Research Statement

Ani Eloyan

My main academic research interest is in the development of statistical methods for exploration and analysis of brain imaging data via multivariate data analysis techniques. Brain imaging data are often collected during several visits by the patients and in different modalities. Hence, the resulting datasets are not only large but also complex. There is a need for development of sound statistical methods that may be applicable for these data. My interest in statistical methods development for brain imaging data was the motivation to investigate one of the methods often used by neuroscientists - Independent Component Analysis (ICA), as a part of my PhD thesis.

Density Estimation: As I started reading the literature on ICA, I found that the problem was not well defined in that the ICA model was not fully identifiable. I proposed a set of conditions that ensure the identifiability of the model by imposing constraints on the moments of underlying source distributions. In doing so, I realized the need for a density estimation methodology for random variables with known moments, or other known characteristics such as the entropy. I developed a non-parametric density estimation method via a mixture distribution that takes into account the known structure of the moments of the density. By incorporating a sieve structure of the means and variances of elements of the mixture distribution, I was able to develop a hypothesis test based method to obtain the number of elements in the estimated mixture density. The developed density estimation method was shown effective in estimating densities of various shapes, most importantly having excellent tail behavior.

Independent Component Analysis: The proposed density estimation methodology was used to develop a likelihood based approach for ICA that takes into account the required constraints for model identifiability. The method estimates the unmixing matrix in the ICA model by a Newton-Raphson based optimization scheme while simultaneously estimating the densities of the underlying independent sources at each update of the optimization routine. The method is robust, as shown both in simulation studies as well as in applications to brain imaging data analysis.

In an effort to extend ICA methods and be able to obtain confidence intervals for the estimated parameters, I developed a novel Bayesian implementation of ICA. As opposed to the available Bayesian ICA algorithms, which are very limited by computation time and modeling assumptions, my proposed method introduces a new type of prior for the mixing matrix to take into account its full-rank structure. The densities of the underlying sources are estimated via mixture densities, with conjugate priors for the hyperparameters. The posterior densities are found analytically, while the sampling for the mixing matrix is performed via a Metropolis-Hastings algorithm.

The group ICA method is an extension of ICA to the case when data from several subjects are available. The model assumes that the structure of the underlying independent sources is the same for all subjects while the manner in which the underlying sources mix differs. Currently available methods for group ICA are limited first in the model formulation, i.e. the model is not identifiable, and secondly in memory requirements and computational scalability for modern brain imaging datasets. With the future of brain imaging data lying in large consortium aggregated datasets, it is paramount to develop scalable and robust methods for this next generation of data. As a part of my postdoctoral work, I extended my likelihood based ICA method for very large group ICA functional brain imaging studies. The important contribution of the proposed work is that the algorithm is applicable to data for any number of subjects and is completely likelihood based, while accounting for the constraints for model identifiability. The research is in review at Biostatistics.

Currently, with my postdoctoral advisor, I am advising doctoral students at Johns Hopkins Bloomberg School of Public Health (JHSPH) to apply the aforementioned group ICA method to a large cohort of subject data. The goal is to develop an atlas of ICA networks in the brain that can be used first to check whether the structure of networks obtained from the analysis of much smaller cohorts remains the same, and secondly the atlas can be used as a prior for future brain network analysis.

Brain Imaging Applications: During my postdoctoral work at the JHSPH, I spent a considerable time learning about the brain imaging data and the common problems in the analysis of these data. My first project in brain imaging was the analysis of functional magnetic resonance imaging (fMRI) data for children with ADHD. As a member of the team from the Johns Hopkins University, I participated in the ADHD 200 competition during the first year of my postdoctoral work. Brain imaging data along with several demographic variables (such as age, various measurements of IQ, gender, etc.) were provided for children from 9 data collection centers all over the world. The task was to develop methods for prediction of ADHD status in the held out test set of about 200 children by using the data for around 900 of the children in the training set. I used several methods first for blending the information from the imaging data with the demographic variables. The methods ranged from clustering techniques for imaging data, the singular value decomposition (SVD) to reduce the dimension of the connectivity matrix obtained from the imaging data, etc. The final prediction for the test set that was constructed by combining the results of four subteams of the Hopkins team was the winner of the competition. I found the competition useful for thinking in the terms of construction of prediction methods where one of the predictors is high dimensional and complex. In addition, it was curious to find that adding so much data to the simple demographic covariates may also result in damaging the prediction, because of the amount of noise introduced in the model.

A lot of my work as a postdoctoral fellow at JHSPH involved working with doctors, neuroscientists and radiologists to help them with developing feasible design for data collection and the analysis of their data. As a result of these collaborations, I appreciate the hardships involved in preparing real datasets for analysis. For example, in order to analyze the localization of lesions in the brains of patients with Multiple Sclerosis, I had to apply different linear and nonlinear registration algorithms to transform the data from 100 patients to the same template space before even starting any statistical analysis. As a result, I have developed a fair amount of technical capacity in the highly intricate domain of brain imaging data preprocessing.

Apart from the MRI imaging data, I also worked with doctors researching epilepsy via analyzing electrocorticography (ECOG) data. By discussing the data collection structure and the questions of interest with the practicing doctor, I used a logistic regression model with nested random effects for prediction of electrode activation patterns during auditory or visual tasks in patients with epilepsy.

Future research: I am currently involved in projects in methods development for brain imaging data as well as applied data analysis. My major research areas of interest are matrix factorization methods using different dimension reduction techniques depending on the structure of the data and the question of interest. I am working in collaboration with neuroscientists to develop methods for disease exploration and prediction for autism based on fMRI data along with other characteristics. I am working on not only developing models for disease prediction based on informative brain regions but also methods development for the analysis of the neurological disorder based on the fMRI connectivity matrix as a predictor as well as an outcome. As an extension to the group ICA methods development I am planning to continue working on atlas development for fMRI brain networks and continue methods development in ICA.