that takes into account the correlation in time. Illustration with real data is presented.

Keywords: dimensionality reduction, principal component analysis, spectral analysis, time series data

The multidimensional temporal (and spatio-temporal) series are observed in the most varied fields of application and are characterized by the correlation structure induced by the sequential order of observations. Let $\mathbf{X}_t$ be a $p-$dimensional time series. The process is said to be stationary if $\mathrm{E}[\mathbf{X}_t]$ and $\mathrm{E}[\mathbf{X}_t\mathbf{X}'_{t+h}]$ exist and don't depend of time t. The $p \times p$ autocovariance function is given by

$$\Gamma_{xx}(h) = \mathrm{E}[\mathbf{X}_t\mathbf{X}'_{t+h}] - \mathrm{E}[\mathbf{X}_t]\mathrm{E}[\mathbf{X}'_{t+h}].$$

If $\sum_h \Gamma_{xx}(h) < \infty$ then the spectral density matrix of $\mathbf{X}_t$ is given by

$$f_{xx}(\omega) = \sum_{h=-\infty}^{+\infty} \Gamma_{xx}(h)\exp(2\pi i h\omega).$$

Therefore, the autocovariance function and the spectral density are Fourier transform pairs and therefore contain the same information. As a consequence, there are two approaches (not necessarily mutually exclusive) to analyse time series data. The time domain approach considers the lagged relationships as most important while in the frequency (or spectral) domain, the periodic information is the most important.
In some multidimensional contexts, the number of observations per series exceeds the total number of time series, so it is of great importance to reduce the dimensionality of the data, extracting the most important information and eliminating noise and redundant correlations. By doing this, graphic representation and subsequent statistical analysis of the dataset are facilitated.

One very popular method for dimensionality reduction is Principal Component Analysis, which allows us to obtain a new set of variables, called Principal Components (PC), that are uncorrelated and ordered so that the first few retain most of the variation presented in the dataset (Jolliffe [2]). In some fields of application, PCA not only reduces the dimensionality of the dataset but also allows for reasonable interpretations of the retained PC.
As referred by Jolliffe [2], most of the inference procedures to be performed for PC are based on independence as well as on multivariate normality of the data, condition that are not satisfied for time series data. Several techniques have been proposed to overcome this issue.
One of the developed methodologies is the so called dynamic PCA, proposed by Brillinger [1] for multivariate time series assuming that the underlying process is stationary. As referred by Shumway and Stoffer [3], it can be considered as a PCA in the frequency domain where classical PCA is performed at each frequency, providing a set of principal components series which are uncorrelated at all time lags, thus allowing inferential procedures.
Formally, dynamic PCA approximates a p vector valued time series X_t by a set of k uncorrelated time series Y_t such that Y_t is the best approximation of X_t in mean squared error sense. While the classical ('static') PCA are linear combinations of the original data, the dynamic PC are linear combinations of past, present and future observations.
Note that classical PCA works with a covariance (or correlation) matrix, but in the time series context we can consider (auto)covariance between variables observed at the same time (given in the matrix $\Gamma_{xx}(0)$) but also between variables at different times (given by the matrices $\Gamma_{xx}(k)$ for $k \neq 0$). Therefore, given the equivalence between the autocovariance and the spectral density functions, it is natural to consider PCA in the frequency domain.
The objective of this work is to describe the Dynamic Principal Component Analysis, discussing its strengths and weaknesses and addressing some implementation issues. In addition, the results of the application of this technique to real datasets are exhibited and compared with classical PCA and MSSA (where the original series is decomposed in a small number of independent and interpretable components that can be thought as trend, oscillatory components and a structureless noise).

Acknowledgements The authors were partially supported by Fundação para a Ciência e a Tecnologia (FCT), within project UID/MAT/04106/2019 (CIDMA).

References

[1] D. R. Brillinger. *Time Series: Data Analysis and Theory.* Classics in Applied Mathematics, 36, SIAM, 2001.

[2] I. T. Jolliffe. *Principal Component Analysis.* Springer-Verlag, 2002.

[3] R. Shumway and D. Stoffer. *Time Series Analysis and Its Applications, With R Examples.* Springer, New York, 4th ed. edition, 2017.