Signal Processing 104 (2014) 381–386

Contents lists available at ScienceDirect

Signal Processing

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

Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched conditions

Maria Greco *, Stefano Fortunati, Fulvio Gini

Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy

ARTICLE INFO

Article history:
Received 29 November 2013
Received in revised form
12 March 2014
Accepted 3 April 2014
Available online 26 April 2014

Keywords:
Matrix estimation
Complex elliptically symmetric distribution
Huber sandwich matrix
Cramér–Rao lower bound

ABSTRACT

This paper deals with the maximum likelihood (ML) estimation of scatter matrix of complex elliptically symmetric (CES) distributed data when the hypothesized and the true model belong to the CES family but are different, then under mismatched model condition. Firstly, we derive the Huber limit, or sandwich matrix expression, for a generic CES model. Then, we compare the performance of mismatched and matched ML estimators to the Huber limit and to the Cramér–Rao lower bound (CRLB) in some relevant study cases.

1. Introduction

A fundamental assumption underlying the analysis of the statistical properties of the maximum likelihood (ML) estimators is that the true data model and the model used for the ML estimator calculation are the same, that is, the model is correctly specified. Unfortunately, this is not always the case and a model mismatch is possible. It is natural to ask ourselves what happens to the ML estimators properties under mismatched conditions. Huber [7] and White [14] have provided an interesting answer to this question. In their work, they proved that the asymptotic distribution of the ML estimator in misspecified models is concentrated on the Kullback–Leibler (KL) divergence minimizing pseudo-true value and it is Gaussian with the "sandwich" covariance matrix, to first asymptotic order.

With a formal and rigorous description, let $\{\mathbf{z}_k\}_{k=1}^K$ be a set of independent and identically distributed (IID) vectors, each one with probability distribution $F(\mathbf{z})$ that admits a measurable probability density function (pdf) $f = dF/d\mu$ with respect to some σ-finite measure μ. Suppose that a model with probability distribution $H(\mathbf{z},\boldsymbol{\theta})$ and measurable pdf $h(\mathbf{z},\boldsymbol{\theta}) = dH(\mathbf{z},\boldsymbol{\theta})/d\mu$, with $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^n$, is assumed, yielding a log-likelihood function (LLF) equal to $\ln L_K(\boldsymbol{\theta}) = \sum_{k=1}^K \ln h(\mathbf{z}_k,\boldsymbol{\theta})$. If $F(\mathbf{z}) \neq H(\mathbf{z},\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta$, then the assumed model is misspecified.

Let $\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} \ln L_K(\boldsymbol{\theta})$ be the ML estimator of $\boldsymbol{\theta}$. Since $K^{-1} \ln L_K(\boldsymbol{\theta})$ is the sample mean estimator for $l_0(\boldsymbol{\theta}) = E\{\ln h(\mathbf{z},\boldsymbol{\theta})\}$, $\hat{\boldsymbol{\theta}}_{ML}$ will be consistent for the value $\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta}} E\{\ln h(\mathbf{z},\boldsymbol{\theta})\}$, where the expectation is calculated with respect to $F(\mathbf{z})$. If $F(\mathbf{z})$ is absolutely continuous with respect to $H(\mathbf{z},\boldsymbol{\theta})$, then

$$l_0(\boldsymbol{\theta}) - E\{\ln f(\mathbf{z})\} = -\int \ln \frac{f(\mathbf{z})}{h(\mathbf{z},\boldsymbol{\theta})} dF(\mathbf{z}) = -KL(F,H), \qquad (1)$$

where $KL(F,H)$ in the KL divergence between the true model $F(\mathbf{z})$ and the assumed model $H(\mathbf{z},\boldsymbol{\theta})$, so $\boldsymbol{\theta}_0$ is also the KL minimizing vector, i.e. $\boldsymbol{\theta}_0 = \operatorname{argmin}_{\boldsymbol{\theta}} KL(F,H)$. In the correctly specified model, $\boldsymbol{\theta}_0$ is the true data generating parameter vector. In misspecified models, this is the "pseudo-true" vector. Moreover, invoking the central-limit theorem (CLT), Huber and White proved that

* Corresponding author.
E-mail address: m.greco@iet.unipi.it (M. Greco).

http://dx.doi.org/10.1016/j.sigpro.2014.04.002
0165-1684/© 2014 Published by Elsevier B.V.

when $K\to\infty$ we have $\sqrt{K}(\hat{\boldsymbol\theta}_{ML}-\boldsymbol\theta_0)\to N(\mathbf{0},\mathbf{H}(\boldsymbol\theta_0))$, where $\mathbf{H}(\boldsymbol\theta_0)=\mathbf{C}(\boldsymbol\theta_0)^{-1}\mathbf{B}(\boldsymbol\theta_0)\mathbf{C}(\boldsymbol\theta_0)^{-1}$, and

$$[\mathbf{C}(\boldsymbol\theta_0)]_{ij}=E\left\{\frac{\partial^2\ln h(\mathbf{z};\boldsymbol\theta_0)}{\partial\theta_i\partial\theta_j}\right\},$$

$$[\mathbf{B}(\boldsymbol\theta_0)]_{ij}=E\left\{\frac{\partial\ln h(\mathbf{z};\boldsymbol\theta_0)}{\partial\theta_i}\frac{\partial\ln h(\mathbf{z};\boldsymbol\theta_0)}{\partial\theta_j}\right\},$$

where the mean values are taken with respect to the true data pdf $f(\mathbf{z})$. $\mathbf{H}(\boldsymbol\theta_0)$ is the Huber "sandwich" matrix. If the model is correctly specified, $\mathbf{B}(\boldsymbol\theta_0)=-\mathbf{C}(\boldsymbol\theta_0)$ and both $\mathbf{B}(\boldsymbol\theta_0)$ and $\mathbf{H}^{-1}(\boldsymbol\theta_0)$ are equal to the Fisher information matrix (FIM). Differently from the Cramér–Rao lower bound (CRLB), that is calculated from the FIM, the Huber limits are not lower nor upper bounds. However, they can help in measuring the effect of model mismatch, at least asymptotically, that is, for a large number K of independent data vectors.

In this work, we derive the Huber limits on the estimation of the covariance matrix of CES distributed data, that is, when the true and the assumed pdf models belong to the CES family, but they are different.

Notations: We use $\mathrm{tr}(\mathbf{A})$, $|\mathbf{A}|$ and $\mathbf{A}^H$ to denote the trace, i.e. the sum of all the elements along the matrix main diagonal, the determinant and the Hermitian of the matrix $\mathbf{A}$, respectively. Moreover $=_{\mathrm{d}}$ means equal in distribution.

2. Huber limits for CES distributed random vectors

CES distributions constitute a wide family of distributions whose complex Gaussian, Cauchy, Generalized Gaussian, compound-Gaussian, such as K-distribution and complex-*t*, are particular cases. The CES distributions are widely applied in many areas, such as radar, sonar, and communications [11,10]. In many applications, where adaptive signal processing is performed, the estimation of the observation vector covariance matrix is required (see e.g. [9,6,1–4,15,16]), that is why we address here the problem of asymptotic performance evaluation of ML matrix estimators under mismatched modeling.

A complex N-dimensional random vector $\mathbf{z}$ is CES distributed, in shorthand notation $\mathbf{z}\in CE_N(\boldsymbol\mu,\boldsymbol\Sigma,g)$, if its pdf is of the form

$$h(\mathbf{z})=c_{N,g}|\boldsymbol\Sigma|^{-1}g((\mathbf{z}-\boldsymbol\mu)^H\boldsymbol\Sigma^{-1}(\mathbf{z}-\boldsymbol\mu)),\qquad(2)$$

where g is the density generator, $c_{N,g}$ is a normalizing constant, $\boldsymbol\mu=E\{\mathbf{z}\}$ and $\boldsymbol\Sigma$ is the full-rank normalized covariance matrix, also called *scatter matrix*, such that $\mathrm{tr}(\boldsymbol\Sigma)=N$. In particular, if $\mathbf{M}=E\{(\mathbf{z}-\boldsymbol\mu)(\mathbf{z}-\boldsymbol\mu)^H\}$ is the covariance matrix of the vector $\mathbf{z}$, then $\boldsymbol\Sigma=N/\mathrm{tr}(\mathbf{M})\cdot\mathbf{M}$. It is important to observe that for some CES distributions the unnormalized covariance matrix $\mathbf{M}=E\{\mathbf{z}\mathbf{z}^H\}$ does not exist, but the scatter matrix $\boldsymbol\Sigma$ is still well defined.

Based upon the stochastic representation theorem, [10] any $\mathbf{z}\in CE_N(\boldsymbol\mu,\boldsymbol\Sigma,g)$ with $rank(\boldsymbol\Sigma)=k\le N$ admits the stochastic representation $\mathbf{z}=_d\boldsymbol\mu+R\mathbf{T}\mathbf{u}=_d\boldsymbol\mu+R\mathbf{T}(\mathbf{w}/R_w)$, where the non-negative random variable (r.v.) $R\triangleq\sqrt{Q}$, the so-called *modular variate*, is a real, non-negative random variable, $\mathbf{u}$ is a k-dimensional vector uniformly distributed on the unit hyper-sphere with $k-1$ topological dimensions such that $\mathbf{u}^H\mathbf{u}=1$, R and $\mathbf{u}$ are independent and $\boldsymbol\Sigma=\mathbf{T}\mathbf{T}^H$ is a factorization of $\boldsymbol\Sigma$, where $\mathbf{T}$ is a $N\times k$ matrix and $rank(\mathbf{T})=k$. In the following derivations, we suppose that $\boldsymbol\Sigma$ is full-rank, then $rank(\mathbf{T})=rank(\boldsymbol\Sigma)=N$, and that it is real. $\mathbf{w}$ is a complex normal distributed random vector, $\mathbf{w}\sim CN(\mathbf{0},\mathbf{I})$, and $R_w^2\triangleq Q_w$ is a Gamma distributed non-negative r.v., independent of $\mathbf{u}$ and $\mathbf{z}$, with shape and scale parameters equal to N and 1, respectively, i.e. $Q_w\sim\mathrm{Gam}(N,1)$. In particular, we have that $E\{Q_w\}=N$ and $E\{Q_w^2\}=N(N+1)$.

Since in many scenarios (e.g. radar and sonar) the mean value of the data vectors can be considered null, we assume in the derivations $\boldsymbol\mu=\mathbf{0}$. Moreover, we suppose that all the characteristic parameters of the CES distributions are known, except the elements of the scatter matrix $\boldsymbol\Sigma$, hence in our case $\boldsymbol\theta=\mathrm{vec}(\boldsymbol\Sigma)$. It is worth noting that the following derivation is valid also in the case that not all the elements of the scatter matrix $\boldsymbol\Sigma$ are unknown, e.g. because the matrix has some a priori known structure (e.g. symmetric, known trace, or autoregressive model). In this case, $\boldsymbol\theta$ is only a subset of the elements of $\boldsymbol\Sigma$. If all the elements are unknown, the Huber limit for CES distributions represents an alternative formulation of the asymptotic covariance of M-estimators, as presented in [10]. Let us now calculate the Huber limits.

Evaluation of $E\left\{\frac{\partial\ln h(\mathbf{z};\boldsymbol\theta)}{\partial\theta_i}\frac{\partial\ln h(\mathbf{z};\boldsymbol\theta)}{\partial\theta_j}\right\}$

By defining $t=\mathbf{z}^H\boldsymbol\Sigma^{-1}\mathbf{z}$, $\mathbf{A}_i=\partial\boldsymbol\Sigma/\partial\theta_i$, and remembering that $(\partial\ln|\boldsymbol\Sigma|/\partial\theta_i)=\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_i)$ and $(\partial(\mathbf{z}^H\boldsymbol\Sigma^{-1}\mathbf{z}))/\partial\theta_i=-\mathbf{z}^H\boldsymbol\Sigma^{-1}\mathbf{A}_i\boldsymbol\Sigma^{-1}\mathbf{z}$ [8, p. 521,13, p. 1401], we have that

$$\frac{\partial\ln h(\mathbf{z};\boldsymbol\theta)}{\partial\theta_i}=-\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_i)-\frac{\partial\ln g(t)}{\partial t}\mathbf{z}^H\boldsymbol\Sigma^{-1}\mathbf{A}_i\boldsymbol\Sigma^{-1}\mathbf{z}.\qquad(3)$$

Then,

$$\begin{aligned}\frac{\partial\ln h(\mathbf{z};\boldsymbol\theta)}{\partial\theta_i}\frac{\partial\ln h(\mathbf{z};\boldsymbol\theta)}{\partial\theta_j}&=\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_i)\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_j)\\&+\frac{\partial\ln g(t)}{\partial t}\mathbf{z}^H\boldsymbol\Sigma^{-1}(\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_i)\mathbf{A}_j+\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_j)\mathbf{A}_i)\boldsymbol\Sigma^{-1}\mathbf{z}\\&+\left(\frac{\partial\ln g(t)}{\partial t}\right)^2\mathbf{z}^H\boldsymbol\Sigma^{-1}\mathbf{A}_i\boldsymbol\Sigma^{-1}\mathbf{z}\mathbf{z}^H\boldsymbol\Sigma^{-1}\mathbf{A}_j\boldsymbol\Sigma^{-1}\mathbf{z}.\end{aligned}\qquad(4)$$

By making use of the stochastic representation of $\mathbf{z}$ [10] in (4), we can state that $t=\mathbf{z}^H\boldsymbol\Sigma^{-1}\mathbf{z}=R^2\mathbf{u}^H\mathbf{T}^H\boldsymbol\Sigma^{-1}\mathbf{T}\mathbf{u}=R^2\triangleq Q$ (the so-called *second-order modular variate*). Defining the vector $\mathbf{x}=\mathbf{T}\mathbf{u}=\mathbf{T}(\mathbf{w}/R_w)$, with R_w independent of $\mathbf{w}$, and the vector $\mathbf{t}=\mathbf{T}\mathbf{w}\sim CN(\mathbf{0},\boldsymbol\Sigma)$ with $\boldsymbol\Sigma=E\{\mathbf{t}\mathbf{t}^H\}$ we can write

$$\begin{aligned}\frac{\partial\ln h(\mathbf{z};\boldsymbol\theta)}{\partial\theta_i}\frac{\partial\ln h(\mathbf{z};\boldsymbol\theta)}{\partial\theta_j}&=\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_i)\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_j)\\&+Q\frac{\partial\ln g(Q)}{\partial Q}\mathbf{x}^H\boldsymbol\Sigma^{-1}(\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_i)\mathbf{A}_j+\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_j)\mathbf{A}_i)\boldsymbol\Sigma^{-1}\mathbf{x}\\&+\left(Q\frac{\partial\ln g(Q)}{\partial Q}\right)^2\mathbf{x}^H\boldsymbol\Sigma^{-1}\mathbf{A}_i\boldsymbol\Sigma^{-1}\mathbf{x}\mathbf{x}^H\boldsymbol\Sigma^{-1}\mathbf{A}_j\boldsymbol\Sigma^{-1}\mathbf{x}.\end{aligned}\qquad(5)$$

Now observing that, for instance [5], $E\{\mathbf{t}^H\boldsymbol\Sigma^{-1}\mathbf{A}_j\boldsymbol\Sigma^{-1}\mathbf{t}\}=E(R_w^2)E\{\mathbf{x}^H\boldsymbol\Sigma^{-1}\mathbf{A}_j\boldsymbol\Sigma^{-1}\mathbf{x}\}$, and using the properties of complex Gaussian vectors [8, p. 564], that $E\{\mathbf{t}^H\boldsymbol\Sigma^{-1}\mathbf{A}_j\boldsymbol\Sigma^{-1}\mathbf{t}\}=\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_j\boldsymbol\Sigma^{-1}E\{\mathbf{t}\mathbf{t}^H\})=\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_j)$ and $E\{\mathbf{t}^H\mathbf{C}\mathbf{t}\ \mathbf{t}^H\mathbf{D}\mathbf{t}\}=\mathrm{tr}(\mathbf{C}\boldsymbol\Sigma)\mathrm{tr}(\mathbf{D}\boldsymbol\Sigma)+\mathrm{tr}(\mathbf{C}\boldsymbol\Sigma\mathbf{D}\boldsymbol\Sigma)$, where $\mathbf{C}$ and $\mathbf{D}$ are Hermitian matrices, we obtain

$$E\left\{\frac{\partial\ln h(\mathbf{z};\boldsymbol\theta)}{\partial\theta_i}\frac{\partial\ln h(\mathbf{z};\boldsymbol\theta)}{\partial\theta_j}\right\}=\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_i)\mathrm{tr}(\boldsymbol\Sigma^{-1}\mathbf{A}_j)$$