Signal Processing 90 (2010) 566–578

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

The Empirical Likelihood method applied to covariance matrix estimation

Frédéric Pascal [a,*], Hugo Harari-Kermadec [b], Pascal Larzabal [c]

[a] SONDRA/Supelec, 3 rue Joliot-Curie, F-91192 Gif-sur-Yvette Cedex, France
[b] ENS Cachan, 61 Av. du Pdt Wilson, F-94235 Cachan Cedex, France
[c] SATIE, ENS Cachan, CNRS, UniverSud, 61 Av. du Pdt Wilson, F-94235 Cachan Cedex, France

ARTICLE INFO

Article history:
Received 18 July 2008
Received in revised form
16 June 2009
Accepted 23 July 2009
Available online 5 August 2009

Keywords:
Empirical likelihood
Maximum likelihood
Structured parameters estimation
Non-Gaussian noise
Covariance matrix estimation

ABSTRACT

This paper presents a new estimation scheme for signal processing problems in unknown noise field. The Empirical Likelihood has been introduced in the mathematical community, but, surprisingly, it is still unknown in the signal processing community. This estimation method is an alternative to estimate unknown parameters without using a model for the probability density function.

The aim of this paper is twofold: first, the Empirical Likelihood theory is presented and revisited thanks to the moment method. Its properties are derived. Second, to emphasize all the potentiality of this method, we address the problem of Toeplitz matrix estimation: this leads us to obtain improved estimates in comparison to conventional ones, as shown in simulations.

1. Introduction

It is often assumed that signals, interferences or noises are Gaussian stochastic processes. Indeed, this assumption makes sense in many applications. Among them, we can cite: sources localization in passive sonar, radar detection where thermal noise and clutter are often modeled as Gaussian processes, digital communications where the Gaussian hypothesis is widely used for interferences and noises.

In these contexts, Gaussian models have been thoroughly investigated in the framework of statistical estimation and detection theory [1–3]. They have led to attractive algorithms. For instance, we can cite the stochastic Maximum Likelihood method for sources localization in array processing [4,5], the matched filter and its adaptive variants in radar detection [6,7], and in digital communications [8].

However, such widespread techniques suffer from several drawbacks when the noise process is a non-Gaussian stochastic process [9]. Therefore, non-Gaussian noise modeling has gained many interest in the last decades and presently leads to active researches in the literature. Higher order moment methods [10] have initiated this research activity while particle filtering [11] is now intensively investigated. In radar applications, experimental clutter measurements, performed by MIT [12], showed that these data are not correctly described by Gaussian statistical models. More generally, numerous non-Gaussian models have been developed in several engineering fields [13–15].

Nevertheless, the question of a model choice for previous applications remains since, most of the time, chosen modeling does not perfectly describe the data behavior. And in these cases, classical estimation methods like for example Maximum Likelihood (ML) based on

* Corresponding author. Tel.: +33 147405320; fax: +33 147402199.
E-mail addresses: frederic.pascal@supelec.fr (F. Pascal), hugo.harari@ens-cachan.fr (H. Harari-Kermadec), larzabal@satie.ens-cachan.fr (P. Larzabal).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.07.028

the data probability density function (pdf) are used, leading as expected to only sub-optimal results. Several non-parametric techniques are proposed in the literature to estimate this unknown pdf. We can cite for example wavelet methods which have been widely investigated. But, most of them are difficult to implement.

An alternative is the Empirical Likelihood (EL) [16]. This method allows to estimate the unknown parameters without assuming a noise modeling. Moreover, prior informations on the data (known moments, parameter structure, etc.) can be integrated in the processing by means of constraints in the optimization procedure. However, surprisingly, this estimation scheme is still unused in the area of signal processing estimation. To the best of our knowledge, we can only cite [17–19] in the corresponding literature.

The main interest of this method is to estimate unknown parameters of interest without assumption on the pdf. Furthermore, we will estimate only parameters of interest but we do not aim in estimating the pdf. This approach differs from robustness considerations, exposed for example by [20], that lead to M-estimators and minimax theory. Here, we do not have in mind a specific model, from which the true pdf could be slightly apart. However, EL will be shown to be based on a moment equation which links the parameters of interest to the observed data, like in the method of moments [1]. Moreover, EL relies on the optimization of a quantity that reminds of a likelihood.

The mathematical properties of EL have already been established in the statistical literature. The aim of this paper is to introduce the EL method to the signal processing community, by means of classical estimation problems. To illustrate the potentiality of the EL method, we analyze the problem of structured covariance matrix estimation. In particular, we focus on the persymmetric structure [21] and on the Toeplitz structure which is widely used [22–26].

Many applications rely on the covariance matrix estimation such as sources localization or Doppler frequencies detection. In these cases, performance directly depends on the accuracy of the covariance matrix estimate, see, e.g. [27] for the particular case of detection in non-Gaussian noise, or more generally, [5] in array processing problems. The choice of the covariance matrix estimation method affects dramatically the accuracy of the estimation in practical settings. In the field of channel identification, [18] compares EL with conventional methods.

The paper is organized as follows. Section 2 presents the estimation problem of interest while Section 3 gives an original presentation of the EL procedure, adapted to signal processing problems. Sections 3.4 and 4 present two applications of the EL method, first without constraint and then, the EL method uses prior informations. In these sections, comparison with the classical ML method will be analyzed through the problem of covariance matrix estimation under Gaussian assumptions. Then, Section 5 contains simulations which illustrate theoretical results of Section 4.

2. Problem formulation

In this section, we introduce the notations used in this paper and the statistical framework.

2.1. Notations

In the following, H denotes the conjugate transpose operator, $\top$ denotes the transpose operator, * denotes the conjugate operator, $E_P[f(\mathbf{x})]$ is the expectation of the function $f(\mathbf{x})$ when the random variable $\mathbf{x}$ is distributed according to the probability P, $E[\mathbf{x}|f(\mathbf{x})]$ means the expectation of $\mathbf{x}$ subject to $f(\mathbf{x})$. $\mathrm{Tr}(\mathbf{M})$ is the trace of matrix $\mathbf{M}$ and $\det(\mathbf{M})$ is the determinant of matrix $\mathbf{M}$. $\mathbb{C}$ (respectively, $\mathbb{R}$) denotes the set of complex (respectively, real) numbers, while for any integer p, $\mathbb{C}^p$ (respectively, $\mathbb{R}^p$) represents the set of p-vectors with complex (respectively, real) elements. For $z \in \mathbb{C}$, we write $\mathscr{R}e(z)$ and $\mathscr{I}m(z)$ its real and imaginary parts.

2.2. Statistical framework

In a lot of signal processing problems, we have to extract the estimator $\widehat{\boldsymbol{\theta}}$ of parameters $\boldsymbol{\theta}$ based on some noisy data $\mathbf{x}$. This leads to the functional relation:

$$\widehat{\boldsymbol{\theta}} = T(\mathbf{x}). \tag{1}$$

To obtain a useful estimator $T(.)$ of parameter $\boldsymbol{\theta}$, a mathematical model on the data has to be introduced. One of the most widespread model, in the area of signal processing, is the following:

$$\mathbf{x}_k = \mathbf{h}(\boldsymbol{\theta}, k) + \mathbf{n}_k \quad \text{for } k = 1, \ldots, K, \tag{2}$$

where $\mathbf{x}_k$ is the kth observation data, $\mathbf{n}_k$ is an additive noise and $\mathbf{h}(\boldsymbol{\theta}, k)$ is the noiseless part of observation which depends on $\boldsymbol{\theta}$ and k. Sometimes, one has the separability property, $\mathbf{h}(\boldsymbol{\theta}, k) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}_k$, where $\mathbf{A}(\boldsymbol{\theta})$ is a transfer matrix and $\mathbf{s}_k$ is a signal.

Recently, the huge increase of computer capacity has allowed the implementation of very sophisticated and performing methods. One of the most famous of them is the ML method which requires, by construction, the knowledge of the data pdf, up to parameters of interest and possibly, additional nuisance parameters. If such a pdf is given, ML methods lead to very performing results, optimal under specific conditions.

The performance of such estimators critically depends on the pdf assumption. This assumption is chosen to be consistent with the problem, but also to be mathematically convenient. Therefore, the choice is restricted to a small number of classical pdf distributions and the real data pdf can apart, at least slightly, from these families. It would be interesting to develop performing methods which do not require an assumption on the pdf family. In this paper, we propose to consider the Empirical Likelihood method, which is not based on the choice of a model for the pdf.

More precisely, let P_0 be the unknown distribution from which data are generated. In practice, this probability is unknown and a classical approach is to