Signal Processing 122 (2016) 190–203

Contents lists available at ScienceDirect

Signal Processing

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

Variance analysis of unbiased least ℓ_p-norm estimator in non-Gaussian noise

Yuan Chen [a,*], Hing Cheung So [b,1], Ercan Engin Kuruoglu [c,1]

[a] School of Computer & Communication Engineering, University of Science & Technology Beijing, Beijing, China
[b] Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China
[c] Italian National Council of Research (ISTI-CNR), Pisa, Italy

ARTICLE INFO

Article history:
Received 21 August 2015
Received in revised form
30 November 2015
Accepted 1 December 2015
Available online 17 December 2015

Keywords:
Variance analysis
Parameter estimation
Non-Gaussian noise
Fractional lower-order moment
ℓ_p-norm minimization

ABSTRACT

Modeling time and space series in various areas of science and engineering require the values of parameters of interest to be estimated from the observed data. It is desirable to analyze the performance of estimators in an elegant manner without the need for extensive simulations and/or experiments. Among various performance measures, variance is the most basic one for unbiased estimators. In this paper, we focus on the estimator based on the ℓ_p-norm minimization in the presence of zero-mean symmetric non-Gaussian noise. Four representative noise models, namely, α-stable, generalized Gaussian, Student's t and Gaussian mixture processes, are investigated, and the corresponding variance expressions are derived for linear and nonlinear parameter estimation problems at $p \geq 1$. The optimal choice of p for different noise environments is studied, where the global optimality and sensitivity analyses are also provided. The developed formulas are verified by computer simulations and are compared with the Cramér-Rao lower bound.

1. Introduction

Parameter estimation [1] is a common task required in many areas of science and engineering such as radar, sonar, speech, image analysis, biomedicine, communications and seismology. It refers to accurately finding the values of parameters of interest from the observed data which consist of two components, viz., signal and noise. Typically, a deterministic model is adopted for the signal while a random process model is employed for the noise. Among numerous estimators developed in the literature, least squares (LS) and maximum likelihood (ML) methods have been widely used.

To assess the quality of an estimator, two fundamental performance measures in the aspect of accuracy are bias and mean square error (MSE). To calculate bias and MSE, approaches such as Taylor series expansion (TSE) of the estimates [2] or TSE on the error function [3] can be used. Although we have only demonstrated the usefulness of the formulas in the presence of Gaussian noise in [4], it is found that the latter has more applicability and may be simpler to derive particularly for nonlinear parameter estimation problems. Despite its theoretical and computational convenience, it is generally understood that the validity of the Gaussian distribution is at best approximate in reality. In fact, the occurrence of non-Gaussian noise has been reported in many fields [5,6]. Note that some non-Gaussian models correspond to impulsive noise. In this case, the LS approach which is based on ℓ_2-norm minimization of errors, fails to provide reliable parameter estimation, since its performance is very sensitive to outliers. The ML estimator may also not be a proper choice. In particular, it is hard to implement for the process whose probability density function (PDF) has a complicated analytical form or lacks an analytical expression. One strategy

* Corresponding author.
E-mail addresses: qchenyuan00@126.com (Y. Chen), hcso@ee.cityu.edu.hk (H.C. So), ercan.kuruoglu@isti.cnr.it (E.E. Kuruoglu).
[1] EURASIP Member.

http://dx.doi.org/10.1016/j.sigpro.2015.12.003
0165-1684/

is to detect and discard the suspicious observations but it may not be feasible for large data sets or complex applications [5].

Alternatively, M-estimator [7], which is based on robust statistics, can resist outliers without preprocessing the data. Its key idea is to replace the squared residuals in the LS methodology by another function which emphasizes large samples less than the square. The least ℓ_p-norm estimator with $p<2$ belongs to the M-estimator family, which is commonly solved by iterative techniques such as the iteratively reweighted least squares (IRLS), Levenberg–Marquardt (LM) and subgradient methods [8–10].

As a follow-up to [4], we study the performance of the least ℓ_p-norm estimator for parameter estimation in additive non-Gaussian noise in this work. Four representative models, namely, symmetric α-stable (SαS) [11], generalized Gaussian (GG) [12], Student's t [13] and Gaussian mixture (GM) [14] processes, are investigated. For simplicity but without loss of generality, all noise models are assumed zero-mean and symmetrically distributed, implying that the least ℓ_p-norm estimator is unbiased and we only analyze the variance formulas. Note that this study assumes that the availability of the noise statistic information, i.e., PDF and density parameters. Therefore, in the case of unknown noise statistics [15,16], they should be estimated [17,18] prior to applying our results. The logarithm moment method [17] can be employed for density parameter estimation from the available noise-only samples. Even when we know nothing about the noise statistics, the GM model can be utilized to approximate the impulsive noise [18]. The parameters of the GM process, that is, numbers of components and their variances, can be adjusted adaptively based on the noise-only observations.

The rest of this paper is organized as follows. In Section 2, we briefly present the bias and MSE formulas based on TSE on the estimator cost function, and then the least ℓ_p-norm estimator. The SαS, GG, Student's t and GM models, are then reviewed in Section 3. Linear and nonlinear signal models are studied with illustrative examples in Sections 4 and 5, respectively. Selection of p for different non-Gaussian noise models which result in minimum variance as well as global optimality and sensitivity analysis are examined in Section 6. In Section 7, computer simulation results are provided to validate the derived variance formulas and contrast with the Cramér–Rao lower bound (CRLB). Finally, conclusions are drawn in Section 8.

2. Variance formula and least ℓ_p-norm estimator

We start with a general signal model:

$$\mathbf{y}=\mathbf{g}(\mathbf{x})+\mathbf{q}, \tag{1}$$

where $\mathbf{y}=[y_1 \cdots y_N]^T \in \mathbb{R}^N$ is the observation vector with ^{T} being the transpose operator, $\mathbf{g}(\cdot)$ is a known function, $\mathbf{x}=[x_1 \cdots x_M]^T \in \mathbb{R}^M$ is the deterministic parameter vector of interest with $M \le N$ and $\mathbf{q}=[q_1 \cdots q_N]^T \in \mathbb{R}^N$ denotes the additive random noise component with zero location parameter. The task of parameter estimation is to find $\mathbf{x}$ from $\mathbf{y}$.

A common approach for estimating $\mathbf{x}$ is to design a cost function $J(\mathbf{x})$ which is constructed from $\mathbf{y}$, and the estimate of $\mathbf{x}$, denoted by $\hat{\mathbf{x}}$, is computed by minimizing $J(\mathbf{x})$:

$$\hat{\mathbf{x}}=\arg\min_{\mathbf{x}} J(\mathbf{x}). \tag{2}$$

Equivalently,

$$\nabla(J(\hat{\mathbf{x}}))=\mathbf{0}_N, \tag{3}$$

where $\nabla(J(\hat{\mathbf{x}}))$ denotes the gradient vector of $J(\mathbf{x})$ at $\mathbf{x}=\hat{\mathbf{x}}$ and $\mathbf{0}_N \in \mathbb{R}^N$ is a column vector with all zeros. In this study, we consider a general class of estimators such that $J(\mathbf{x})$ is twice differentiable. When $\hat{\mathbf{x}}$ is located at a reasonable proximity to $\mathbf{x}$, which is valid when the noise is small enough and/or observation number is sufficiently large, the truncated TSE of $\nabla(J(\hat{\mathbf{x}}))$ around $\mathbf{x}$ is

$$\nabla(J(\hat{\mathbf{x}})) \approx \nabla(J(\mathbf{x}))+\mathbf{H}(J(\mathbf{x}))(\hat{\mathbf{x}}-\mathbf{x}), \tag{4}$$

where $\mathbf{H}(J(\mathbf{x}))$ is the Hessian matrix. Assuming that $\mathbf{H}(J(\mathbf{x}))$ is smooth enough to have $\mathbf{H}(J(\mathbf{x})) \approx E\{\mathbf{H}(J(\mathbf{x}))\}$, (4) can be utilized to compute the bias and MSE of $\hat{\mathbf{x}}$:

$$\text{bias}(\hat{\mathbf{x}})=E\{\hat{\mathbf{x}}\}-\mathbf{x} \approx -\left(E\{\mathbf{H}(J(\mathbf{x}))\}\right)^{-1} E\{\nabla(J(\mathbf{x}))\}, \tag{5}$$

$$\begin{aligned}\mathbf{M}(\hat{\mathbf{x}}) &= E\{(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^T\} \\ &\approx \left(E\{\mathbf{H}(J(\mathbf{x}))\}\right)^{-1} E\{\nabla(J(\mathbf{x}))\nabla^T(J(\mathbf{x}))\}\left(E\{\mathbf{H}(J(\mathbf{x}))\}\right)^{-1},\end{aligned} \tag{6}$$

where E denotes the expectation operator and $(\cdot)^{-1}$ is the matrix inverse. The MSE of $\hat{x}_m$, $m=1,\ldots,M$, is given by the (m,m) entry of $\mathbf{M}(\hat{\mathbf{x}})$. Note that the expressions in (5) and (6) are exact when $J(\mathbf{x})$ is a quadratic function. Furthermore, the validity of (5) and (6) may not relate to the parameter dimension, namely, M, but depends on whether $\mathbf{x}$ is linear in the observation vector or not. It is because nonlinear estimators nearly always exhibit the threshold effect [1].

For an unbiased estimator where $\text{bias}(\hat{\mathbf{x}})=\mathbf{0}_M$, the MSE is in fact the variance. Then the covariance matrix for $\hat{\mathbf{x}}$, denoted by $\mathbf{C}(\hat{\mathbf{x}})$, is approximated by (6) while the variance of $\hat{x}_m$, denoted by $\text{var}(\hat{x}_m)$, is provided by the (m,m) entry of $\mathbf{C}(\hat{\mathbf{x}})$.

Denote $\mathbf{g}(\mathbf{x})=[g_1(\mathbf{x}) \cdots g_N(\mathbf{x})]^T$. A typical choice for $J(\mathbf{x})$ is the LS cost function:

$$J(\mathbf{x})=\sum_{n=1}^{N}|y_n-g_n(\mathbf{x})|^2, \tag{7}$$

which corresponds to ℓ_2-norm minimization. It is well known that the LS solution is equivalent to the ML estimate when $\mathbf{q}$ is a zero-mean white Gaussian process. In fact, (5) and (6) have been verified in [4] for LS-based parameter estimation in the presence of white Gaussian noise. Nevertheless, when the noise is non-Gaussian distributed, particularly if $\mathbf{q}$ is impulsive, unreliable parameter estimation will result since the performance of the ℓ_2-norm minimizer is very sensitive to outliers. To achieve robust estimation, ℓ_p-norm minimization with $p<2$ is widely used since it is less sensitive to outliers than the square function. In this work, we focus on the ℓ_p-norm