11-9-2011

LIKELIHOOD BASED POPULATION INDEPENDENT COMPONENT ANALYSIS

Ani Eloyan
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, aeloyan@jhsph.edu

Ciprian M. Crainiceanu
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Brian S. Caffo
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Likelihood Based Population Independent Component Analysis

Ani Eloyan, Ciprian M. Crainiceanu and Brian S. Caffo

October 11, 2011

Abstract

Independent component analysis (ICA) is a widely used technique for blind source separation, used heavily in several scientific research areas including acoustics, electrophysiology and functional neuroimaging. We propose a scalable two-stage iterative true group ICA methodology for analyzing population level fMRI data where the number of subjects is very large. The method is based on likelihood estimators of the underlying source densities and the mixing matrix. As opposed to many commonly used group ICA algorithms the proposed method does not require significant data reduction by a twofold singular value decomposition. In addition, the method can be applied to a large group of subjects since the memory requirements are not restrictive. The performance of our approach is compared with commonly used group ICA algorithms is shown by using simulation studies. Furthermore, the proposed method is applied to a large collection of resting state fMRI datasets. The results show that the postulated brain networks are recovered by the proposed algorithm.

Keywords: functional MRI, signal processing

1 Introduction

Independent component analysis (ICA, Jutten and Herault, 1991) is a source separation technique that assumes linear mixing and independent source signals. ICA is commonly used in a variety of fields, including: acoustics, electrophysiology and neuroimaging. Particularly, in resting state (rs) functional connectivity (fc) functional magnetic resonance imaging (fMRI), it has become the standard tool for discovery, exploration and modeling of brain networks. An active scientific discussion is ongoing as to the exact (patho-) physiological interpretation and importance of rs-fc-fMRI brain networks discovered via ICA. It is clear that large scale implementations of ICA will be an important component of resolving these issues. Large scale databases of resting state fMRI scans are becoming increasingly common; ideally allowing rs-fMRI to become a part of population-based research. As an example, the 1,000 Connectome Project combines scans from several sites resulting in a database of over 1,400 scans of healthy adults. In addition, the ADHD 200 dataset has resting state scans of roughly 200 attention deficit hyperactive children and 500 control children where some of the children have several scans resulting in over 1,000 fMRI scans. Moreover, the US National Institutes of Health has spearheaded the Human Connectome Project, a 30 million dollar venture to compile a comprehensive database of connectivity data, including rs-fMRI. In addition to addressing the importance of such large scale implementations of ICA to rs-fc-fMRI, our work generalizes to any high dimensional implementation of ICA.

ICA is an umbrella term that includes several different algorithmic implementations and theoretical foundations. At their core, the primary commonality of ICA algorithms is a linear factor analytic model (Harman, 1967) with the assumption of independence of underlying factors. We focus on the so-called noise-free ICA, a version of ICA that simply results in an "unmixing" (non-singular linear multiplication) of the input data

matrix. This results in including the measurement error or other noise in the data as a part of the independent components. The estimation of the linear un-mixing matrix involves iterative algorithms. A common starting point for all algorithms is a first stage singular value decomposition where dimension reduction is performed, after de-meaning, to avoid an overdetermined system. Thus, the data input to ICA have mean zero and are uncorrelated. Hence, Gaussian distributional assumptions provide little further insight to linear reorganizations. This motivates the search for solutions that are as non-Gaussian as possible. Hyvarinen et al. (2001) present an extensive overview of such algorithms. Most notably, Hyvarinen and Oja (1997) introduced a fast fixed point algorithm (called fastICA) for finding the independent signals by maximizing an approximation to the negentropy. Cardoso (1990) introduced the JADE algorithm which is based on cumulant tensors. The Bell-Sejnowski algorithm (Bell and Sejnowski, 1995) finds maximum likelihood (ML) estimates of the underlying independent signals.

As Bell and Sejnowski (1995), we focus on the ML implementations of ICA, which require a fully specified likelihood. Eloyan and Ghosh (2011a) discuss parameter identifiability in ICA and present a set of sufficient conditions that ensure model identifiability. This manuscript builds on their work, extending it to high dimensional applications, focusing on fMRI.

Calhoun et al. (2001a) introduced the use of ICA for group inferences of fMRI data. The proposed algorithm is based on reducing the dimensions of the original images by using principal component analysis (PCA) and then applying fastICA to obtain the underlying sources. In related work, Backmann and Smith (2005) presented tensorial extensions to ICA for group fMRI data analysis and Guo and Pagnoni (2008) provided an EM algorithm based ICA method for the case when there is a Gaussian noise term in the model.

We propose a two stage maximum likelihood algorithm for group independent component analysis applicable to datasets where the number of subjects is very large. The densities of the underlying sources are modeled using finite mixtures of smooth densities. The time courses for each subject are updated using an optimization algorithm. The method is based on iteratively updating the time courses of each subject in the group along with estimating the underlying densities of the independent components to obtain the individual time courses for each subject and the common spatial map.

Most methods developed for group ICA require high memory capacity, since the data are usually concatenated to obtain a full matrix consisting of the observed images from each subject. If the number of subjects is very large, then loading the concatenated matrix requires excessive memory and becomes implausible. A solution to the problem of dimensionality uses a two stage SVD approach for dimension reduction. In this approach, a first stage SVD is applied to the image for each subject, where a few vectors are retained before concatenating the matrices. Next, a second SVD is applied to the now concatenated (over subjects) spatial singular vectors and the first few spatial singular vectors are retained to force a determined linear system for the noise-free group ICA model. Notice that in this process the reduction of the dimension from having components for a group of subjects to one subject is done using the SVD. The ICA algorithm is applied to the resulting twice projected data to find the independent components. In this standard approach to group ICA, one SVD is required by the algorithm, while another is done purely for computational convenience. In the meantime, if the number of subjects is very large the concatenation of the matrices may not be computationally feasible, which limits direct applications of current group ICA methods to only a few subjects. Our methods are linear in the number of subjects and are scalable to high dimensional data because they require sequential access to subject-specific data instead of the entire group data matrix.

Our primary application data is the 1,000 Functional Connectomes Project hosted on the NITRC web site (http://www.nitrc.org/projects/fcon_1000). Biswal et al. (2010) describe the data collection and acquisition parameters from contributing sites and provide an analysis of functional connectivity. The data set contains structural and rs-fMRI images for 1,414 healthy adults collected independently from 35 centers worldwide. The time of repetition, total scan time and other experimental parameters differ across sites. Hence, we focus on subsets of the data that share the same scanning characteristics; the largest number of subjects considered was 150, but our methods are linear in the number of subjects and can easily be applied to thousands or tens of thousands of subjects. Such large groups of subjects cannot reasonably be handled by the commonly used group ICA algorithms. These are the reasons why we label our method "likelihood-based population ICA".

2 Methods

2.1 Likelihood Based Group Independent Component Analysis for fMRI

We present the model in generality. Suppose that for each subject, indexed by $i = 1, \ldots, I$, a $T \times V$ dimensional matrix $\boldsymbol{X}_i$ is observed. Here the columns, indexed by $v = 1, \ldots, V$, are the independent repetitions of each observed mixture while the rows, indexed by $t = 1, \ldots, T$, are the mixed signals. Contextually, v represents voxels in a fMRI series while t indexes the scans.

When necessary we use $\boldsymbol{X}_i(t, v)$ to represent row t, column v of $\boldsymbol{X}_i$ and apply the same convention to other vectors and matrices. We assume that a group ICA decomposition

implies the equation

$$\boldsymbol{X}_i(t,v) = \sum_{q=1}^{Q} \boldsymbol{A}_i(t,q)\boldsymbol{S}(q,v) \tag{1}$$

for all $i = 1, \ldots, I$. Model (1) assumes that the spatio-temporal process, $X_i(t,v)$, for each subject, i, can be decomposed into a finite sum of products between subject-specific time series, $A_i(t,q)$, and subject-independent spatial maps, $S(q,v)$. This equation is equivalent to $\boldsymbol{X}_i = \boldsymbol{A}_i\boldsymbol{S}$ or $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}$ where $\boldsymbol{X} = [\boldsymbol{X}_1^T \ldots \boldsymbol{X}_I^T]^T$ and $\boldsymbol{A} = [\boldsymbol{A}_1^T \ldots \boldsymbol{A}_I^T]^T$ are the $IT \times V$ and $IT \times Q$ matrices obtained by stacking the $\boldsymbol{X}_i$ and $\boldsymbol{A}_i$ respectively. Here the V dimensional vectors, $\boldsymbol{S}(q,\cdot)$, are the underlying independent components (assumed random over v) and $\boldsymbol{A}_i(t,\cdot)$ are the subject-specific fixed effect linear mixing vectors. In the context of fMRI, the $\boldsymbol{S}(q,\cdot)$ are spatial maps that are often interpreted as brain networks while the $\boldsymbol{A}_i$ are the subject specific temporal mixing matrices.

In order for the ICA model to be fully identifiable we assume that the square mixing matrices, $\boldsymbol{A}_i$, are of full rank and hence we define $\boldsymbol{W}_i = \boldsymbol{A}_i^{-1}$. We define the densities of the underlying sources as $f_1, \ldots, f_Q$. That is, we assume that $\{\boldsymbol{S}(q,v)\}_{v=1}^{V}$ is a collection of iid draws from f_q, for all $q = 1, \ldots, Q$.

Group ICA makes the parsimony assumption that the (random) independent components are common across subjects, while how they mix is a fixed effect that can differ among subjects. This is exactly a standard ICA model on the data concatenated across subjects to obtain an $IT \times V$ matrix. Allowing for separate independent components (IC) and mixing matrices across subjects is equivalent to simply applying ICA separately to each subject and is discussed here. Having separate ICs across subjects and a common mixing matrix is another possible parsimony assumption. This is analogous to an ICA model on the data having concatenated subjects to obtain a $T \times IV$ matrix.

In the context of single-subject fMRI, assuming spatial ICs and temporal fixed effect mixing matrices results in the so-called spatial ICA (sICA) (Calhoun et al., 2001b). Alter-

natively, assuming temporal ICs and spatial fixed effect mixing matrices is often referred to as temporal ICA (tICA). The assumptions in different fMRI experiments leading to the choice of using sICA or tICA have been widely discussed. For instance, Calhoun et al. (2001b) show different paradigm related fMRI experiments where the tICA may perform poorly compared with sICA when the assumptions on the temporal independence are violated. Concatenating to obtain an $IT \times V$ matrix and hence using the sICA, has been settled upon for group ICA analysis of rs-fMRI. For resting state data, only this variation of concatenation is sensible, since subjects are spatially co-registered into a common template space, whereas they are not temporally registered. In other words, time 1 for subject 1 is not the same as time 1 for subject 2. We develop our method for group sICA, however it can easily be modified to obtain the temporal ICA model if necessary.

There is a technical consideration in that (1) is overdetermined. Hence, we first preprocess the data at the subject level by centering and whitening the observed matrices via an SVD and retaining only the first Q components for each subject. Henceforth, we assume that the number of components to estimate is $Q = T$, i.e. the data are projected on the first Q singular vectors. Since we are assuming a noise-free model, noise in the data is absorbed into the estimated ICs and the mixing matrix.

2.2 Density Estimation in High Dimensions

In the early literature on maximum likelihood based estimation of the independent sources, well known distribution functions were used to model the f_q (Hyvarinen et al., 2001, p. 204). Boscolo et al. (2004) suggested using kernel density estimation to model the underlying densities non-parametrically. We take a similar approach by using mixture density estimates (MDE) introduced by Eloyan and Ghosh (2011b).

To elaborate, we parameterize the density as:

$$f_q(s) = \sum_{j=1}^{J_q} \theta_{qj} \frac{1}{\sigma_q} \phi\left(\frac{s - \mu_{qj}}{\sigma_q}\right), \tag{2}$$

where $\phi(\cdot)$ is the standard normal density function. Our treatment fixes the means, μ_{qj}, and variance, σ_q for each mixture density. We define the μ_{qj} to lie on a grid of not necessarily equidistant points. To illustrate, consider a strict density estimation problem where $\boldsymbol{s}(q, \cdot)$ are observed with empirical mean zero and variance one.

As a starting set of fixed parameters for the mixture densities define the number of densities in the mixture as $J_q = 1 + \frac{2}{3}\text{Range}_v\{\boldsymbol{S}(q, v)\}$. The set of fixed means is given by $\mu_{qj} = \min_v \boldsymbol{S}(q, v) + \frac{j-1}{J_q - 1}\text{Range}_v\{\boldsymbol{S}(q, v)\}$ for $j = 1, \ldots, J_q$, and the variance component $\sigma_q^2 = 2(\mu_{qV} - \mu_{q1})/\{3(J_q - 1)\}$. The rationale behind these choices is to set the μ_{qj} as an equally spaced grid between the extremes of the data and to set σ_q^2 such that $\sigma_q = o(1)$ as $J_q \to \infty$. Suppose $\mathcal{M}_{J_q} = \{\mu_{q1} < ... < \mu_{qJ_q}\}$ is the set of fixed means of the mixture densities. As the number of mixture densities J_q increases the set $\mathcal{M}_{J_q+1}$ is constructed by adding the median of one of the intervals $[\mu_{q,j}, \mu_{q,j-1}]$. Hence the sequence $\mathcal{M}_1, \mathcal{M}_2, \ldots$ maintains the sieve structure, i.e. each consecutive set contains the previous sets as subsets.

The weights of the mixture densities in (2) given by $(\theta_{q1}, \ldots, \theta_{qJq})$ are estimated using a constrained EM algorithm. The resulting density estimates satisfy the moment constraints required for full identifiability of the model given by

$$E[\boldsymbol{S}(q, \cdot)] = 0, \; E[\boldsymbol{S}(q, \cdot)^2] = 1, \text{ and} \tag{3}$$

$$0 < E[\boldsymbol{S}(1, \cdot)^3] < \ldots < E[\boldsymbol{S}(Q, \cdot)^3],$$

for $q = 1, \ldots, Q$.

The nonparametric estimation of the density of a vector $\boldsymbol{S}(q, \cdot)$, which has a large sample size ($\approx 70,000$ voxels in this case), can be computationally problematic. To

address this issue, we propose a binning algorithm for the density estimation, essentially looking at the approximation to the histogram of the data. Choose $p << V$ and suppose for each independent component, q, the set $(r_{q1}, r_{q2}, \ldots, r_{qp}, r_{q,p+1})$ consists of the $p+1$ quantiles of $\boldsymbol{S}(q, \cdot)$. Next, bins, $B_{q1}, B_{q2}, \ldots, B_{qp}$, are constructed using the above quantiles as the endpoints for the bins. In addition, let $\boldsymbol{c}_q = (c_{q1}, c_{q2}, \ldots, c_{qp})$ denote the counts of the observations in each of the bins. The underlying density of the components can be found by using the midpoints, say $\boldsymbol{M}_q = (M_{q1}, M_{q2}, \ldots, M_{qp})$, of the bins and the counts by slightly modifying the proposed constrained EM algorithm. In other words, the updates of the mixture weights are computed as

$$\widehat{\theta}_{qj}^{(k+1)} = \frac{\sum_{i=1}^{n} w_{ij}(\boldsymbol{\theta}_q^{(k)}, \boldsymbol{c}_q, \boldsymbol{M}_q)}{\widehat{\lambda}_1 + \widehat{\lambda}_2 \mu_{qj} + \widehat{\lambda}_3 \mu_{qj}^2}. \tag{4}$$

where the Lagrange multipliers are computed by the following system of equations

$$\begin{aligned}
\sum_{j=1}^{N} \frac{\sum_{i=1}^{n} w_{ij}(\boldsymbol{\theta}_q^{(k)}, \boldsymbol{c}_q, \boldsymbol{M}_q)}{\lambda_1 + \lambda_2 \mu_{qj} + \lambda_3 \mu_{qj}^2} &= 1 \\
\sum_{j=1}^{N} \frac{\sum_{i=1}^{n} w_{ij}(\boldsymbol{\theta}_q^{(k)}, \boldsymbol{c}_q, \boldsymbol{M}_q)\mu_{qj}}{\lambda_1 + \lambda_2 \mu_{qj} + \lambda_3 \mu_{qj}^2} &= 0 \\
\sum_{j=1}^{N} \frac{\sum_{i=1}^{n} w_{ij}(\boldsymbol{\theta}_q^{(k)}, \boldsymbol{c}_q, \boldsymbol{M}_q)\mu_{qj}^2}{\lambda_1 + \lambda_2 \mu_{qj} + \lambda_3 \mu_{qj}^2} &= 1 - \sigma_q^2,
\end{aligned} \tag{5}$$

where

$$w_{ij}(\boldsymbol{\theta}_q^{(k)}, \boldsymbol{c}_q, \boldsymbol{M}_q) = \frac{\theta_{qj}^{(k)} \phi([M_{qi} - \mu_{qj}]/\sigma_q) c_{qi}}{\sum_{j=1}^{J_q} \theta_{qj}^{(k)} \phi([M_{qi} - \mu_{qj}]/\sigma_q) c_{qi}}.$$

see Eloyan and Ghosh (2011b) for more details on the construction of the constrained EM algorithm.

The estimated densities maintain the constraints on the moments of the densities (mean and variance). By (3) it is stated that the independent components are ordered so that the third moments are in increasing order. This is necessary to avoid label switching issues, as the model is invariant to permutations of the independent components (provided

the mixing matrix is permuted correspondingly). The ordered third moment assumption is a straightforward fix to this issue. Thus, within each iteration, the independent components are permuted to preserve the order of the estimated third moments.

2.3 Semiparametric Iterative Algorithm for Group ICA

Based on the density estimation algorithm described in Section 2.2 we develop a semiparametric iterative algorithm for true group independent component analysis model (1).

By the independence of the underlying sources $\boldsymbol{S}(q,\cdot)$ the likelihood of the unmixing matrix $\boldsymbol{W} = [\boldsymbol{W}_1 \ \ldots \ \boldsymbol{W}_I]$ is given by

$$L(\boldsymbol{W}, \boldsymbol{f}) = \prod_{i=1}^{I} \prod_{v=1}^{V} \prod_{q=1}^{Q} f_q \left(\sum_{l=1}^{Q} w_{iql} x_{ilv} \right) |\det \boldsymbol{W}_i|. \tag{6}$$

Assuming that the densities of the underlying sources are estimated using the finite mixtures of Gaussian densities defined as $\hat{f}_1, \ldots, \hat{f}_Q$, the likelihood function of the unmixing matrix $\boldsymbol{W}$ can be constructed analytically as

$$L(\boldsymbol{W}, \widehat{\boldsymbol{f}}) = \sum_{i=1}^{I} \left\{ \sum_{v=1}^{V} \sum_{q=1}^{Q} \log \left[\widehat{f}_q \left(\sum_{l=1}^{Q} w_{iql} x_{ilv} \right) \right] + V \log |\det \boldsymbol{W}_i| . \right\}, \tag{7}$$

where

$$\widehat{f}_q(s) = \sum_{j=1}^{J_q} \widehat{\theta}_{qj} \phi \left(\frac{s - \mu_{qj}}{\sigma_q} \right) \frac{1}{\sigma_q}.$$

Notice that by construction of the densities of the underlying sources the derivative and Hessian matrices of the loglikelihood can also be found analytically.

We need to obtain starting values for the unmixing matrices to start the algorithm. This can be done by choosing the subject specific unmixing matrices given by $\widehat{\boldsymbol{W}_i}^{(0)}$. Alternatively, we can find starting values that satisfy the condition that the underlying independent components are the same for all subjects. We can use the population value decomposition (Crainiceanu et al., 2010) of the full matrix $\boldsymbol{X} = [\boldsymbol{X}_1^T, \ldots, \boldsymbol{X}_I^T]^T$ given by

$$\boldsymbol{X} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^T \tag{8}$$

The starting values of the $\boldsymbol{W}_i$ then are chosen as the *ith* block of the rows of $\boldsymbol{U\Sigma}$.

Table 1: *HDICA algorithm.*

	For $M \in \{1, 2, \ldots\}$
1.	Let $\boldsymbol{S}_i^{(M)} = \boldsymbol{W}_i^{(M-1)} \boldsymbol{X}_i$, for each $i = 1, \ldots, I$
2.	For each IC q construct the set of midpoints $M_{q1}, M_{q2}, \ldots, M_{qp}$ of the bins
	and the corresponding counts $c_{q1}, c_{q2}, \ldots, c_{qp}$.
3.	For each $q = 1, \ldots, Q$ construct the set of means $\mathcal{M}_{J_q^{(M)}} \supseteq \mathcal{M}_{J_q^{(M-1)}}$.
	and the variance component σ_q.
4.	By using MDE estimate $(\theta_{q1}^{(M)}, \ldots, \theta_{qJ_q^{(M)}}^{(M)})$.
5.	For each $i = 1, \ldots, I$ compute the gradient $L'(\widehat{\boldsymbol{W}_i}^{(M)})$ and hessian matrix $L''(\widehat{\boldsymbol{W}_i}^{(M)})$.
6.	For each $i = 1, \ldots, I$ update the unmixing matrix
	$\widehat{\boldsymbol{W}_i}^{(M+1)} = \widehat{\boldsymbol{W}_i}^{(M)} - L''(\widehat{\boldsymbol{W}_i}^{(M)})^{-1} L'(\widehat{\boldsymbol{W}_i}^{(M)})$.
7.	$\delta = \max \lvert \widehat{\boldsymbol{W}_i}^{(M+1)} - \widehat{\boldsymbol{W}_i}^{(M)} \rvert$. If $\delta > \epsilon$ return to step 1

Even though the derivative and Hessian matrices of the loglikelihood can be computed analytically we can also use an approximation to the Hessian given by $L''(w) \approx L'(w)L'(w)^T$ to have a more robust optimization algorithm.

The iterative algorithm for finding the maximum likelihood estimate of the unmixing matrices $\boldsymbol{W}_1, ..., \boldsymbol{W}_{N_s}$ is given in Table 1. One of the striking differences of our method compared with other group ICA algorithms is that at each iteration only one $Q \times V$ dimensional subject-specific matrix is loaded in memory to compute the update for $\boldsymbol{W}$ matrix. In addition, the densities are estimated using a binning algorithm hence the increase in the sample size does not affect the speed of the density estimation part of the algorithm.

3 Results

3.1 Simulation Results

In order to illustrate the performance of the proposed method we conducted simulation studies where data were generated using distributions of different shapes. Three different cases were used in the study. The results are compared with the commonly used group ICA algorithm by Calhoun et al. (2001a) which is based on fastICA (Hyvarinen and Oja, 1997).

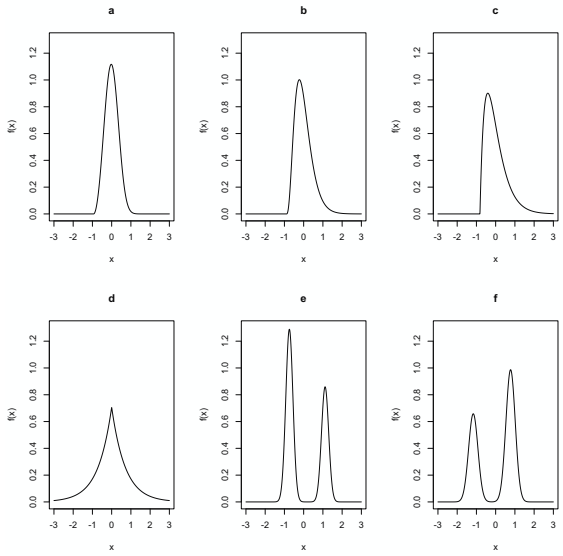

Figure 1: The distributions used to generate data for the underlying sources.

Suppose the number of subjects is $N_s = 3$. First suppose that the number of underlying sources $Q = 2$. The data are generated by the ICA model $\boldsymbol{X}_i = \boldsymbol{A}_i\boldsymbol{S}$ with $T = 2$ and $V = 2000$. We further assume that

$$A_1 = \begin{pmatrix} 0.75 & 0.25 \\ 0.5 & -0.5 \end{pmatrix}, A_2 = \begin{pmatrix} 1 & 0 \\ 0.5 & -0.5 \end{pmatrix}, A_3 = \begin{pmatrix} 1 & 0.5 \\ 0.75 & 1 \end{pmatrix}$$

with $\det(A_1) = -0.5$, $\det(A_2) = -0.5$ and $\det(A_3) = 0.625$. The shapes of the densities used for as the densities for independent components are shown in Figure 1. For each of the densities two independent components are generated and the individual mixture matrices are constructed by using A_1, A_2, A_3 above.

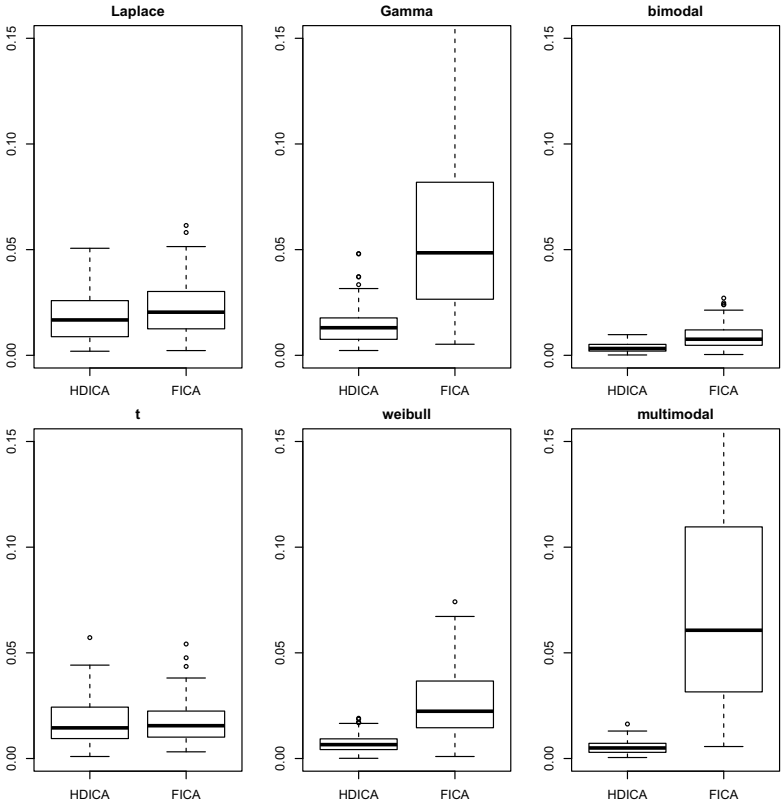

Figure 2: The Amari errors of the estimated $\boldsymbol{W}$ matrix using the proposed Iterative Group ICA and Fast ICA. The sources are generated using Weibull densities.

The Amari error (introduced by Amari, 1998) is used for evaluating the performance of the ICA methods. It is given by

$$AE(A, \widehat{W}) = \frac{1}{2Q} \sum_{i=1}^{Q} \left(\sum_{j=1}^{Q} \frac{|p_{ij}|}{max_k |p_{ik}|} - 1 \right) + \frac{1}{2Q} \sum_{j=1}^{Q} \left(\sum_{i=1}^{Q} \frac{|p_{ij}|}{max_k |p_{kj}|} - 1 \right),$$

where $P = A\widehat{W}$ and $\widehat{W}$ is the estimated unmixing matrix for each subject by each method. The Amari error: is invariant to matrix permutations, sign changes, and is a value between

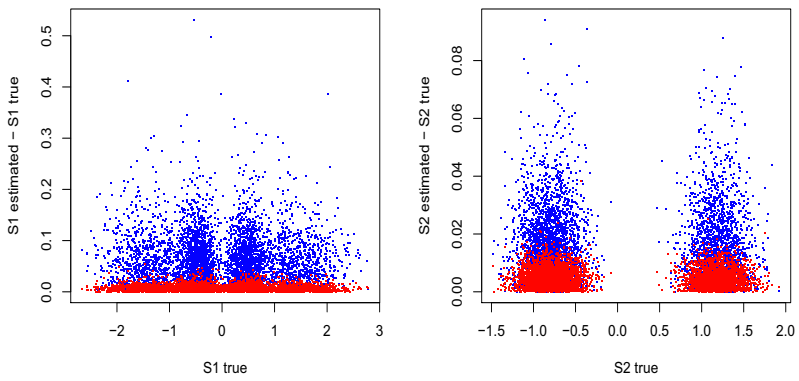

Figure 3: The scatterplots showing the difference of the estimated components from the true data in the simulation study with $m = 8$. The x-axis shows the true component and the y-axis shows the absolute difference of the true component with the estimated component (by HDICA (red) and by fastICA (blue)).

0 and $Q - 1$.

The boxplots of the Amari errors are shown in Figure 2. When the distributions of the underlying sources are symmetric and unimodal (t, Laplace) we observe that our method is competitive with fastICA. However, for the other distributions our method seriously outperforms fastICA in terms of minimizing the Amari error.

The scatterplots in Figure 3 show the absolute difference of the true independent component with the estimates found by using the HDICA (in red) and fastICA (in blue). Two of the components generated using Gamma (left) and multimodal densities (right)

are plotted. As can be seen from this figure the independent components are almost completely recovered by HDICA. In other words the absolute difference of the estimate by HDICA with the truth is less variable than the absolute difference of the estimate found by fastICA with the truth.

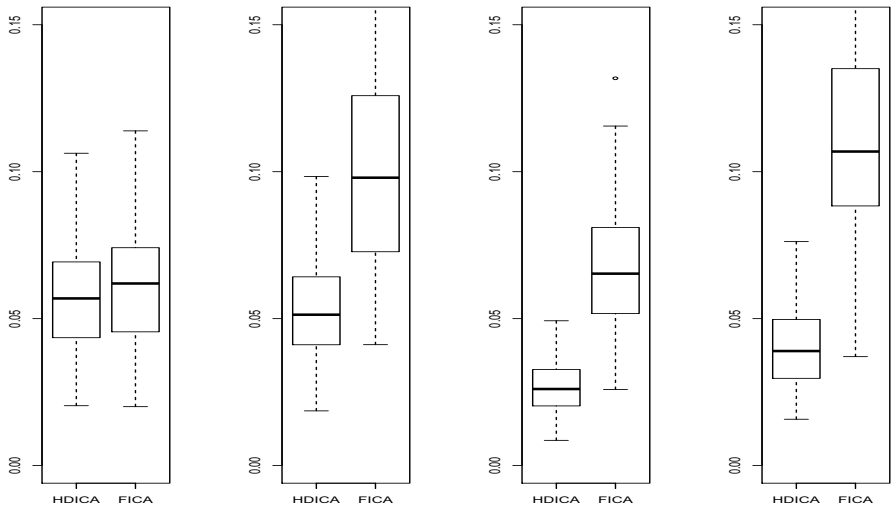

Figure 4: The Amari errors of the estimated $\boldsymbol{W}$ matrix using the proposed Iterative Group ICA and Fast ICA. The sources are generated using Weibull densities.

For the second scenario suppose that the number of underlying sources is $Q = 4$ with $V = 5000$ voxels. Again the number of subjects is $N_s = 3$. The mixing matrices used are given by:

$$A_1 = \begin{pmatrix} 2 & 1 & 2 & 3 \\ 3 & 3 & 1 & 0.5 \\ 1 & 2 & 2 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 2 & 3 & 2 & 1 \\ 3 & 4 & 1 & 0.5 \\ 3 & 2 & 3 & 4 \\ 2 & 3 & 3 & 1 \end{pmatrix}, A_3 = \begin{pmatrix} 1 & 2 & 2 & 1 \\ 3 & 4 & 1 & 0.5 \\ 3 & -1 & 3 & 4 \\ 2 & 1 & 3 & 1 \end{pmatrix},$$

with $\det(A_1) = -8$, $\det(A_2) = 7.5$ and $\det(A_3) = 47$. The shapes of the densities used

for generating the values of the independent components $S_k, k = 1, 2, \ldots, m$ are plotted in Figure 1. Three different cases are considered as shown in Table 2.

	IC 1	IC 2	IC 3	IC 4
a.	$Laplace(\mu, \sigma)$	$Laplace(\mu, \sigma)$	$Laplace(\mu, \sigma)$	$Laplace(\mu, \sigma)$
b.	$Gamma(a, b)$	$Gamma(a, b)$	$Laplace(\mu, \sigma)$	$Laplace(\mu, \sigma)$
c.	$Gamma(a, b)$	$Gamma(a, b)$	$MixNorm(\mu, \sigma)$	$MixNorm(\mu, \sigma)$
d.	$t(df)$	$Gamma(a, b)$	$Weibull(a, b)$	$MixNorm(\mu, \sigma)$

Table 2: The choice of the independent components in simulations.

The boxplots of the Amari Errors are presented in Figure 4. Here again, the performance of HDICA is comparable to that of fastICA using the Amari Error criterion. In Case a., where the true underlying densities have Laplace distribution the results obtained by HDICA are similar to that of fastICA. As has been observed before when the underlying sources have nonsymmetric or bimodal densities as in b.-d. HDICA is outperforming fastICA in minimizing the Amari Error.

Finally, we illustrate a simulation study using $Q = 8$ components with $V = 5000$ voxels. All of the distributions given in Figure 1. The Amari errors for this case are shown in Figure 5. Here again our method performs significantly better than fastICA.

The simulations suggests that the proposed method, including data reduction steps that remove information, resulting in an algorithm that is comparably as good as the most popular group ICA algorithm. More importantly, however, is that the HDICA algorithm is immediately scalable to hundreds or thousands of subjects.

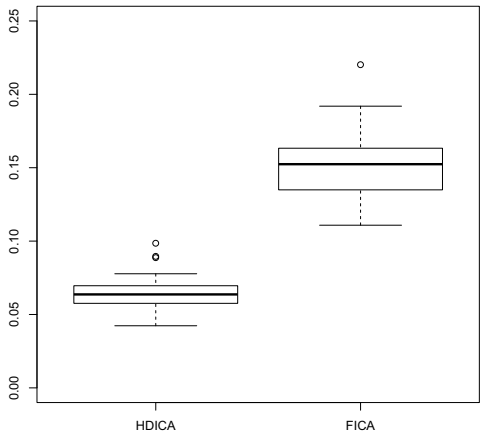

Figure 5: The Amari errors of the estimated $\boldsymbol{W}$ matrix using the proposed Iterative Group ICA and Fast ICA. The sources are generated using Weibull densities.

3.2 Application to the 1,000 Functional Connectomes fMRI Dataset

To illustrate our algorithm we use one of the freely available large fMRI datasets. The 1000 Functional Connectomes Project dataset (Biswal et al., 2010) is aimed at discovery science of brain function. As discussed in Section 1, applying the current group ICA methods to a large number of subjects is computationally impossible for regular computers. As discussed, the motivation of this paper is to develop a group ICA algorithm that can be applied to use for any number of subjects without excessive data reduction steps. Because there are several covariates in these data we illustrate our method for a subset of 50 subjects from the nitrc collection. Again our method does not require the concatenation of the subject specific data matrices. Moreover, at each step of the algorithm only one of the subject specific data matrices is loaded into the memory. Hence, the HDICA algorithm is linear in the number of subjects and can be scaled up relatively easily.

The scans are collected when the subjects are in resting state. Each dataset is a 4D

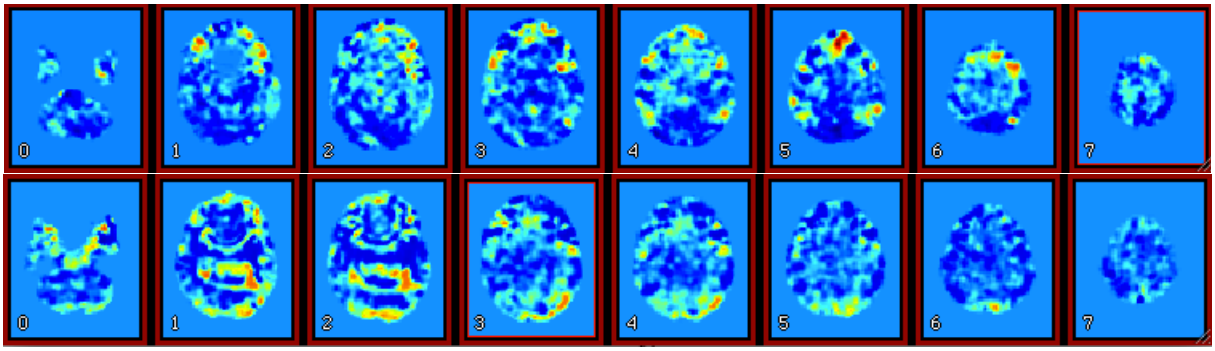

Figure 6: Temporal snapshot of the data in 7 slices from the fMRI images for 2 subjects (subject 1 is shown in the top panels and subject 2 in the bottom panels).

array of intensities. Each subject was in the scanner for 2.2-20 minutes. For the subset used in this analysis the number of time points was $T = 123$. Notice that even if the number of time points varies among the subjects the algorithm can still be applied, since the first PCA step will reduce the dimensions of the datasets to the same Q. The scans are collected using a 3T scanner. Selected fMRI scans of the brain from two subjects are shown in Figure 6 to illustrate the structure of the data. Standard image processing was performed to register the data to the MNI standard brain space. However, no smoothing is done on the data before applying group ICA.

The MNI152_T1_3mm_brain_mask.nii mask was used to extract the background of the images and obtain the voxels that are in the actual brain. For each time point, the 3D array is vectorized to obtain a V dimensional vector of intensities that are then concatenated over time. Hence we obtain a $T \times V$ dimensional matrix $\boldsymbol{X}_i$ for each subject. The HDICA algorithm is then applied using these $\boldsymbol{X}_i$ matrices. The algorithm is applied without smoothing the data first.

We first chose a subset of $I = 50$ subjects to compare the performance of HDICA and fastICA. Following Biswal et al. (2010) we use group ICA to obtain $m = 20$ components.

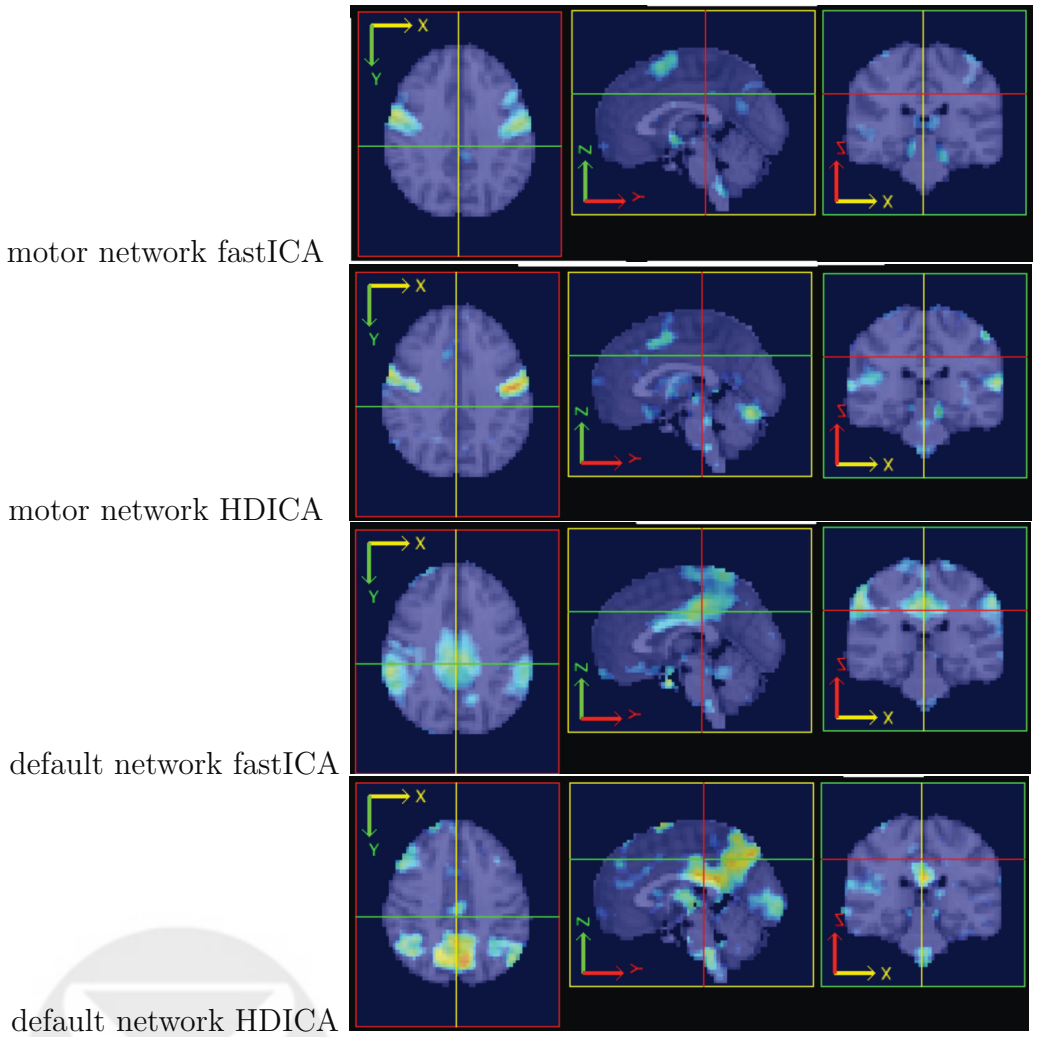

Figure 7: Motor (top two) and default (bottom two) networks computed for $I = 50$ subjects using fastICA and HDICA.

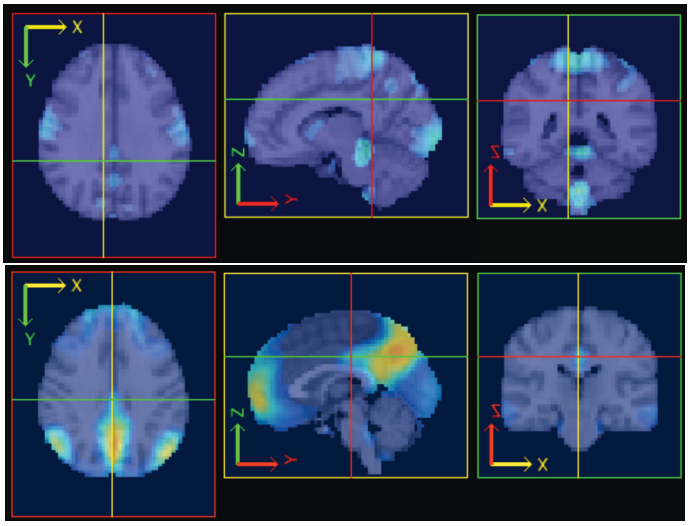

Figure 8: Motor (top) and default (bottom) networks computed for $I = 150$ subjects using the proposed HDICA algorithm.

The slices of two independent components overlayed on the brain template are shown in Figure 7. The first two rows correspond to the motor network estimated by fastICA (above) and HDICA (below). The third and forth rows show the default brain network.

Finally, HDICA was used to compute $m = 20$ independent components by using fMRI data of $I = 150$ subjects from Cambridge site (a subset of the nitrc data). Figure 8 shows two of the independent components overlayed on the brain template. Again the motor network and the default network are found by HDICA.

4 Discussion

In this paper we present a group ICA algorithm based on nonparametric estimation of the densities of the underlying sources using finite mixtures of continuous densities. The mixing matrix is simultaneously estimated using an iterative optimization algorithm. The proposed algorithm is scalable to large datasets. As a byproduct of the algorithm we

obtain the estimates of the densities of the underlying spatial maps.

We first develop a density estimation method based on binning the data and using a mixture of continuous densities for approximating the histogram of the data. The density estimates are computed using a constrained EM algorithm to satisfy moment constraints for identifiability of the model. The estimated densities are then used to model the distributions of the underlying spatial maps.

The performance of the proposed algorithm is presented by simulation studies showing that our method performs at least as well as other commonly used methods. The algorithm was applied to a set of resting state fMRI data. The method can be used for large groups if fMRI data in different subpopulations to obtain the brain networks and study differences within the subpopulations.

References

S Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2): 251–276, 1998.

C.F. Backmann and S. M. Smith. Tensorial extensions of independent component analysis for multisubject fmri analysis. *Neuroimage*, 25:295–311, 2005.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind seperation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

Bharat B Biswal, Maarten Mennes, Xi-Nian Zuo, Suril Gohel, Clare Kelly, Steve M Smith, Christian F Beckmann, Jonathan S Adelstein, Randy L Buckner, Stan Colcombe, Anne-Marie Dogonowski, Monique Ernst, Damien Fair, Michelle Hampson, Matthew J Hoptman, James S Hyde, Vesa J Kiviniemi, Rolf Kötter, Shi-Jiang Li, Ching-Po Lin, Mark J Lowe, Clare Mackay, David J Madden, Kristoffer H Madsen, Daniel S Margulies, Helen S Mayberg, Katie McMahon, Christopher S Monk, Stewart H Mostofsky, Bonnie J Nagel, James J Pekar, Scott J Peltier, Steven E Petersen, Valentin Riedl, Serge A R B Rombouts, Bart Rypma, Bradley L Schlaggar, Sein Schmidt, Rachael D Seidler, Greg J Siegle, Christian Sorg, Gao-Jun Teng, Juha Veijola, Arno Villringer, Martin Walter, Lihong Wang, Xu-Chu Weng, Susan Whitfield-Gabrieli, Peter Williamson, Christian Windischberger, Yu-Feng Zang, Hong-Ying Zhang, F Xavier Castellanos, and Michael P Milham. Toward discovery science of human brain function. *Proc Natl Acad Sci U S A*, 107(10):4734–9, 2010.

R. Boscolo, H. Pan, and V.P. Roychowdhury. Independent component analysis based on nonparametric density estimation. *Neural Networks, IEEE Transactions on*, 15(1):55 –65, 2004.

V. D. Calhoun, T. Adali, G.D. Pearlson, and J. J. Pekar. A method for making group inferences from functional mri data using independent component analysis. *Human Brain Mapping*, 14:140–151, 2001a.

V. D. Calhoun, T. Adali, G.D. Pearlson, and J.J. Pekar. Spatial and temporal independent component analysis of functional mri data containing a pair of task-related waveforms. *Human Brain Mapping*, 13:43–53, 2001b.

J.-F. Cardoso. Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'90), Abuquerque, New Mexico*, pages 2655–2658, 1990.

C. M. Crainiceanu, B. S. Caffo, S. Luo, and V. Zipunnikov. Population value decomposition, a framework for the analysis of image populations. *Johns Hopkins University, Dept. of Biostatistics Working Papers*, 2010.

A. Eloyan and S. K. Ghosh. A semiparametric approach to source separation using independent component analysis. *Technical Report 2635, Department of Statistics, NCSU*, 2011a.

A. Eloyan and S. K. Ghosh. Smooth density estimation with moment constraints using mixture distributions. *Journal of nonparametric statistics*, 23(2):513–531, 2011b.

Y. Guo and G. Pagnoni. A unified framework for group independent component analysis for multi-subject fmri data. *Neuroimage*, 42:1078–1093, 2008.

H. H. Harman. *Modern Factor Analysis*. University of Chicago Press, 2nd edition, 1967.

A. Hyvarinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, New York, 2001.

C. Jutten and J. Herault. Blind seperation of sources, part i: An adaptive algorithm based on neuromimethic architecture. *Signal Processing*, 24:1–10, 1991.