ABSTRACT

Analysis of Resting-State fMRI using Pearson-VII Mixture Modeling

True Price

Director: Erich Baker, Ph.D.

In the past two decades, functional magnetic resonance imaging (fMRI) has been widely used to research and characterize neural activity in the brain based on measuring the hemodynamic response correlated to neuronal firing. A main goal of fMRI analysis is to characterize functional connectivity—correlation in activation pattern—between different regions of the brain for different task- or disease-related events. Here, we present a novel, data-driven method for identifying functionally connected regions of the brain using a Pearson-VII mixture model learning algorithm (p7-means). This method is complementary to independent component analysis (ICA), which derives underlying source signals, rather than probabilistic distributions, to characterize brain components. The p7-means algorithm is powerful in its ability to model a range of leptokurtic to Gaussian components, which makes it robust in identifying core functional components in noisy images. Additionally, p7-means has the advantage of learning the number of components from the data set, rather than returning a fixed number of components matching dimensionality. We apply the algorithm to resting-state monkey fMRI and compare the discovered components to those found in ICA. Correlational analysis shows consistent activation components between the two methods, although p7-means appears to result in more spatially localized groups.

APPROVED BY DIRECTOR OF HONORS THESIS:

Dr. Erich Baker, Department of Computer Science

APPROVED BY THE HONORS PROGRAM:

Dr. Andrew Wisely, Director

DATE: _____

ANALYSIS OF RESTING-STATE FMRI USING PEARSON-VII MIXTURE MODELING

A Thesis Submitted to the Faculty of

Baylor University

in Partial Fulfillment of the Requirements for the

Honors Program

By

True Price

Waco, Texas

May 2013

TABLE OF CONTENTS

LIST OF FIGURES iv

ACKNOWLEDGMENTS vi

1 Introduction 1

1.1 fMRI Analysis 2

1.2 Functional and Effective Connectivity 4

1.3 Methods of analysis 5

1.3.1 The General Linear Model 5

1.3.2 Hypothesis-Driven Techniques 6

1.3.3 Data-Driven Analysis 8

1.4 Principle Component Analysis 8

1.5 Independent Component Analysis 9

2 p7-Means Algorithm 12

2.1 The Pearson VII Distribution 12

2.2 Pearson VII EM Algorithm 14

2.2.1 E-step 14

2.2.2 M-step 16

2.3 The p7-means Algorithm 17

3 Experimental Results 21

3.1 Image Acquisition and Pre-processing 21

3.2 Probabilistic Components 22

3.3 Comparison with ICA 23

4 Discussion and Conclusion 31

BIBLIOGRAPHY 36

LIST OF FIGURES

2.1 Pearson-VII type distributions with unit variance for different values of m. 13

2.2 Performance of p7-means on a simulated data set of 6000 points, 5000 coming from random Gaussian distributions, and one consisting of a uniform distribution. 20

3.1 Result of the p7-means algorithm on a smoothed, resting-state monkey fMRI image (Scan #1). Using PCA dimensionality reduction to 20 dimensions, the p7-means algorithm discovers 34 generative components for this scan. Each image shows four of these components, in no specific order, with a probability threshold cutoff of $p(\mathbf{t} \mid \Theta_k) > 0.5$. 25

3.2 Result of the p7-means algorithm on a smoothed, resting-state monkey fMRI image (Scan #2). Using PCA dimensionality reduction to 20 dimensions, the p7-means algorithm discovers 23 generative components for this scan. Each image shows four of these components, in no specific order, with a probability threshold cutoff of $p(\mathbf{t} \mid \Theta_k) > 0.5$. 26

3.3 Result of the p7-means algorithm on a smoothed, resting-state monkey fMRI image (Scan #3). Using PCA dimensionality reduction to 20 dimensions, the p7-means algorithm discovers 33 generative components for this scan. Each image shows four of these components, in no specific order, with a probability threshold cutoff of $p(\mathbf{t} \mid \Theta_k) > 0.5$. 27

3.4 Probability maps from Melodic ICA components after PCA dimensionality reduction to 20 dimensions. Each image shows two of these components, with any overlap indicated in purple, using a threshold cutoff of $p(\mathbf{t} \mid \Theta_k) > 0.5$. 28

3.5 Pearson's r scores between p7-means and MELODIC ICA probability maps. The components (each row and column) are given in no specified order. Qualitatively, we see that nearly every component from p7-means corresponds at least somewhat with a component from ICA, and vice-versa. 29

3.6 Four paired ICA (red) and p7-means (blue) components having the highest Pearson's r scores. Areas of overlap between the components appear as purple (threshold: $p(\mathbf{t} \mid \Theta_k) > 0.5$). r values are: (a) 0.6, (b) 0.458, (c) 0.445, and (d) 0.428. 30

ACKNOWLEDGMENTS

The time it takes to write, and much less read, a set of acknowledgements is a gross injustice to the formidable influence and importance of the people listed therein. To sum up my career at Baylor in a few pithy "thank you's" is to omit a great many people who invested in me and to largely ignore four incredible years of growth, both academic and personal, as well as probably a few too many work-filled nights. In the interest of showing appreciation for Baylor and the Honors College, however, I will merely say that I would not have traded my undergraduate experience for anything.

My first expression of gratitude must go to my thesis mentor and academic advisor, Dr. Erich Baker. If nothing else, Dr. Baker has shown me how to approach academic research without losing the ability to personally connect with those around me. He is an incredible mentor who is always willing to help, instruct, and importantly, be a friend even while being professional. I am indebted to him for how far I have come in these last four years.

I thank Dr. Young-Rae Cho for many of the same reasons; his mentorship and belief in my ability has helped me become a confident academic. Thanks to Dr. Greg Hamerly for his excellent teaching ability, and for helping me get started with pg-means. To Dr. Michael Poor, thank you for sitting on my thesis committee, and to Dr. Jeff Donahoo, thank you for providing the template format for this document.

Many thanks to Kathleen Grant and Christopher Kroenke at the Oregon National Primate Research Center, who provided the monkey fMRI scans for this project. The scans were funded under grant number AA13510 (K. Grant).

Finally, I could not completed this thesis without my family and friends. Mom, Dad, and Joanna, thank you for loving me so unconditionally and for supporting me in every way. Jared and Austin, thanks for being such excellent roommates. And of course, thank you, Luci, for being the best and only companion for my life.

CHAPTER ONE

Introduction

The past few decades have seen remarkable growth in the field of medical imaging. Our understanding of the human body is being continuously updated as novel, more sensitive imaging methods are developed. As the quality and scope of imaging acquisition techniques increases, however, new methods for extracting meaningful information from the data must also be developed. This is especially true when analyzing the activity of functionally complex and widely varying biological structures such as the human brain. Neuroimaging is a fascinating field of research because it extends not only to structural analysis in the localized space of the skull, but to phenotypic traits, such as social behavior and general body functions, as well. The introduction of medical imaging to neuroscience research is revolutionizing our perspective on topics like cognition and neurological disease that previously could only be evaluated on a qualitative level. Neuroimaging analysis techniques seek to understand how regions of the brain differ in functionality, the patterns of activity these regions exhibit in different conditions, and how these regions interact to drive large-system biological function and form a complex, self-cognizant individual.

The purpose of this honors thesis is to demonstrate a new computational process, referred to as p7-means, for the functional segmentation of brain activity with specific application to fMRI. This method deviates from current analysis methods by using a Pearson VII mixture model to describe functional components, rather than characterizing such regions using signal processing techniques. In applying Pearson VII distributions, which are capable of accommodating outlier-heavy data sets in their modeling, the approach provides robust performance against sub-Gaussian noise factors inherent in fMRI scans. Effectively, the mixture model is a method for functional

segmentation of (not necessarily contiguous) regions across the entire brain that is consistent for a given neurological state. p7-means has implications in many areas of neuroimaging research, including between-subject registration, structural segmentation, simulation, and disease-state analysis.

1.1 fMRI Analysis

Functional magnetic resonance imaging (fMRI) is an important imaging technique that has greatly influenced neuroscience research since its first use in the early 1990s. This method iteratively applies the magnetic resonance techniques of MRI to collect a spatial-temporal signal mapping of the Blood-Oxygenation Level Dependency (BOLD) signal in the brain. In the resulting image, the scan is discretized into three-dimensional positional units called voxels, usually on the order of one to three cubic millimeters, with each voxel storing a vector of BOLD signal fluctuations in over time. In neuroscience theory, variations in neural activity are trailed by changes in the flow of oxygenated blood — known as the hemodynamic response — by one to two seconds. As such, the BOLD signal measure in a given voxel can be interpreted as a time-delayed observation of the firing of neurons contained within the space of that voxel. On the scale of the entire neural system, the patterns of voxel activation over time give an indication of how the brain performs certain tasks, such as observing an image, tapping a finger, or just being in an unstimulated, resting state.

The process by which fMRI translates brain activity is into a volumetric time series is complex and outside the scope of this paper. However, in the interest of providing an insight into the type of data fMRI collects, a short overview follows. MRI technology is based on the quantum mechanical property of atomic nuclei spin (Logothetis 2008; Huettel 2009), which may be thought of as the angular rotation and orientation of an elementary particle. When spin-$\frac{1}{2}$ protons, such as the hydrogen atoms found in water molecules, are placed in the presence of an external magnetic

field, the spin orientations begin to align parallel to the field. The orientations are said to *precess* around the magnetic field vector, rotating in a circle like a spinning top. Application of a radiofrequency field rotating at a frequency equal to the frequency of precession then causes the magnetization of the spinning protons to point perpendicular to the orientation of the magnetic field. After this excitation in the transverse plane, the protons return to their parallel configuration in a process known as *relaxation*. There are three different time constants of relaxation corresponding to different types of environments for the protons: T_1, which is the time recovery for spins returning to a low-energy (non-perpendicular) state; T_2, which characterizes the rate of magnetic resonance signal decay due to spin-spin interactions of excited protons; and T_2^*, which is a combined measurement of the T_2 decay and loss of signal coherency due to fluxuations in the applied magnetic field. These relaxation times are often used as measures of MRI contrast; for example, MRI T_1-weighted scans are able to highlight the different relaxation rates of grey and white matter, generating structural contrast. In fMRI, T_2^* relaxation is typically used to detect subtle variations in blood oxygenation levels. The end result of an MRI or fMRI image, therefore, is a formation of the magnetic gradient, or spatially varying magnetic field (Huettel 2009). The magnitude of this field directly correlates, with some factor of noise, to the amount of oxygenated blood in a given voxel.

fMRI has found widespread use in neuroscientific research because of several advantages over other well-known methods. First, fMRI is capable of measuring deep-brain activity, unlike electroencephalograph techniques, which are only able to measure signal changes around the neural cortex. Second, unlike MRI analysis, fMRI is able to detect the temporal, rather than merely structural, origins of blood flow gradients that contribute to the hemodynamic response for a given activity. Finally, fMRI is a non-invasive procedure unlike positron emission tomography, which uses a radioactive metabolite to detect blood flow in the brain.

1.2 Functional and Effective Connectivity

Historically, the principle of functional segregation has been a popular mechanism for understanding task-related neural activity (Friston 1994; Stephan, Riera, Deco, and Horwitz 2008). This concept is derivative of the practice of mapping brain regions to associated roles – for example, the amygdala controlling emotional response or the hippocampus being a center for memory storage. While the idea that brain activation organizes into structure-function associations is undoubtedly important for understanding neural activation at a high level, at best it is only a partial model for explaining event-driven brain function (Stephan, Riera, Deco, and Horwitz 2008). In practice, observed brain function follows a causal network model.

The idea of perception as a process illustrates this concept: First, some external stimulus causes a localized area of the brain (e.g. the visual coretex) to activate. In turn, the neurally-encoded interpretation of the stimulus is transmitted to other areas of the brain for processing via neural connections. Each unit of brain activity is sequential, and the overall function being performed is determined by the activity of several distinct regions simultaneously processing the information transmitted by one or several source regions. As the capabilities of neuroimaging and subsequent neuronal modeling have increased, explicitly system-based, information-theoretical approaches have become the focus of neuroscience research. Current methods in neuroscience seek to answer the question of functional integration, that is, interpreting brain function as an emergent property of interacting components of neuronal networks. Consequently, the topic of brain connectivity has become the leading factor in modeling neural system dynamics (Lee, Harrison, and Mechelli 2003; Stephan, Riera, Deco, and Horwitz 2008).

The advancement of neuroanalytical techniques has led to two core definitions of connectivity: functional connectivity and effective connectivity (Friston 1994; Sporns 2012). Functional connectivity is defined as the temporal correlation between spa-

tially remote neurophysiological events. In other words, two brain regions having highly correlated patterns of activation are said to be functionally connected. Importantly, observed functional connectivity makes no assumption about underlying neuroanatomical structure; it is a model-free statistical measure. Effective connectivity, on the other hand, is defined as the influence one neuronal system exerts over another. Anatomical and functional connectivity derive this link; two brain regions are effectively connected if they have a neural pathway connecting them and the excitement of one region causes an activation of the other region via this pathway. The effective relationship between the two regions may be uni- or bi-directional. Intuitively, this sequence of activation is manifested in observed functional connectivity.

1.3 Methods of analysis

The distinction between functional and effective connectivity is an example of the difficulty in converting data analysis to knowledge discovery. As stated in (Lee, Harrison, and Mechelli 2003), functional connectivity measures are descriptions of patterns of neural activity, while effective connectivity seeks to find explanations of the origins of these patterns. Thus, discovery of effective connectivity is tied to both functional connectivity and an *a priori* model of anatomical connections. Many methods have been proposed for characterizing effective connectivity models but, in general, fMRI data are analyzed using some manner of hypothesis-driven or data-driven techniques (Frackowiak 2004).

1.3.1 The General Linear Model

Introduced by Karl Friston in 1995 (Friston, Holmes, Worsley, Poline, Frith, and Frackowiak 1994), the general linear model (GLM) has long been the underlying framework for modeling the hemodynamic response. This model asserts that the observed data (the "response" variable) in an MRI signal comes from a linear mixture

of K components which may be protocol-dependent (e.g. task-related or disease-related) or protocol-independent (relating to scanner artifacts or physiological noise), plus some independent error:

$$x_{ij} = g_{i1}\beta_{1j} + g_{i2}\beta_{2j} + \ldots + g_{iK}\beta_{Kj} + e_{ij} \tag{1.1}$$

for some individual observation x_{ij}, where i and j index a specific observation (time-point, subject scan, etc.) and a specific voxel, respectively. Each coefficient g_{ik} is an explanatory variable relating the level of influence a source component β_{kj} has on the given observation; β_{kj} and K are unknown parameters to be estimated from the data. e_{ij} is a source of random error found in the signal. In matrix notation, the formulation of the GLM given in Equation 1.1 becomes:

$$\mathbf{X} = \mathbf{G}\beta + \mathbf{e}. \tag{1.2}$$

Here, $\mathbf{X}$ is a data matrix comprised of i rows corresponding to each scan and j columns corresponding to each scan. β becomes a parameter matrix whose rows represent underlying source components and whose columns reflect the associated parameter vector for each voxel. The I by K matrix $\mathbf{G}$ is commonly referred to as the design matrix, as it reflects the weight of the influence of each effect (its columns) in each dimension (its rows). Finally, $\mathbf{e}$ is a matrix of error terms normally distributed around the mean of the data.

1.3.2 Hypothesis-Driven Techniques

Hypothesis-driven methods of fMRI analysis seek to model effective connectivity by considering activated brain regions with *a priori* knowledge of function. Often, these methods analyze patterns of activity between a set of pre-selected voxels or Regions of Interest (ROIs) to find significant patterns of correlation. Increasingly, ROI selection in clinical studies is driven by exploratory analysis such as independent component analysis (see below), where selected regions show consistent activation

across subjects. ROIs present a simple way to explore activity patterns across conditions, although the use of ROI connectivity to describe brain activity as a whole is inappropriate because of the statistical bias introduced when the ROIs were first selected. (Poldrack 2007) gives an excellent overview of the benefits and issues in using ROIs to study functional imaging.

A priori models can also be based around task or disease analysis. Statistical Parametric Mapping (Friston, Holmes, Worsley, Poline, Frith, and Frackowiak 1994), for instance, can be used to assess the differences in brain activation patterns of a single scan based on task activation and deactivation events. The design matrix of the GLM in this case might consist of a pattern such as $[1, 1, 1, \ldots, -1, -1, -1, \ldots]$, where a value of 1 indicates task activation and -1 indicates resting state. Voxel activation is necessarily tied to this pre-defined paradigm.

The power of hypothesis-driven techniques is severely limited when generalized to large-scale analysis, whether in disease-state models or among large groups of subjects (Poldrack 2007). Type I errors are a very real possibility, especially when working with ROIs, where the selected areas are almost guaranteed to be significantly connected. In neuroinformatics contexts, too, hypothesis-driven methods become nearly intractable – that is, the concept of a "normal" brain falls away as the number of subject scans increases. Variability of anatomy and even effective connectivity make manual curation of ROI assessment infeasible in this context.

Finally, disease-state analysis can be very problematic in hypothesis-driven techniques, as regions other than those selected can be activated, and even the activation patterns observed in ROIs may not match correctly to the design matrix. For example, recent exploratory studies have found evidence that schizophrenia patients may display task activation patterns during resting state and have asymmetric activation in ROIs between hemispheres (Zhou, Liang, Tian, Wang, Hao, Liu, Liu, Jiang, *et al.* 2007; Hoptman, Zuo, D'Angelo, Mauro, Butler, Milham, and Javitt 2012).

1.3.3 Data-Driven Analysis

Because of the inherent drawbacks to *a priori* modeling approaches, much theoretical, and increasingly clinical, fMRI research of the last decade has focused on data-driven methods that approximate the GLM without forcing predetermined design onto the data. Without any further processing, data-driven methods are a form of exploratory analysis and, to this end, are commonly used to discover or confirm patterns and areas of activation in real fMRI data. While data-driven methods can be powerful, they may also may suffer from dependence on user-defined parameters or uninterpretable results. (For instance, these are two open issues with independent component analysis – model order selection can greatly affect the resulting description of activity, and the order of components returned is non-specific (Li, Adalı, and Calhoun 2007; Lindquist 2008).) However, their statistical power and ability to more completely describe neuroimages outweigh these caveats. The next two sections of this chapter briefly describe two of the most common data-driven fMRI analysis techniques, principle component analysis and independent component analysis. Chapter two defines our proposed data-driven analysis method, p7-means.

1.4 Principle Component Analysis

Principle Component Analysis (PCA) is a statistical method of extracting key features of a data set across all dimensions. These features (the "components") are ordered such that the largest components account for the greatest sources of variance in the original data set. Often, a small number of the largest components are able to sufficiently account for most of the data distribution in a set of observations. In other words, a small number of cross-dimensional components may be able to model the data of a higher-dimensional data set, especially if redundant (i.e. strongly correlated) dimensions exist. As such, PCA is often used as dimensionality reduction technique in data-driven analysis to decrease dimensional and computational complexity of a

data set. In addition to dimensionality reduction, PCA on a single-subject level has denoising benefits in choosing the most informative subspace for the data (Erhardt, Rachakonda, Bedrick, Allen, Adali, and Calhoun 2011). Algorithm 1 outlines the process of PCA:

Algorithm 1 Principal Component Analysis

Input: n by d data matrix X, number of principal components to select p
Output: n by p PCA-reduced data matrix X'

1: Zero-mean the dimensions (columns) of X.
2: Compute the d by d covariance matrix of the transpose of X.
3: Find the eigenvalues W and eigenvectors V of the covariance matrix.
4: Rank the W and their corresponding V from highest to lowest eigenvalue.
5: Let $V' =$ the first p eigenvectors from the ranked list selected into a p by d matrix, where each row of the matrix represents one eigenvector.
6: Let $X' = transpose(V' \times transpose(X))$.

Several methods exist for choosing the reduced dimensionality p, including Akaike Information Criterion (AIC), Kullback Information Criterion (KIC), and the Minimum Description Length (MDL) (Erhardt, Rachakonda, Bedrick, Allen, Adali, and Calhoun 2011). Choosing an optimal value for p is dependent on the manner of analysis. For single-subject scans, matching the optimal p will best fit the data; in group-level comparisons of scans, however, it is necessary to select a common p for all scans. Methods for group-level selection include performing PCA on the mean of the normalized dataset and on a spatially concatenated matrix of each subject's scans.

1.5 Independent Component Analysis

First introduced in (Bell and Sejnowski 1995), independent component analysis (ICA) is a method for data-driven analysis that incorporates unsupervised information and signal processing techniques to discover underlying source signals in a set of signal observations. ICA is a process of blind source separation, where the set of unknown source signals is derived under the assumption that the signal in a given observation is a linear weighted sum of all source signals. While the method is relatively new, it has

a rich history of application to fMRI (Mckeown, Makeig, Brown, Jung, Kindermann, Kindermann, Bell, and Sejnowski 1998; Calhoun, Adali, Hansen, Larsen, and Pekar 2003).

ICA has a corollary definition to the GLM. Given an observation x_i with t time indices, we seek to "unmix" x into the linear weighted sum of statistically independent signals:

$$x_i(t) = \sum_{j=1}^{m} a_{ij} s_j(t), \tag{1.3}$$

where s_j is an underlying signal, a_{ij} is the weight of the effect of that signal on x_i, and m is the number of distinct source signals (Hyvärinen 2013).

In more general terms, it is adequate to model all x_i in one n by t matrix X, where n is the number of distinct observations. The coefficients become a mixing matrix A of size n by n, and therefore the sources assume a matrix of the shape n by t. Importantly, this formulation of ICA requires the number of distinct sources m to be equal to the number of observations n.

$$X = AS. \tag{1.4}$$

Unlike PCA, which organizes data into orthagonal components, ICA assumes that the data consist of statistically independent components, meaning they have no systematic overlap in time or space (Lindquist 2008). Additionally, ICA models components as non-Gaussian sources, allowing for at most one Gaussian component. In application to fMRI, these assumptions have been justified by the observation that functional sources in the brain tend to be super-Gaussian signals, although artifactual signals may take different forms, such as Gaussian or sub-Gaussian (Calhoun and Adali 2006). A final assumuption of ICA is that the mixture matrix A is both square and invertible, which may be relaxed in certain instances (Hyvärinen 2013).

Many heuristic methods for estimating A and S exist, including the infomax, fastICA, and JADE algorithms (Calhoun, Adali, Hansen, Larsen, and Pekar 2003).

Numerous approaches build off these methods, particularly in multi-subject applications, known as Group ICA (GICA) (Calhoun, Liu, and Adalı 2009; Erhardt, Rachakonda, Bedrick, Allen, Adali, and Calhoun 2011). Still other methods, such as Probabilistic ICA (PICA), which we use as a point of comparison in this study, seek to tailor ICA models specifically to fMRI (Beckmann and Smith 2004). PICA, for one, allows for a nonsquare mixing process (i.e. m does not necessarily equal n) and assumes background Gaussian noise is inherently present in the data. The general formulation for this approach is

$$x_i = As_i + \mu + e_i, \tag{1.5}$$

where μ is the mean of x_i and e is some source of Gaussian noise. This equation is intented to more closely match the GLM.

CHAPTER TWO

p7-Means Algorithm

Mixture model estimation is a data-driven analysis technique where a set of K probability distributions are estimated from the data for purposes of describing the data set as a whole. The method is useful in a variety of contexts where data can be interpreted by a statistical model, including data mining, clustering/segmentation, and computer vision.

With the specific goal of functionally segmenting fMRI, this research applies an existing maximum likelihood estimator defined for Pearson type VII mixture models. This method is wrapped into an iterative K-learning algorithm, which we call p7-means. The algorithm is intended to provide a robust method for component derivation from noisy data sets, with the additional feature of estimating the number of components K from natural model-fitting, rather than running with a fixed number of components decided *a priori* or estimated from the data prior to analysis. Another main difference in the modeling principles of p7-means when compared to methods such as ICA is to treat sub-Gaussian components as sources of noise confounds, rather than attempting to characterize them as separate functional sources.

2.1 The Pearson VII Distribution

The Pearson type VII family of distributions (Pearson 1916) is a class of symmetrical frequency curves that includes both the normal- and t-distributions. In statistical modeling, Pearson type VII distributions have been used for robust mixture modeling because of their ability to adopt both Gaussian and super-Gaussian forms and for their power against outliers in noisy data sets (Sun, Kaban, and Garibaldi 2010). The probability density function for the multivariate Pearson type VII distri-

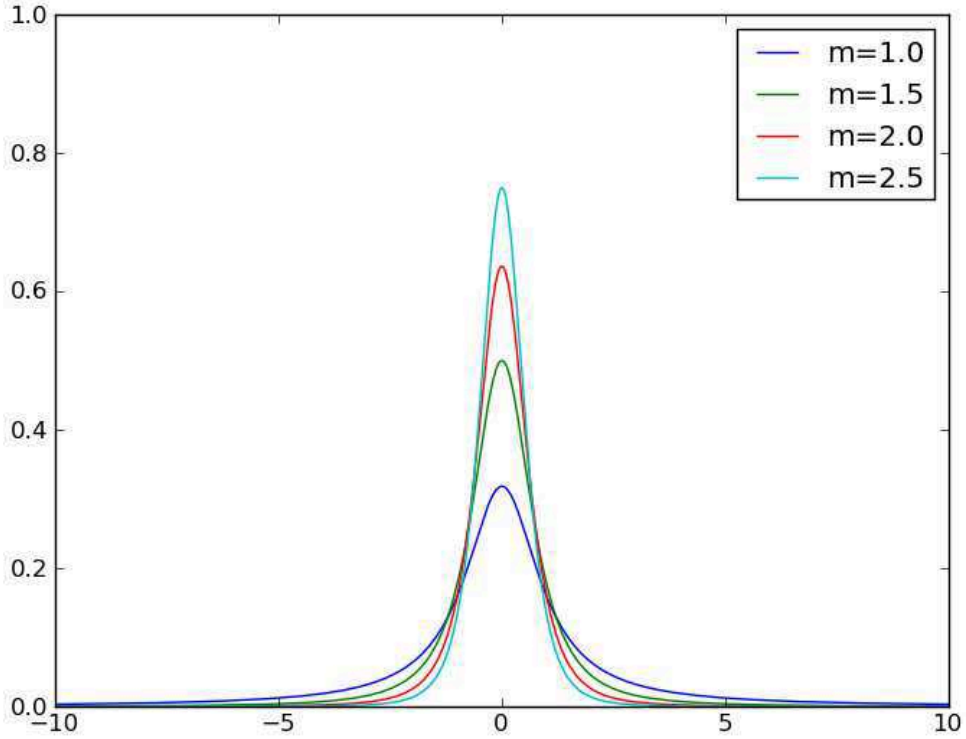

Figure 2.1: Pearson-VII type distributions with unit variance for different values of m.

bution in d dimensions is

$$p(\mathbf{t};\mu,\Lambda,m) = \frac{\Gamma(m)}{\pi^{\frac{d}{2}}\Gamma(m-\frac{d}{2})}\,|\Lambda|^{-\frac{1}{2}}\left[1+\Delta^2\right]^{-m}, \tag{2.1}$$

where $\Delta^2 = (\mathbf{t}-\mu)^T\Lambda^{-1}(\mathbf{t}-\mu)$ is the Mahalanobis squared distance, μ is the d-dimensional distribution mean, Λ is a scaling matrix, and m is the degrees of freedom. Figure 2.1 shows the range of robustness the Pearson VII distribution can assume for different degrees of freedom.

Interestingly, a Pearson type VII distribution can be expressed as a scale mixture of normals. Following the formulations given in (Sun, Kaban, and Garibaldi 2010), we observe the following Pearson VII property:

$$p(\mathbf{t};\mu,\Lambda,m) = \int_0^\infty \mathcal{N}\left(\mathbf{t}\mid\mu,\frac{\Lambda}{u}\right)\mathcal{G}\left(u\mid m-\frac{d}{2},\frac{1}{2}\right)du, \tag{2.2}$$

where $\mathcal{G}$ is the gamma distribution, given parameters a and b and latent variable u.

$$\mathcal{G}(u \mid a, b) = b^a u^{a-1} \frac{e^{-bu}}{\Gamma(a)}. \tag{2.3}$$

We can then rephrase the distribution model as the dependence of our observed variable over the integral of our (non-negative) latent variable:

$$p(\mathbf{t}; \mu, \Lambda, m) = \int_0^\infty p(\mathbf{t} \mid u) p(u) du, \tag{2.4}$$

where $p(\mathbf{t} \mid u)$ and $p(u)$ correspond to the normal and gamma scale distributions of our model, respectively. In a mixture model context, we incorporate a prior probability for each of K independent models, given an observation t:

$$p(t \mid \Theta_k) = p(t, u, \Theta_k) = p(t \mid u, \Theta_k) p(u \mid \Theta_k) p(\Theta_k), \tag{2.5}$$

where $\Theta_k = \{\mu_k, \Lambda_k, m_k\}$ is one of the K Pearson VII models, and $p(\Theta_k)$ is the prior probability of any given model ($\sum_K p(\Theta_k) = 1$).

The log-likelihood of Θ given a finite data set $\mathcal{Y} = \{t_1, \ldots, t_N\}$ is thereby given as:

$$\mathcal{L}(\Theta \mid \mathcal{Y}) = \sum_N \sum_K \log \Big(p(t_n \mid u_n, \Theta_k) p(u_n \mid \Theta_k) p(\Theta_k) \Big). \tag{2.6}$$

2.2 Pearson VII EM Algorithm

To estimate a Pearson VII mixture model from a data set, we use the expectation-maximization procedure for maximum likelihood estimation outlined in (Sun, Kaban, and Garibaldi 2010). In this section, we give an overview of their formulation of the algorithm. The maximum likelihood algorithm works by, after initializing parameters for each of the K specified models, iteratively running the expectation, or E-step, and maximization, or M-step, procedures until the likelihood of the model converges.

2.2.1 E-step

The E-step of the modelling algorithm computes the posterior probabilities, or expectations, of each data point given the set of estimated components, Θ. From

this, we are able to calculate the likelihood of the estimated model.

In our explanation, let us say that $p(\Theta_k|t_n)$ represents the probability that model Θ_k has some influence in the value of observation t_n. This is commonly referred to as the "responsibility" or "expectation" of t_n for model Θ_k. We then have:

$$p(\Theta_k, u_n \mid t_n) = p(\Theta_k \mid t_n)p(u_n \mid t_n, \Theta_k). \tag{2.7}$$

Bayes' rule allows us to reformulate the posteriors as:

$$p(\Theta_k \mid t_n) = \frac{p(t_n \mid \Theta_k)p(\Theta_k)}{\sum_{k'} p(t_n \mid \Theta_{k'})p(\Theta_{k'})} \tag{2.8}$$

and

$$\begin{aligned} p(u_n \mid t_n, \Theta_k) &= \frac{p(t_n \mid u_n, \Theta_k)p(\Theta_k)}{p(t_n \mid \Theta_k)} \\ &= \frac{\mathcal{N}\Big(t_n \mid \mu_k, \frac{\Lambda_k}{u_n}\Big)\mathcal{G}\Big(u_n \mid m_k - \frac{d}{2}, \frac{1}{2}\Big)}{p(t \mid \mu_k, \Lambda_k, m_k)} \\ &= \frac{\left(\frac{1}{(2\pi)^{d/2}}\frac{u_n^{d/2}}{|\Lambda_k|^{1/2}}e^{\frac{1}{2}u_n\Delta_k^2}\right)\left(\frac{1}{2}^{m_k-\frac{d}{2}}u_n^{m_k-\frac{d}{2}-1}\frac{e^{-\frac{u_n}{2}}}{\Gamma(m_k-\frac{d}{2})}\right)}{\frac{\Gamma(m_k)}{\pi^{\frac{d}{2}}\Gamma(m_k-\frac{d}{2})}|\Lambda_k|^{-\frac{1}{2}}[1+\Delta_k^2]^{-m_k}} \\ &= \left(\frac{[1+\Delta_k^2]}{2}\right)^{m_k} u_n^{m_k-1}\frac{e^{-\frac{[1+\Delta_k^2]}{2}u_n}}{\Gamma(m_k)} \\ &= \mathcal{G}\left(a_{nk} = \frac{[1+\Delta_k^2]}{2}, b_{nk} = m_k\right), \end{aligned} \tag{2.9}$$

where $\Delta_k^2 = (t_n - \mu_k)^T\Lambda_k^{-1}(t_n - \mu_k)$.

Using the hyperparameters a_{nk} and b_{nk} derived in the above equation, we can define the following estimated latent variables:

$$u_{nk} = \int u_n p(u_n \mid t_n, \Theta_k)du_n = \frac{a_{nk}}{b_{nk}} \tag{2.10}$$

and

$$\log u_n k = \int [\log u_n]p(u_n \mid t_n, \Theta_k)du_n = \psi(a_{nk}) - \log b_{nk}, \tag{2.11}$$

where ψ is the digamma function.

Finally, we are able to compute the log-likelihood of the estimated model:

$$\begin{aligned}
\mathcal{L}(\Theta) =& \sum_N \sum_K p(\Theta_k \mid t_n) \log \left[p(t_n \mid u_n, \Theta_k) p(u_n \mid \Theta_k) p(\Theta_k) \right] \\
=& \sum_N \sum_K p(\Theta_k \mid t_n) \log [p(t_n \mid u_n, \Theta_k)] \\
&+ \sum_N \sum_K p(\Theta_k \mid t_n) \log [p(u_n \mid \Theta_k)] \\
&+ \sum_N \sum_K p(\Theta_k \mid t_n) \log [p(\Theta_k)] \\
=& \sum_N \sum_K p(\Theta_k \mid t_n) \log \mathcal{N}(t_n \mid \mu_k, \frac{\Lambda_k}{u_{nk}}) \\
&+ \sum_N \sum_K p(\Theta_k \mid t_n) \log \mathcal{G}(u_{nk} \mid m_k - \frac{d}{2}, \frac{1}{2}) \\
&+ \sum_N \sum_K p(\Theta_k \mid t_n) \log p(\Theta_k)
\end{aligned} \tag{2.12}$$

Applying the log of the normal and gamma distributions gives us:

$$\begin{aligned}
\log \mathcal{N}(t_n \mid \mu_k, \frac{\Lambda_k}{u_{nk}}) &= \log \left(\frac{1}{(2\pi)^{d/2}} \frac{u_{nk}^{d/2}}{|\Lambda_k|^{1/2}} e^{\frac{1}{2} u_{nk} \Delta_k^2} \right) \\
&= \frac{d}{2} u_{nk} - \frac{1}{2} \log |\Lambda_k| - \frac{1}{2} u_{nk} \Delta_k^2 - \frac{d}{2} \log(2\pi)
\end{aligned} \tag{2.13}$$

and

$$\begin{aligned}
\log \mathcal{G}(u_{nk} \mid m_k - \frac{d}{2}, \frac{1}{2}) =& \log \left(\left(\frac{1}{2} \right)^{m_k - \frac{d}{2}} u_n^{m_k - \frac{d}{2} - 1} \frac{e^{-\frac{u_{nk}}{2}}}{\Gamma(m_k - \frac{d}{2})} \right) \\
=& (m_k - \frac{d}{2}) \log \frac{1}{2} + (m_k - \frac{d}{2} - 1) \log u_{nk} \\
&- \frac{1}{2} u_{nk} - \log \Gamma(m_k - \frac{d}{2})
\end{aligned} \tag{2.14}$$

2.2.2 M-step

The M-step of the algorithm seeks to find the model components Θ_k that maximize $\mathcal{L}(\Theta)$. Model priors can simply be updated as the average expectation for each component:

$$p(\Theta_k) = \frac{1}{N} \sum_N p(\Theta_k | t_n) \tag{2.15}$$

The mean and scale matrices for each component can be computed through weighted least squares estimation; that is, the parameters can be estimated from the data (considered as independent samples), weighted by our currently estimated component priors and latent u variables:

$$\mu_k = \frac{\sum_N p(\Theta_k \mid t_n) u_{nk} t_n}{\sum_N p(\Theta_k \mid t_n) u_{nk}} \tag{2.16}$$

$$\Lambda_k = \frac{\sum_N p(\Theta_k \mid t_n) u_{nk} (t_n - \mu_k)(t_n - \mu_k)^T}{\sum_N p(\Theta_k \mid t_n)} \tag{2.17}$$

Calculating the degrees of freedom forces us to take the derivative of Equation 2.12 w.r.t. m_k and setting it to zero, leaving us with:

$$\frac{\partial}{\partial m_k} \sum_N p(\Theta_k \mid t_n) \left[(m_k - \frac{d}{2}) \log \frac{1}{2} + (m_k - \frac{d}{2} - 1) \log u_{nk} - \log \Gamma(m_k - \frac{d}{2}) \right] = 0$$

$$\sum_N p(\Theta_k \mid t_n) \left[\log u_{nk} - \log(2) - \psi \left(m_k - \frac{d}{2} \right) \right] = 0$$

$$\psi^{-1} \left(\frac{\sum_N p(\Theta_k \mid t_n) \left[\log u_{nk} - \log(2) \right]}{\sum_N p(\Theta_k \mid t_n)} \right) + \frac{d}{2} = m_k \tag{2.18}$$

2.3 The p7-means Algorithm

Now that we have defined the EM approach for learning our Pearson VII mixture model, we merely need to wrap our algorithm in an appropriate K estimator. Several methods have been proposed for estimating the number of components in a data set, including pre-processing methods such as Minimum Description Length (MDL), which estimates optimal model order based on log-likelihood of the scaled data eigenvalues balanced by some penalty model, and K-learning methods such as pg-means, which iteratively adds components to the model and, for each K-component model, projects the data and model onto one dimension and uses the Kolmogorov-Smirnov statistic and Monte Carlo estimation to determine whether the model fits the data (Feng and Hamerly 2007). For our algorithm, p7-means, we employ a simple iterative approach to learning K that begins with one component and grows until at least one component

is estimated that has a prior probability, $p(\Theta_k)$ less than a user-specified threshold t. In our experiments we use $t = 0.01$, which means that the algorithm exits when it discovers a component that explains less than one percent of the data.

To initialize our starting component, we estimate μ_1 as the mean of the data across every dimension, and we set m_1, and Λ_1 according to (Nagahara 2004):

$$m_1 = \frac{5 \times \text{kurtosis} - 9}{2 \times \text{kurtosis} - 6} \tag{2.19}$$

$$\Lambda_1 = \Sigma \times (2m_1 - 3), \tag{2.20}$$

where Σ is the covariance matrix of the data.

When generating a new component for inclusion in the model, p7-means actually runs a pre-determined number (in our experiments, ten) of EM runs and selects the result with the highest log-likelihood as the new mixture model. To seed these runs, p7-means all voxels by ascending e value, as defined in (Sun, Kaban, and Garibaldi 2010):

$$e_n = \sum_K p(\Theta_k \mid t) u_{nk}. \tag{2.21}$$

The purpose of running multiple EM runs is to find the best new model for the set of data points that most poorly fit the current mixture model. As the EM approach optimizes only to local maxima (Moon 1996), the use of multiple model initializations allows for the possibility of approaching a better fit than if only the lowest probability point was used as a seed. This is similar to the approach used in pg-means, except that we choose to use only the least-likely data points (making the algorithm deterministic), rather than a mixture of unlikely and random data points.

In calculating m and Λ values for the new component, the algorithm computes the values just as in the initialization step, except that it only samples the first N/K voxels when computing kurtosis and covariance, where K is the number of components with the newest component included. Model means are assigned for each of the r runs by selecting the first r voxels with the lowest e values.

We implemented p7-means in the Python programming language using the NumPy (http://www.numpy.org) and SciPy (http://www.scipy.org) libraries (Ascher, Dubois, Hinsen, Hugunin, and Oliphant 1999; Jones, Oliphant, Peterson, *et al.* 01). Figure 2.2 shows the performance of p7-means on a simulated data set of five Gaussian components with a uniform background noise component. The top four figures are, from left to right, the progression of the algorithm from one to four centers. Each color represents a different estimated component, and data points are assigned to a cluster based on the highest expectation. The bottom picture shows the final estimation of the algorithm. We see that p7-means correctly distinguished the five Gaussian components from the uniform noise (shown in red), but two components (shown in light blue) were incorrectly estimated as one single component.

As with most mixture modeling algorithms, p7-means becomes computationally difficult as the data dimensionality increases. Additionally, the covariance matrices of high-dimensional data, particularly fMRI, tend to approach singularities because the distinction between two given dimensions may be hard or impossible to define. Therefore, it is common practice in mixture model to reduce high-dimensional data along the first d components of variance in the data set. In other words, we apply principal component analysis to high-dimensional data sets and run p7-means using the first d (a pre-selected number) dimensions of variance, instead.

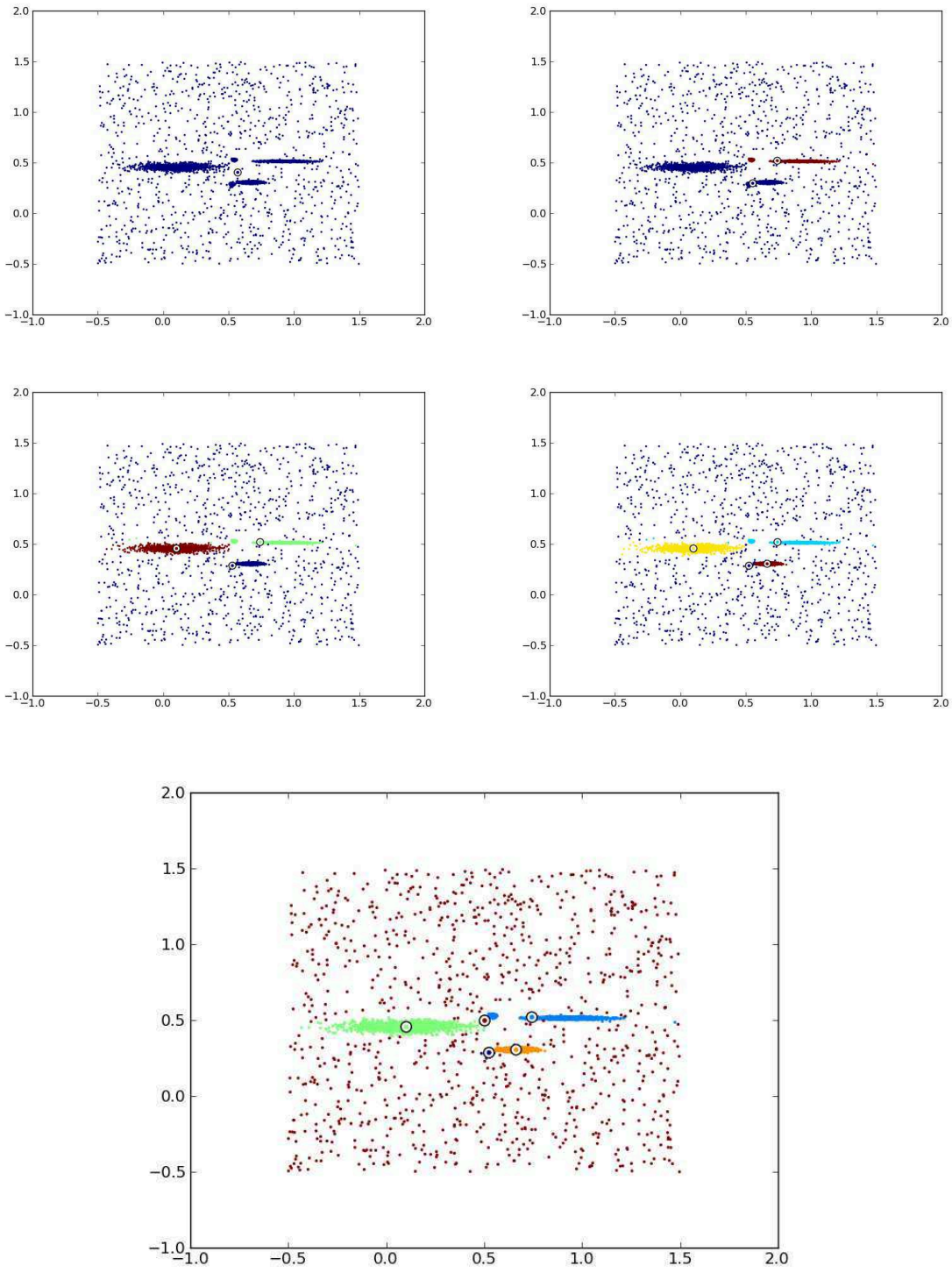

Figure 2.2: Performance of p7-means on a simulated data set of 6000 points, 5000 coming from random Gaussian distributions, and one consisting of a uniform distribution.

CHAPTER THREE

Experimental Results

To analyze the performance of p7-means on real fMRI data, we apply the algorithm to resting-state monkey fMRI. We qualitatively evaluate the components returned by the method and compare the results to the components returned by ICA.

3.1 Image Acquisition and Pre-processing

Three resting-state monkey fMRI were obtained Oregon National Primate Research Center on a Siemens Tim-Trio (3T) MRI scanner. We refer to these scans as Scans #1, #2, and #3. Scans #1 and #3 are images of normal control monkeys, while Scan #2 is an image of a monkey suspected to exhibit alcohol addiction.

Resting state echo-planar images recorded BOLD fluctuation over 864 time-points (excluding the first four acquisitions) approximately two seconds apart (TR=2.07). Pre-processing of the functional images was performed using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (www.fmrib.ox.ac.uk/fsl), FMRIB's Software Library (Jenkinson, Beckmann, Behrens, Woolrich, and Smith 2012; Woolrich, Jbabdi, Patenaude, Chappell, Makni, Behrens, Beckmann, Jenkinson, Smith, *et al.* 2009; Smith, Jenkinson, Woolrich, Beckmann, Behrens, Johansen-Berg, Bannister, De Luca, Drobnjak, Flitney, *et al.* 2004). The image was (1) motion corrected using MCFLIRT (Jenkinson, Bannister, Brady, Smith, *et al.* 2002); (2) slice-timing corrected; (3) masked using brain extraction with BET, f=0.3 (Smith 2002); (4) spatially smoothed (FWHM, 5mm); and (5) high-pass filtered (sigma=50 seconds).

3.2 Probabilistic Components

Before applying p7-means to the monkey scans, the data were first voxel-wise de-meaned and subsequently reduced to 20 dimensions through PCA. Figures 3.1, 3.2, and 3.3 show the components returned by p7-means for Scans #1, #2, and #3. Scan #1 shows 34 strongly segmented components (that is, having little overlap) that also tend to be spatially localized. Components also appear to correlate strongly between hemispheres of the brain for the same functional regions, although some complementary areas are divided into two components (for example, components e-blue and f-blue).

Scans #2 and #3 show strong segmentation of components, as well, with p7-means generating 23 components for Scan #2 and 33 components for Scan #3. It is difficult to qualitatively evaluate the overall performance of p7-means using only three scans. However, the number and location of components predicted for Scans #1 and #3 appear much more similar than the components returned for Scan #2. For example, component c-green for Scan #1 and component e-red for Scan #3 appear to match structurally to some degree around the putamen and caudate nucleus. Scan #2 also has a putamen- and caudate nucleus-related component (a-blue); however, the component also incorporates posterior white matter regions, as well, which is significantly different from the Scan #1 and Scan #3 components.

It is interesting that Scan #2, a brain suspected to exhibit alcoholism, displays fewer regions of distinct functional activation when compared to Scans #1 and #3. In an exploratory but limited fashion, this suggests that p7-means may have been quite successful in distinguishing functionally important patterns of activation in the brain. While more trials are necessary to identify the specificity of the components returned by p7-means for different brain states, the fewer number of components returned for Scan #2 indicate that the algorithm is able to successfully characterize a marked difference in activation for disease-state scans.

3.3 Comparison with ICA

To validate the performance of our approach against existing methods, we compared the output of our single-scan results for Scan #1 to the FSL program MELODIC, which estimates functional similar regions using probabilistic independent component analysis (Beckmann and Smith 2004). In applying this method, voxels were de-meaned and variance normalized, and the data were reduced to 20 dimensions using PCA dimensionality reduction.

Initially, we used MELODIC's automatic estimation tool for calculating the number of independent components to detect. However, this ended up generating 631 separate components, which is impossible to interpret and yields no components particularly interesting in size or location. Therefore, we set the number of components to a predetermined 20, which is around the number that has commonly been used in literature when applying group ICA (Calhoun, Adali, Hansen, Larsen, and Pekar 2003, for example), which requires the number of components be consistent. Figure 3.4 shows the results of this analysis.

Figure 3.5 shows the results of matching p7-means probability maps to ICA maps on a per-voxel basis. The ICA maps generated by MELODIC contain probabilistic involvement in each component based on a mixture model of a single gamma and a single normal distribution. The data modeled by MELODIC in this case is the spatial map of per-voxel mixing coefficients for each component, normalized by z-statistics. To quantify the amount of similarity between the two, we used Pearson's correlation coefficient, which ranges from -1, meaning completely anticorrelated, to 1, meaning completely correlated. The formula for correlation is:

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

Figure 3.6 shows the four ICA/p7-means component pairs with the highest r scores. Images (a), (b), and (d) appear to indicate a more coherent, localized performance of p7-means over ICA in terms of component specificity ("noisiness"). In

image (c), we see the same effect; however, it also appears that ICA has predicted additional activation in the frontal cortex that was not detected by p7-means. This may not be a false positive or false negative for either algorithm – computationally, each algorithm perceives the data in a very different way, and as always, the parameters given to each algorithm could affect its performance. Nevertheless, the observation does reaffirm the notion that p7-means returns a set of components that are more spatially segmented when compared to ICA. On the other hand, ICA components are more likely to predict similar patterns of activation in different (that is, not mirrored across hemispheres) regions of the brain.

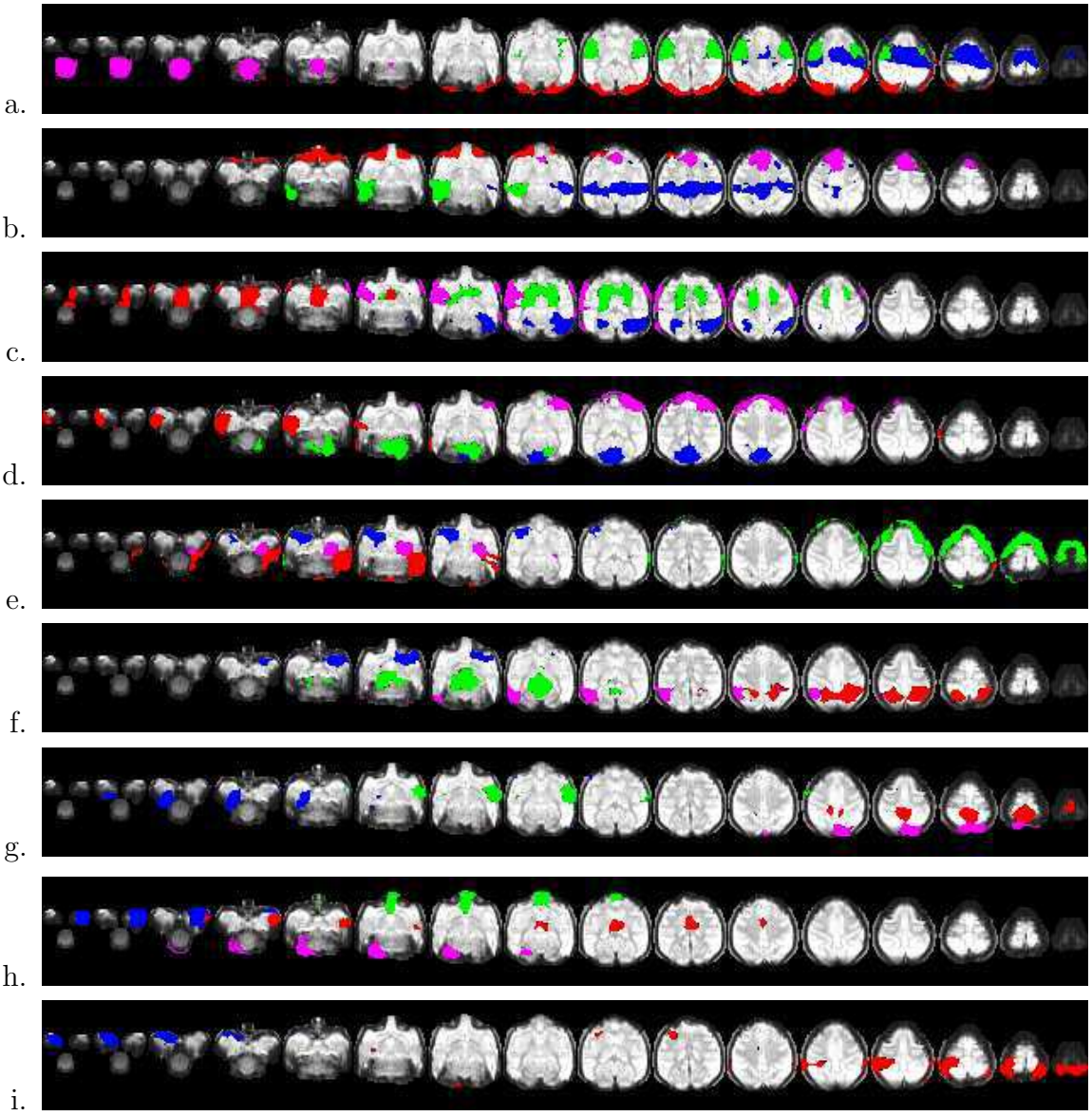

Figure 3.1: Result of the p7-means algorithm on a smoothed, resting-state monkey fMRI image (Scan #1). Using PCA dimensionality reduction to 20 dimensions, the p7-means algorithm discovers 34 generative components for this scan. Each image shows four of these components, in no specific order, with a probability threshold cutoff of $p(\mathbf{t} \mid \Theta_k) > 0.5$.

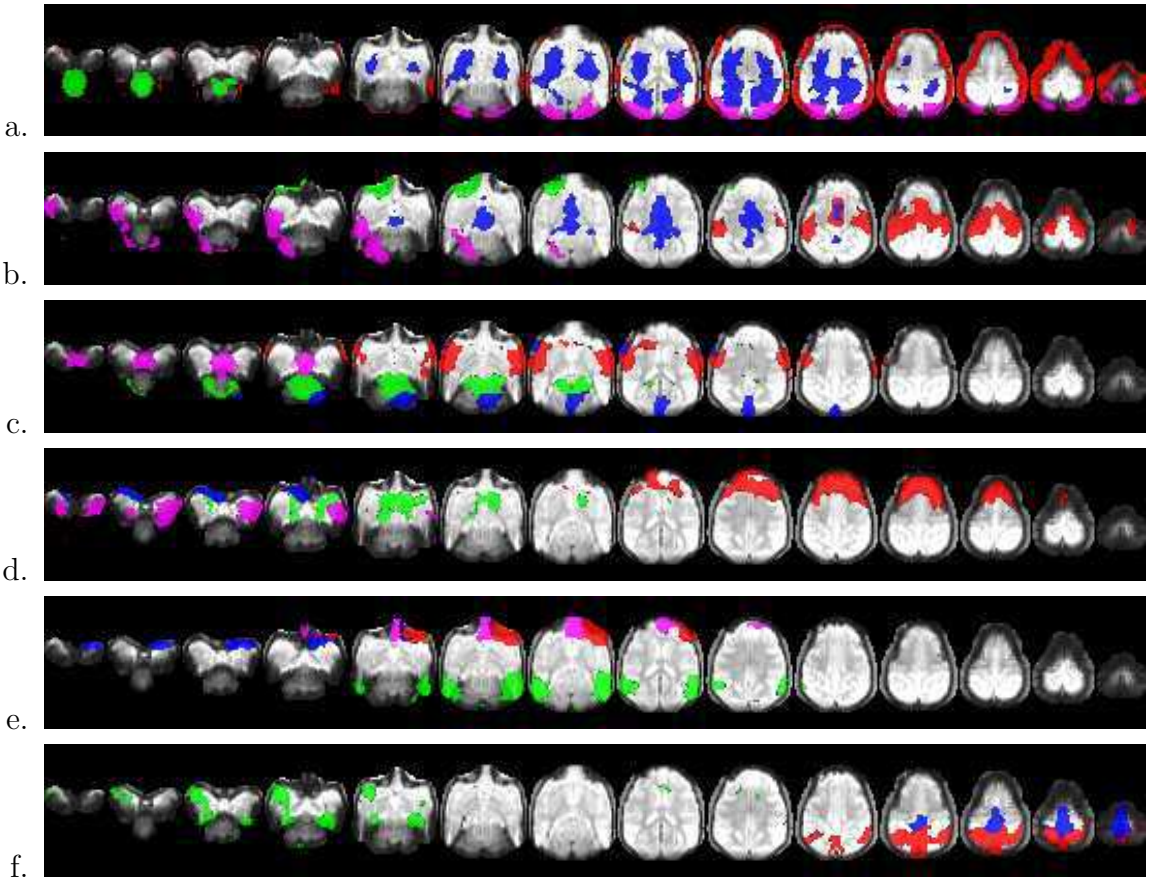

Figure 3.2: Result of the p7-means algorithm on a smoothed, resting-state monkey fMRI image (Scan #2). Using PCA dimensionality reduction to 20 dimensions, the p7-means algorithm discovers 23 generative components for this scan. Each image shows four of these components, in no specific order, with a probability threshold cutoff of $p(\mathbf{t} \mid \Theta_k) > 0.5$.

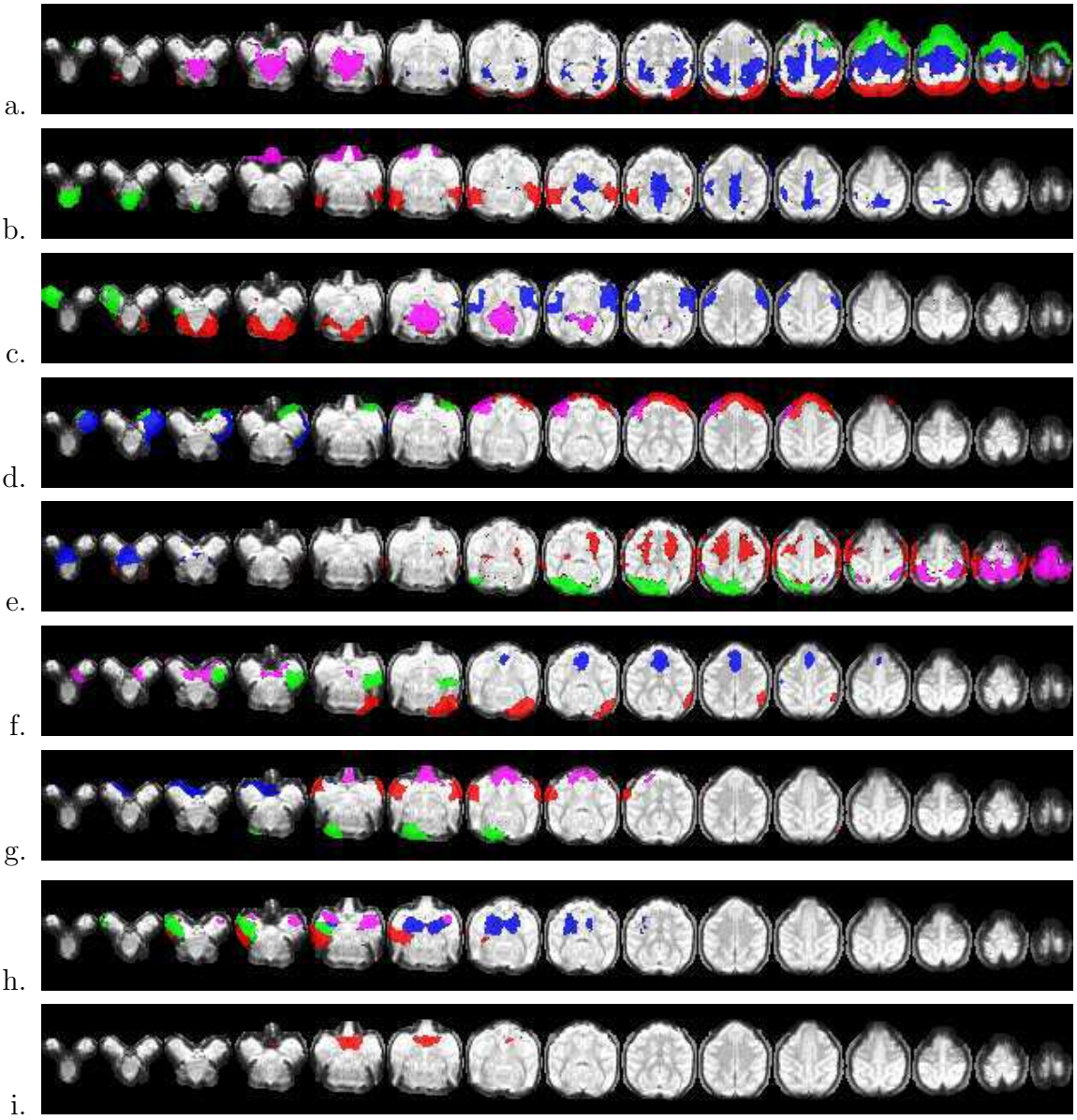

Figure 3.3: Result of the p7-means algorithm on a smoothed, resting-state monkey fMRI image (Scan #3). Using PCA dimensionality reduction to 20 dimensions, the p7-means algorithm discovers 33 generative components for this scan. Each image shows four of these components, in no specific order, with a probability threshold cutoff of $p(\mathbf{t} \mid \Theta_k) > 0.5$.

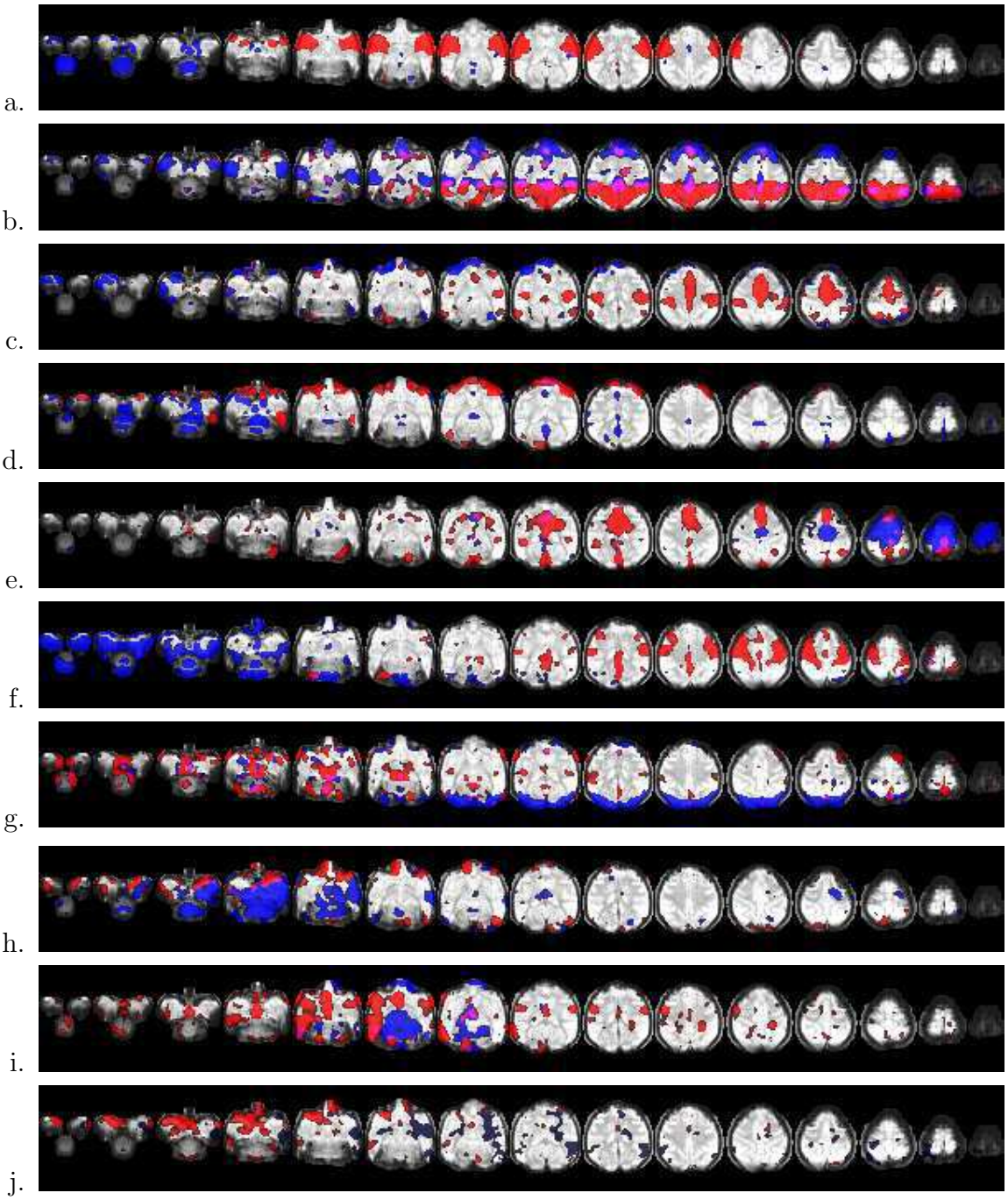

Figure 3.4: Probability maps from Melodic ICA components after PCA dimensionality reduction to 20 dimensions. Each image shows two of these components, with any overlap indicated in purple, using a threshold cutoff of $p(\mathbf{t} \mid \Theta_k) > 0.5$.

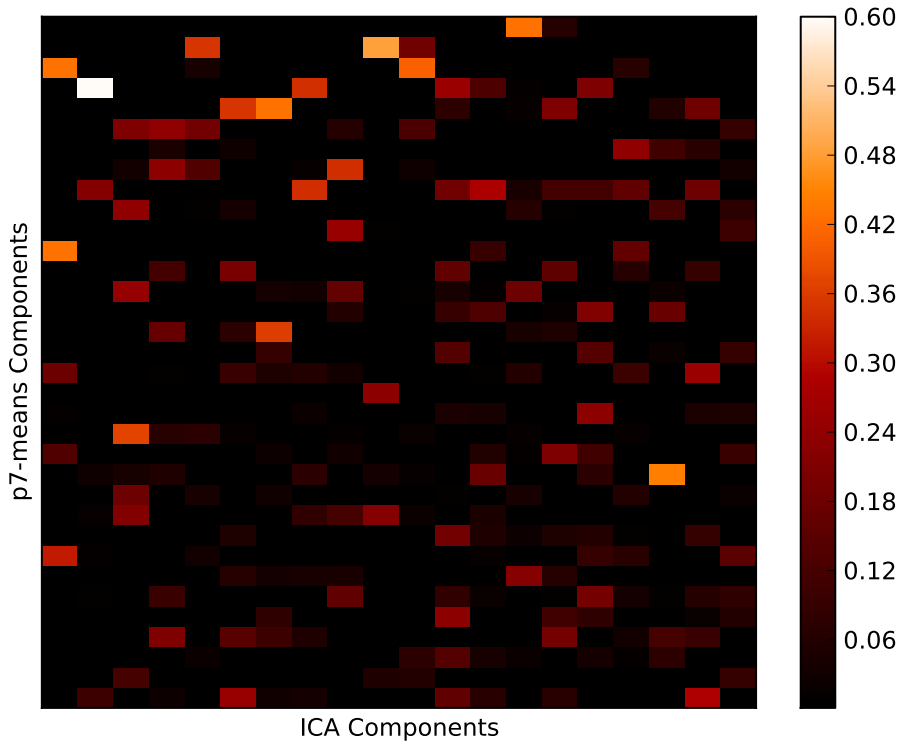

Figure 3.5: Pearson's r scores between p7-means and MELODIC ICA probability maps. The components (each row and column) are given in no specified order. Qualitatively, we see that nearly every component from p7-means corresponds at least somewhat with a component from ICA, and vice-versa.

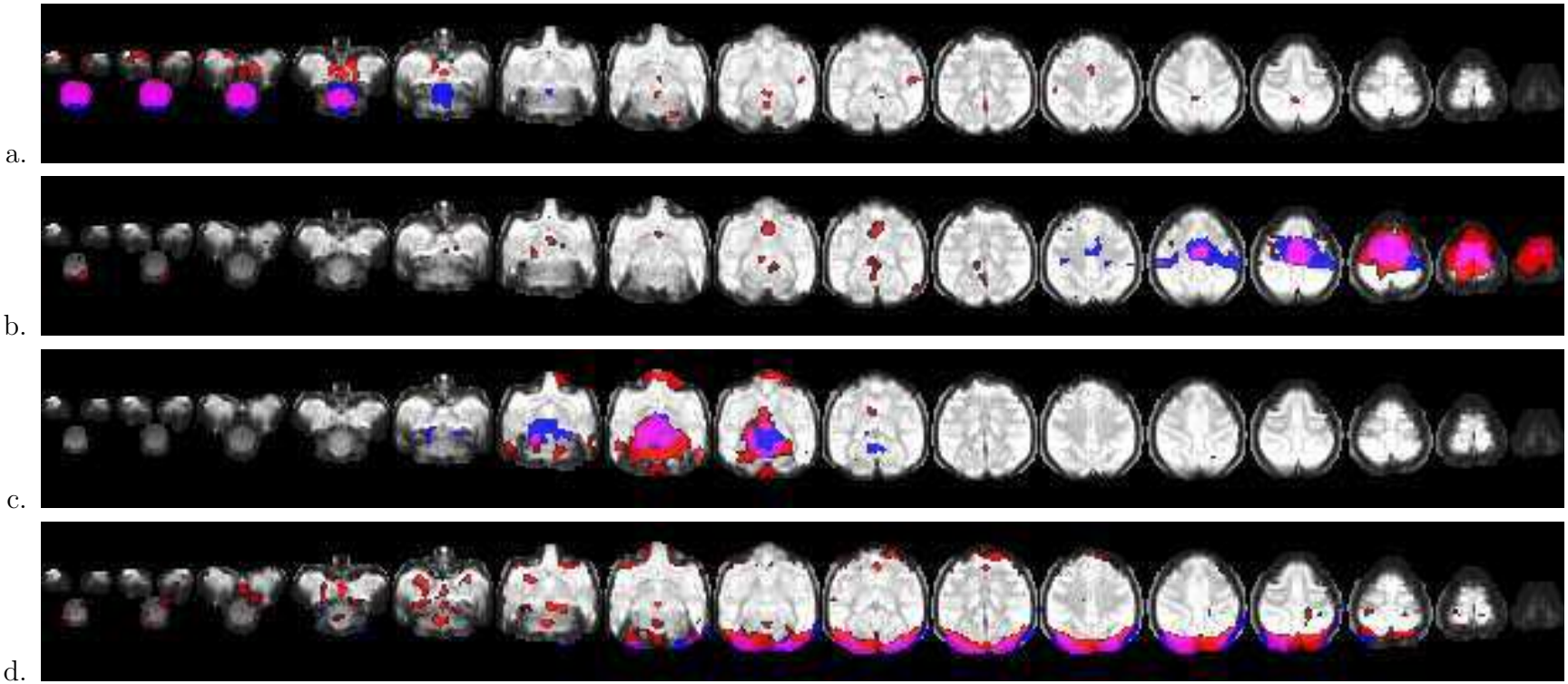

Figure 3.6: Four paired ICA (red) and p7-means (blue) components having the highest Pearson's r scores. Areas of overlap between the components appear as purple (threshold: $p(\mathbf{t} \mid \Theta_k) > 0.5$). r values are: (a) 0.6, (b) 0.458, (c) 0.445, and (d) 0.428.

CHAPTER FOUR

Discussion and Conclusion

The purpose of this thesis was to lay the framework for a new method for functional connectivity analysis of fMRI that has robust performance over existing methods. While p7-means is shown merely as a proof-of-concept approach here, there are many avenues of research still available for the algorithm. We have shown, however, that p7-means is able to perform comparatively to independent component analysis on real-world data, although the lack of specific ground truth data makes this comparison difficult to quantify.

Historically, there has been little research in applying clustering-type algorithms to functionally segment fMRI, especially when compared to independent component analysis, which has a panoply of algorithms and implementations throughout the literature. The research presented here is therefore novel in its application to fMRI, and as such, we believe that p7-means is an innovative method of fMRI analysis. If nothing more, we hope this research opens the door for discussion about how robust methods can be applied in data-driven fMRI, as well as challenging the current field of algorithm development to consider the advantages and pitfalls of popular methods in a new light.

With that said, p7-means (and whatever may come from it) still has a long way to go before it is even comparable to the vastly more researched analysis methods currently in use. Several optimizations in the algorithm should be the focus for future research:

1. Smarter thresholding: The current implementation of p7-means uses a very rudimentary threshold for algorithm completion. In future versions of the algorithm, it could be beneficial to use different methods for checking whether

the components returned by the algorithm adequately describe the data.

2. Noise-agnostic modeling: One limitation to applying an iterative K-learning algorithm is that noise components of the data are still included in one or more components. Methods could be developed to adjust the mixture model such that noise components do not as significantly affect the placement of new components.

3. Alternatives to PCA dimensionality reduction: p7-means, like almost all mixture model algorithms, is restricted in the number of dimensions it can handle, especially in data sets where many dimensions have high redundancy, such as is the case in fMRI. Estimation of the covariance matrix (with some data loss, of course) and its inverse may allow us to avoid dimensionality reduction and possibly decrease the alogorithm runtime. (Sun, Kaban, and Garibaldi 2010), for instance give a method in their paper for estimating the covariance matrix with the Wishart distribution.

4. Ordering components: ICA and p7-means both suffer from not having a specified order of components that makes sense in terms of functional significance. Further research into effective ordering strategies could benefit not just our own research, but the general data-driven community, as well.

5. u-map incorporation: One idea, not addressed in this thesis, is to perform analysis on the estimated latent variables u_{nk}, which are associated with each voxel for each component. u values give a less modular, but possibly more informative, map of the brain for each component that includes localized white-matter tracts (data not shown). The exact use of these maps in our analysis has yet to be determined, but at the very least, they offer an additional layer of information which we can analyze.

There exist several open applications of p7-means to be explored, as well, the chief of these being between-subject registration and comparison. Because the algorithm inherently provides a statistical model of functional activity, registration applications can be applied using distribution similarity tests. In other words, when comparing the results of p7-means on two (or more) different scans, the components could be matched using a statistical test. The Kolmogorov-Smirnov test, for example, could be used to measure the empirical distributions of voxel values between all pairs of components between two different scans. A similar pair of components would have a smaller Kolmogorov-Smirnov statistic, and therefore component pairs could be matched by the best statistic. In addition, the statistic gives a weak assertion of the degree of matching for any pair of components, implying that multiple functional components in one scan could be associated with one or more components in the other. Importantly, this statistical matching would allow for registration in probabilistic functional space, rather than anatomical space – a non-atlas (that is, not spatially normalized) between-subject comparison. Especially for neuroinformatics research, this anonymous registration method could be very powerful in elucidating functional characteristics of the brain on a population-wide level.

An alternative to the Kolmogorov-Smirnov test that abstracts the model analysis even further is to use information metrics such as the Kullback-Leibler divergence (Kullback and Leibler 1951). This measure would allow us to match component pairs between scans based solely on the predicted Pearson VII distribution of each component, rather than using the set of voxels most strongly associated with a given component as a comparative sample. When comparing the probability distributions associated with components from two scans, a distribution from the second scan having low divergence from a given distribution in the first scan can be thought of as well described by the first scan's distribution. Effectively, the measure of divergence functions for measuring differences in component distributions just as the

Kolmogorov-Smirnov statistic does for measuring differences in component samples. The implications of this are that full functional registration can be achieved as a post-processing step using only computed statistical models. Developing divergence analysis methods for scan comparison is a key focus of future research.

The approach also paves the way for the development of functional atlases, where the voxels of other scans are mapped to the pre-defined mixture model of the atlas. This would give a "null-hypothesis map" for (atlas-based) normal functioning. In turn, this ground-truth could be analyzed against the results of running p7-means on the non-atlas scan. Component-wise correlation or model likelihood comparison could give an idea of how well the scan matches the atlas control. Another method could be to apply the Kolmogorov-Smirnov test or Kullback-Leibler divergence using the set of atlas distributions as references and the set of predicted components to match against (using the associated voxels for each predicted component as samples in the case of the Kolmogorov-Smirnov test). In this case, predicted components having no strongly corresponding atlas component would be marked as potentially abnormal in function.

These approaches solve issues of non-standard structural damage, as well. For example, a traumatic brain injury patient may have severe structural damage in one area of the brain, causing *a priori* and structural atlas-based methods poorly model the functional capabilities across the entire brain, especially if neuroplasticity has allowed another region (or other regions) to adopt and maintain the functionality of the affected region. Applying the mixture model in a non-atlas way, on the other hand, could allow us to detect this alternative functional structure.

The mixture model approach also lends itself well to the simulation of fMRI, as the probability makeup of each voxel can be randomly sampled to try and simulate the patterns of neuronal firing across the entire brain. Comparison of these randomly-induced processes to actual fMRI signals could help us better understand

the underlying mechanisms as well as confounds that form the fMRI signal. Moving forward, we could weight our random samplings with signal-related biases based on *a priori* models of neuronal firings. Matching these adjusted random processes to actual observations would aid us in validating the accuracy of our neuronal models.

Our long-term intention is to apply p7-means as a robust method for disease-state and group analysis, particularly on a neuroinformatics level. We believe that the algorithm shows particular promise as far as comparitive results go, as the data are being modeled as probability distributions, rather than arbitrary signals. We are excited by the range of potential influence of p7-means, and we hope this thesis is the first step toward a new perspective for the field of data-driven fMRI analysis.

BIBLIOGRAPHY

Ascher, D., P. F. Dubois, K. Hinsen, J. Hugunin, and T. Oliphant (1999). *Numerical Python* (UCRL-MA-128569 ed.). Livermore, CA: Lawrence Livermore National Laboratory.

Beckmann, C. F. and S. M. Smith (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *Medical Imaging, IEEE Transactions on 23*(2), 137–152.

Bell, A. J. and T. J. Sejnowski (1995). An information-maximization approach to blind separation and blind deconvolution. *NEURAL COMPUTATION 7*, 1129–1159.

Calhoun, V., T. Adali, L. Hansen, J. Larsen, and J. Pekar (2003). Ica of functional mri data: an overview. In *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*. Citeseer.

Calhoun, V. D. and T. Adali (2006). Unmixing fmri with independent component analysis. *Engineering in Medicine and Biology Magazine, IEEE 25*(2), 79–90.

Calhoun, V. D., J. Liu, and T. Adalı (2009). A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data. *Neuroimage 45*(1 Suppl), S163.

Erhardt, E. B., S. Rachakonda, E. J. Bedrick, E. A. Allen, T. Adali, and V. D. Calhoun (2011). Comparison of multi-subject ica methods for analysis of fmri data. *Human Brain Mapping 32*(12), 2075–2095.

Feng, Y. and G. Hamerly (2007). Pg-means: learning the number of clusters in data. In *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*, Volume 19, pp. 393. The MIT Press.

Frackowiak, R. S. J. (2004). *Human Brain Function*. Amsterdam Boston: Elsevier Academic Press.

Friston, K., A. Holmes, K. Worsley, J. Poline, C. Frith, and R. Frackowiak (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping 2*(4), 189–210.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Human brain mapping 2*(1-2), 56–78.

Hoptman, M. J., X.-N. Zuo, D. D'Angelo, C. J. Mauro, P. D. Butler, M. P. Milham, and D. C. Javitt (2012). Decreased interhemispheric coordination in schizophrenia: A resting state fmri study. *Schizophrenia Research*.

Huettel, S. (2009). fmri: Bold contrast. In E. in Chief: Larry R. Squire (Ed.), *Encyclopedia of Neuroscience*, pp. 273 – 281. Oxford: Academic Press.

Hyvärinen, A. (2013). Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371*(1984).

Jenkinson, M., P. Bannister, M. Brady, S. Smith, *et al.* (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage 17*(2), 825–841.

Jenkinson, M., C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith (2012). Fsl. *NeuroImage 62*(2), 782–790.

Jones, E., T. Oliphant, P. Peterson, *et al.* (2001–). SciPy: Open source scientific tools for Python.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics 22*(1), 79–86.

Lee, L., L. M. Harrison, and A. Mechelli (2003). A report of the functional connectivity workshop, dusseldorf 2002. *NeuroImage 19*(2), 457 – 465.

Li, Y.-O., T. Adalı, and V. D. Calhoun (2007). Estimating the number of independent components for functional magnetic resonance imaging data. *Human brain mapping 28*(11), 1251–1266.

Lindquist, M. A. (2008). The statistical analysis of fmri data. *Statistical Science 23*(4), 439–464.

Logothetis, N. K. (2008). What we can do and what we cannot do with fmri. *Nature 453*(7197), 869 – 878.

Mckeown, M. J., S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, R. S. Kindermann, A. J. Bell, and T. J. Sejnowski (1998). Analysis of fmri data by blind separation into independent spatial components. *Human Brain Mapping 6*, 160–188.

Moon, T. K. (1996). The expectation-maximization algorithm. *Signal Processing Magazine, IEEE 13*(6), 47–60.

Nagahara, Y. (2004). A method of simulating multivariate nonnormal distributions by the pearson distribution system and estimation. *Computational statistics & data analysis 47*(1), 1–29.

Pearson, K. (1916). Mathematical contributions to the theory of evolution. xix. second supplement to a memoir on skew variation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 216*(538-548), 429–457.

Poldrack, R. A. (2007). Region of interest analysis for fmri. *Social cognitive and affective neuroscience 2*(1), 67–70.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping 17*(3), 143–155.

Smith, S. M., M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, *et al.* (2004). Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage 23*, S208.

Sporns, O. (2012). From simple graphs to the connectome: Networks in neuroimaging. *NeuroImage 62*(2), 881 – 886.

Stephan, K. E., J. J. Riera, G. Deco, and B. Horwitz (2008). The brain connectivity workshops: Moving the frontiers of computational systems neuroscience. *NeuroImage 42*(1), 1 – 9.

Sun, J., A. Kaban, and J. M. Garibaldi (2010). Robust mixture modeling using the pearson type vii distribution. In *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pp. 1–7. IEEE.

Woolrich, M. W., S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, S. M. Smith, *et al.* (2009). Bayesian analysis of neuroimaging data in fsl. *Neuroimage 45*(1 Suppl), S173–S186.

Zhou, Y., M. Liang, L. Tian, K. Wang, Y. Hao, H. Liu, Z. Liu, T. Jiang, *et al.* (2007). Functional disintegration in paranoid schizophrenia using resting-state fmri. *Schizophrenia research 97*(1), 194–205.