Available online at www.sciencedirect.com
ScienceDirect
Physica A 387 (2008) 4077–4098
PHYSICA A
www.elsevier.com/locate/physa

Kernel methods and flexible inference for complex stochastic dynamics

Enrico Capobianco

CRS4 Bioinformatics Laboratory, Science and Technology Park of Sardinia, Pula (Cagliari) - Sardinia, Italy

Received 30 November 2007; received in revised form 17 February 2008
Available online 8 March 2008

Abstract

Approximation theory suggests that series expansions and projections represent standard tools for random process applications from both numerical and statistical standpoints. Such instruments emphasize the role of both sparsity and smoothness for compression purposes, the decorrelation power achieved in the expansion coefficients space compared to the signal space, and the reproducing kernel property when some special conditions are met. We consider these three aspects central to the discussion in this paper, and attempt to analyze the characteristics of some known approximation instruments employed in a complex application domain such as financial market time series. Volatility models are often built *ad hoc*, parametrically and through very sophisticated methodologies. But they can hardly deal with stochastic processes with regard to non-Gaussianity, covariance non-stationarity or complex dependence without paying a big price in terms of either model mis-specification or computational efficiency. It is thus a good idea to look at other more flexible inference tools; hence the strategy of combining greedy approximation and space dimensionality reduction techniques, which are less dependent on distributional assumptions and more targeted to achieve computationally efficient performances. Advantages and limitations of their use will be evaluated by looking at algorithmic and model building strategies, and by reporting statistical diagnostics.

Keywords: Sparse approximation; Reproducing Kernel Hilbert spaces; Wavelets; Overcomplete atomic dictionaries; Frames

1. Introduction

It is often the case in applications about stochastic phenomena that a statistical model has to be more or less heavily constrained in order to deal with non-standard probabilistic conditions characterizing the random variables involved. Examples are offered by self-similar and long range dependent processes, typically investigated in network traffic data and finance, where also non-Gaussianity, non-stationarity and heteroskedasticity are known to characterize the observations.

In experiments where the signals of interest have to be either recovered from noisy environments (extraction and filtering problems) or estimated from indirect observations (deconvolution and blind separation problems), a well-known model strategy is that of switching from the original function space to an image space obtained by projection operators. The image space of transformed values consists of the series expansion coefficient sequences computed from splines, discrete cosines, wavelet, frames, and other functional families.

E-mail address: enrico.capobianco@gmail.com.

0378-4371/$ - see front matter
doi:10.1016/j.physa.2008.03.003

Most of these techniques have certain properties, in particular sparse approximation, reproducing kernel, and decorrelation power, which are by no means the only important aspects, but nevertheless represent important tools for interpreting experimental results and application findings.

For instance, the possibility of representing random functions and signals in a sparse way addresses the importance of compression and reconstruction power. Thus, having a relatively small number of expansion coefficients is useful for encompassing the smoothness degree of the original function space.

With regard to the reproducing kernel property, this embedding holds in the input–output space map allowed by the expansion or projection operators applied to a random function or signal, and this result is very useful when non-linearity and high-dimensionality are involved.

Then, the decorrelation power which can be achieved in the expansion coefficients space, particularly when compared to the correlation structure of the underlying stochastic processes, is key for pursuing statistical inference, especially asymptotically.

It is through these properties that we attempt to analyze the characteristics of learning instruments in the complex application domain of finance. Volatility studies have revealed that many statistical models can hardly deal with the typical features or stylized facts which are empirically observed, i.e. non-Gaussianity, covariance non-stationarity, long memory and persistence, intermittency and clustering effects, just to name a few.

Our efforts are directed to the application of greedy approximation techniques and space dimensionality reduction strategies which are not dependent on distributional assumptions, but instead targeted to achieve a computationally efficient performance while learning the data structure and detecting volatility features.

The paper is organized as follows: Section 2 introduces general approximation aspects, while Sections 3 and 4 illustrate, respectively, orthogonal and non-orthogonal expansions and algorithms; Sections 5–7 show statistical analysis and diagnostics, performance measures, and numerical aspects. Section 8 is for the conclusions.

2. General approximation aspects

2.1. Background

Random processes whose probabilistic properties are not completely known usually lead to the construction of non-parametric or semi-parametric statistical models. Particularly important is the second order statistical information available through the estimated autocovariance (autocorrelation) functions of the observed process realizations. It is by expanding these functions in optimal bases, and then by diagonalizing them, that one knows how well the process may be decomposed and approximated.

The problems of what basis to use and how well to approximate with it, are of course of crucial relevance. Linear and non-linear approximation schemes may be explored. Given a process $X \in \mathcal{H}$ (Hilbert space) and a generic approximating basis function sequence $\{e_1, e_2, \ldots e_N\}$ such that $P_N X = \sum_{k=n}^{N} \langle X, e_k \rangle e_k$ represents the projection of X onto the N-dimensional space where we believe the optimal model belongs, one can measure through the basis functions the approximation error. An immediate example is provided by the typical mean square error (MSE) criterion:

$$\mathrm{MSE}_N = E \left\| X - \sum_{k=n}^{N} \langle X, e_k \rangle e_k \right\|_2^2 \tag{1}$$

with the goal of achieving its minimum $\forall N \geq 0$ among all orthonormal approximating families.

This is the feasible way of measuring the approximation error through a norm, and then checking that its limit goes to 0 when $N \to \infty$. The orthogonal projection onto this fixed N-dimensional subspace spanned by the basis functions delivers a linear approximation. It remains to be established how fast the convergence of the selected N-term approximation is, and what will change when, instead of the first N terms, other selection schemes are used.

The best basis that one can choose to minimize the approximation error of stationary random vector processes (fully specified by second order statistics) is given by a set of eigenvectors ordered by decreasing eigenvalues. This identifies the so-called Karhunen–Loeve transform, as we explain next.