Signal Processing 138 (2017) 53–62

Contents lists available at ScienceDirect

Signal Processing

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

Enhanced bootstrap method for statistical inference in the ICA model

Shahab Basiri*, Esa Ollila, Visa Koivunen

Dept. of Signal Processing and Acoustics, Aalto University, P.O. Box 13000, FI-00076 Aalto, Finland

A R T I C L E I N F O

Article history:
Received 28 October 2016
Revised 14 February 2017
Accepted 6 March 2017
Available online 7 March 2017

Keywords:
Bootstrap
Fast-and-robust bootstrap
Independent component analysis
FastICA

A B S T R A C T

In this paper, we develop low complexity and stable bootstrap procedures for FastICA estimators. Our bootstrapping techniques allow for performing cost efficient and reliable bootstrap-based statistical inference in the ICA model. Performing statistical inference is needed to quantitatively assess the quality of the estimators and testing hypotheses on mixing coefficients in the ICA model. The developed bootstrap procedures stem from the fast and robust bootstrap (FRB) method [1], which is applicable for estimators that may be found as solutions to fixed-point (FP) equations. We first establish analytical results on the structure of the weighted covariance matrix involved in the FRB formulation. Then, we exploit our analytical results to compute the FRB replicas at drastically reduced cost. The developed enhanced FRB method (EFRB) for FastICA permits using bootstrap-based statistical inference in a variety of applications (e.g., EEG, fMRI) in which ICA is commonly applied. Such an approach has not been possible earlier due to incurred substantial computational efforts of the conventional bootstrap. Our simulation studies compare the complexity and numerical stability of the proposed methods with the conventional bootstrap method. We also provide an example of utilizing the developed bootstrapping techniques in identifying equipotential lines of the brain dipoles from electroencephalogram (EEG) recordings.

1. Introduction

Independent component analysis (ICA) [2,3] is a widely used multivariate analysis technique for extracting unobserved independent source signals from their observed multivariate mixture recordings. During the past two decades, a myriad of methods for separating source signals have been developed [2], whereas less attention has been paid to developing statistical inference framework (standard errors of the estimators, hypothesis testing) for ICA.

In many problems where ICA is commonly applied, statistical inference is needed for either assessing the stability of the estimated independent components [4,5] or validating prior hypothesis posed by practitioners [6]. For example, in many sensing applications, propagation of a source signal-of-interest may be local and limited only to a few number of sensors in the vicinity of the source. Magnetoencephalography (MEG) is a good example of such a sensing modality. Therefore, tests for hypotheses on the coefficients of the mixing matrix are needed to identify contribution of a specific source signal-of-interest onto a specific mixture variable (sensor). This in turn reveals sparsity of the ICA mixing matrix which can be exploited as a prior information to improve upon the estimation accuracy and/or reducing the dimensionality.

Bayesian inferential methods [7–9] and inference based on asymptotic statistics [10,11] are proposed for performing statistical inference in the ICA model. Bayesian approaches require knowledge of prior distributions of the model parameters. As pointed out in [7, Section 5], performance of such probabilistic methods vastly depends on the validity of the assumed prior distributions. In addition, the number of Bayesian model components grow exponentially as a function of the latent space dimensionality [7, Section 5].

Also statistical tests that are based on asymptotic distribution of the mixing matrix estimator as in [10,11] perform poorly when the sample size is not several orders of magnitude larger than the dimension. Moreover, for real-world data, the ICA model (e.g., linear mixing) can be at best only approximately true, and thus validity of tests based on asymptotic normality of estimators can be questioned.

The main objective of this paper is on developing a stable and practical bootstrap procedure which can be used to perform statistical inference on the FastICA estimates of mixing matrix [2,12] in the linear ICA model. Such statistical inferences are used to identify which sources contribute to a specific observed mixture signal.

Bootstrap is an increasingly important statistical inference tool, widely used in testing hypotheses, characterizing empirical distributions of the parameters and assessing the quality of estimators in terms of standard error, variance, confidence intervals, etc. [13–15]. Constructing bootstrap distributions by *conventional bootstrap* (CB) method [13] requires the calculation of the mixing matrix es-

* Corresponding author.
E-mail address: shahab.basiri@aalto.fi (S. Basiri).

http://dx.doi.org/10.1016/j.sigpro.2017.03.005
0165-1684/

timator $\hat{\mathbf{A}}$ for several thousands bootstrap samples which may be computationally infeasible in many applications.

Techniques developed in this paper stem from the fast and robust bootstrap (FRB) method [1,16,17] which is applicable for estimators that may be found as solutions to fixed-point (FP) estimation equations. We establish analytical results for the weighted covariance matrix involved in the FRB formulation. This allows us to compute the FRB replicas of the demixing matrix at drastically reduced cost. Such an Enhanced FRB for FastICA is referred to as EFRB in the sequel. The proposed EFRB procedure provides further substantial reductions in computation times of bootstrap-based estimates and quantitative performance measures such as computing standard errors of the estimates or testing hypotheses.

The non-convergence runs of the FastICA algorithms occur more frequently as $\kappa = d/n$ grows, where n and d denote the number of distinct observations and latent space dimensionality respectively i.e., see Table I in [18] and Table III-VI in [19]. Such a convergence problem increasingly arises when FastICA algorithms are run on bootstrap samples. This is due to the fact that, in the process of sampling with replacement from the observed data set only about 63% of the original data points appear in a bootstrap sample [14]. This reduced number of distinct data points in bootstrap samples leads to more frequent convergence problems with the fastICA algorithm.

One of the key advantages of the EFRB procedures developed in this paper is that they do not require the computation of the FastICA estimator of the mixing matrix for each bootstrap sample. As a consequence, the EFRB procedure is fast to compute and it avoids the above mentioned convergence problems.

As a practical example, we develop bootstrap tests for hypothesis $\mathcal{H}_0 : a_{ij} = 0$ vs $\mathcal{H}_1 : a_{ij} \neq 0$, where a_{ij} denotes the (i, j)th element of the unknown mixing matrix parameter $\mathbf{A}$ of the ICA model. The motivation is to identify which sources contribute to the observed mixtures in the ICA model. This is of high interest when the mixing modality may be local as in many biomedical measurements such as MEG. Utility of such inference procedure is illustrated in the context of analysing electroencephalogram (EEG) recordings.

The paper is organized as follows. In Section 2, the ICA model and the FastICA estimators are briefly reviewed. In Section 3, two approaches to bootstrapping the ICA model are described. The new bootstrap method, EFRB, is proposed in Section 4. In Section 5 numerical examples are provided and performance evaluations are performed. An example of utilizing the new bootstrap method in analysing EEG signals is provided in Section 6. Section 7 concludes.

2. ICA model and the FastICA estimator

Recall that in the linear ICA model the random vector, $\mathbf{y} \in \mathbb{R}^p$ is a *linear* mixture of unobserved random *source* vector $\mathbf{s} = (s_1, \ldots, s_d)^\top$ possessing statistically independent components (IC's), i.e.,

$$\mathbf{y} = \mathbf{A}\mathbf{s} = \mathbf{a}_1 s_1 + \cdots + \mathbf{a}_d s_d, \tag{1}$$

where $\mathbf{A} = \begin{pmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_d\end{pmatrix}$ is the unknown $p \times d$ *mixing matrix* whose element $a_{ij} = [\mathbf{A}]_{ij}$ represents the contribution of the jth source s_j onto the ith mixture variable y_i, where $i \in \{1, \ldots, p\}$, $j \in \{1, \ldots, d\}$ and $p \geq d$. A common preprocessing step in ICA is whitening transform that decorrelates the observed data and performs dimensionality reduction. Let $\mathbf{D}$ denote the *whitening matrix*, $\mathbf{D} = \mathbf{L}^{-1/2}\mathbf{E}^\top$, where $\mathbf{L} = \text{diag}(l_1, \ldots, l_d)$ consists of the d non-zero eigenvalues of the *covariance matrix*, $\mathcal{C}(\mathbf{y}) = \mathbb{E}[\mathbf{y}\mathbf{y}^\top]$, and $\mathbf{E}$ is the $p \times d$ matrix of respective eigenvectors as its column vectors, so $\mathcal{C}(\mathbf{y}) = \mathbf{E}\mathbf{L}\mathbf{E}^\top$. The whitened data $\mathbf{x} = \mathbf{D}\mathbf{y}$ satisfies $\mathbf{s} = \mathbf{W}\mathbf{x}$ for an orthogonal full rank $d \times d$ *demixing matrix* $\mathbf{W}$. In other words, the whitened data $\mathbf{x}$ follows the ICA model

$$\mathbf{x} = \mathbf{W}^\top \mathbf{s} = \mathbf{w}_1 s_1 + \cdots + \mathbf{w}_d s_d, \tag{2}$$

where the unknown $d \times d$ mixing matrix $\mathbf{W}^\top = \mathbf{D}\mathbf{A}$ is orthogonal. Thus the problem of source extraction is considerably simpler for whitened data.

From this point onwards we assume that the data is prewhitened and centered, i.e. has zero mean, which also implies that $\mathbb{E}[\mathbf{s}] = \mathbf{0}$. We seek for an orthogonal *demixing matrix* $\mathbf{W} = \begin{pmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_d\end{pmatrix}^\top$, whose (transposed) row vector $\mathbf{w}_j$ is called the jth *demixing vector*. In ICA, the demixing matrix can be estimated up to sign, scale and permutation ambiguities of the demixing vectors.

Let us now present the necessary definitions for the FastICA estimator. We define the *inner product* in the vector space $\mathbb{R}^d$ w.r.t. to $\mathcal{C} = \mathcal{C}(\mathbf{x}) = \mathbb{E}[\mathbf{x}\mathbf{x}^\top]$ as $\langle \mathbf{w}, \mathbf{v}\rangle_{\mathcal{C}} = \mathbf{w}^\top \mathcal{C}\mathbf{v}$. Let $\|\mathbf{w}\|_{\mathcal{C}} = \mathbf{w}^\top \mathcal{C}\mathbf{w}$ denote the induced norm. When $\mathcal{C} = \mathbf{I}$, we often write $\|\mathbf{x}\|$ for brevity, i.e., $\|\cdot\|$ without subscript $\mathcal{C}$ should be read as regular Euclidean norm. It should be noted that for whitened data, $\mathcal{C} = \mathbf{I}$. However, we keep the covariance matrix in our derivations, since later on the bootstrap samples are formed from whitened data samples for which the (sample) covariance matrix is no longer an identity matrix. The *1-unit* FastICA estimator finds a FastICA demixing vector $\mathbf{w}$ as a local maxima of a non-Gaussianity measure $\left|\mathbb{E}\left[G(\mathbf{w}^\top\mathbf{x})\right]\right|$ under the unit-norm constraint $\|\mathbf{w}\|^2_{\mathcal{C}} = \mathbf{w}^\top \mathcal{C}\mathbf{w} = 1$, where G can be any twice continuously differentiable nonlinear and non-quadratic function with $G(0) = 0$; see [20, Chapter 8]. Thus the *1-unit* FastICA estimator maximizes the Lagrangian

$$\mathcal{L}_{1U}(\mathbf{w}; \lambda) = \left|\mathbb{E}\left[G(\mathbf{w}^\top\mathbf{x})\right]\right| - \frac{\lambda}{2}(\mathbf{w}^\top \mathcal{C}\mathbf{w} - 1), \tag{3}$$

where λ is the Lagrange multiplier. We write $g = G'$ and $g' = G''$ for the 1st and 2nd derivative of G respectively, where g is referred to as ICA *nonlinearity*.

When more than one sources need to be extracted, there exists two approaches: In *deflation-based* FastICA the demixing vectors $\mathbf{w}_{k,\,g}$, $(k = 1, \ldots, d)$, are estimated one-by-one by iterating the following steps until convergence:

1. $\mathbf{w}_{k,g} \leftarrow \mathcal{C}^{-1}\mathbb{E}[g(\mathbf{w}_{k,g}^\top\mathbf{x})\mathbf{x}] - \mathbb{E}[g'(\mathbf{w}_{k,g}^\top\mathbf{x})]\mathbf{w}_{k,g}$
2. $\mathbf{w}_{k,g} \leftarrow \Pi^\perp_{k-1}\mathbf{w}_{k,g}$
3. $\mathbf{w}_{k,g} \leftarrow \mathbf{w}_{k,g}/\|\mathbf{w}_{k,g}\|_{\mathcal{C}}$

where

$$\Pi^\perp_{k-1} = \mathbf{I} - \sum_{i=1}^{k-1}\mathbf{w}_{i,g}\mathbf{w}_{i,g}^\top\mathcal{C} = \mathbf{I} - \sum_{i=1}^{k-1}\langle \mathbf{w}_{i,g}, \cdot\rangle_{\mathcal{C}}\,\mathbf{w}_{i,g},$$

is an orthogonal projection operator that projects onto the orthogonal complement of the subspace (of the inner product space) spanned by the previously found FastICA demixing vectors $\mathbf{w}_{1,g}, \ldots, \mathbf{w}_{k-1,g}$.

In *symmetric FastICA* approach demixing vectors are estimated simultaneously with a symmetric orthogonality constraint on demixing vectors ($\mathbf{W}^\top\mathcal{C}\mathbf{W} = \mathbf{I}$). The symmetric FastICA algorithm computes the demixing matrix $\mathbf{W}_g = (\mathbf{w}_{1,g} \;\cdots\; \mathbf{w}_{d,g})^\top$ by iterating the following steps:

1. $\mathbf{w}_{k,g} \leftarrow \mathcal{C}^{-1}\mathbb{E}[g(\mathbf{w}_{k,g}^\top\mathbf{x})\mathbf{x}] - \mathbb{E}[g'(\mathbf{w}_{k,g}^\top\mathbf{x})]\mathbf{w}_{k,g}$ for $k = 1, \ldots, d$
2. $\mathbf{W}_g \leftarrow \left(\mathbf{W}_g\mathcal{C}\mathbf{W}_g^\top\right)^{-1/2}\mathbf{W}_g$

until further iterations do not update the previous results.

Finally, after the orthogonal $d \times d$ demixing matrix $\mathbf{W}_g = \begin{pmatrix}\mathbf{w}_{1,g} & \cdots & \mathbf{w}_{1,g}\end{pmatrix}^\top$ for whitened data $\mathbf{x} = \mathbf{D}\mathbf{y}$ has been found, the FastICA estimate $\mathbf{A}_g$ of the mixing matrix $\mathbf{A}$ of original (non-whitened) data $\mathbf{y}$ is computed as

$$\mathbf{A}_g = \mathbf{D}^- \mathbf{W}_g^\top, \tag{4}$$