Discovering signals in fMRI data; a Bayesian nonparametric approach

Final project for STAT308

Ahmed Bou-Rabee, Wanrong Zhu, Zheng Xu, Mo Zhou

March 11, 2017

1 Introduction

In this paper we formulate and test a method which can be used to adaptively identify clusters of signals in functional magnetic resonance imaging (fMRI) data. Roughly, fMRI measures the change in brain blood flow associated with mental activity [HSM04]. The brain is divided into tiny blocks known as voxels, and the intensity of the blood flow over each voxel is recorded at evenly spaced time intervals while the subject is stimulated. For example, suppose researchers wanted to identify regions of the brain associated with hunger or craving. To aid in this, fMRI readings can be taken while hungry subjects are shown pictures of food.

An advantage of using fMRI is that it's a noninvasive procedure. Due to this, there are many publicly available datasets [PBM+13]. However, analyzing fMRI data poses many statistical challenges, one of them being the multiple comparisons problem. Any sampling procedure involves error. Because there are typically thousands voxels, it's likely that individual voxels may have high readings, but not be statistically significant. Identifying significant clusters (not just individual voxels) introduces an additional challenge.

In this paper, we explore a way of estimating significant clusters in fMRI using a Bayesian method. The method and some theory is presented in the first section. In the second section, we test the method on simulated data and real data.

2 Method

2.1 Generative prior

First some notation. Let $\mathcal{P} = [0, 1]$ be our sample space and let $\mathcal{D}$ represent location-time space, (for example $\mathcal{D} = \mathbf{R}^3 \times \mathbf{R}^+$). Suppose we have n data points $(p_1, d_1), \ldots, (p_n, d_n)$ where $p_i \in \mathcal{P}$ and $d_i \in \mathcal{D}$. The p-value p_i represents our belief in d_i being a null feature.

That is, our belief that the brain voxel at that point in time is not correlated with the stimulus.

Signals in fMRI data typically tend to be sparse. That is, most of the p_i are likely null features and so are drawn from a Uniform$(0, 1)$ distribution. There is a small, but unknown, number of unknown signal clusters which are located spatially in $\mathcal{D}$. Each of the p-values within each cluster is drawn from a probability distribution on $\mathcal{P}$ which places most of its mass around 0. We can encode these approximate beliefs about fMRI data in the following generative prior.

$$\begin{aligned}
\text{number of signal clusters: } & k \sim \text{Truncated Poisson}(\lambda, 1, k_{max}) \\
\text{signal centers: } & c_j \sim \text{Uniform}(\mathcal{D}) \text{ for } j = 1, \ldots, k \\
\text{signal radius: } & r_j \sim \text{Truncated Normal}(\mu, \sigma, r_{min}, r_{max}) \text{ for } j = 1, \ldots k \\
\text{signal strength: } & \beta_j \sim \text{Uniform}(\beta_{min}, \beta_{max}) \text{ for } j = 1, \ldots, k \\
\text{p-values in signal clusters: } & p_i \sim \text{Beta}(\frac{1}{\beta_j}, \beta_j), \text{when } x_i \text{ is in cluster } j \\
\text{p-values not in signal clusters: } & p_i \sim \text{Uniform}(0, 1).
\end{aligned} \tag{1}$$

For lack of time, we did not put priors on the hyper-parameters in the generative prior. Also, note that in this model, it is possible for clusters to overlap. This is an intentional choice asit allows for greater freedom in cluster shape. Furthemore, biologically, if each cluster corresponds to a role, it's plausible that regions of the brain play multiple roles, and hence are in multiple clusters.

2.2 Inventing the chain

Having established our prior, we sample from the posterior by using a method heavily inspired by the one in [Ste00]. In the aforementioned paper, Matthew Stephens constructs a Markov birth-death process to determine the number of unknown components in a mixture model. Our prior is not a mixture model, so we can't use his method directly. However, we can construct a birth-death Markov process similar to his which (assuming the generative prior) has stationary distribution given by the posterior likelihood of the data.

Some notation first: let $\mathcal{C}$ denote the set of clusters (spheres parameterized by centers and radii) and let $\mathcal{B}$ denote the set of β_i from which we parameterize the β distributions. Then, the likelihood of a set of n p-values, p^n is given by

$$L(\mathcal{C}, \mathcal{B}) = P(p^n | \mathcal{C}, \mathcal{B}) = \prod_{p_{i_j} \text{ in cluster } j} \frac{p_{i_j}^{\frac{1}{\beta_j}-1}(1-p_{i_j})^{\beta_j - 1}}{B(\frac{1}{\beta_j}.\beta_j)},$$

To construct the chain, we view each configuration of the model parameters as a point in an infinite dimensional parameter space. We construct a walk over this infinite dimensional parameter space which converges to what is known as a marked point process. A marked point process is a formal way to denote a posterior distribution over an infinite state space.

? This is a probabilistic way to denote the posterior distribution. You can think of it as the posterior likelihood.

We simulate real time using Gillespie algorithm.

2.3 Proving it works

Theorem 1. *Hey micol*

Proof. trivial □

3 Discovering signals

3.1 Toy-data

3.2 Real data

4 Extensions

Don't need to work with just p-values, can work with intensities directly by specifying appropriate priors for null distributions and for signal distributions.

Update the radius prior to allow for a vector in arbitrary dimensions instead of just one fixed.

References

[HSM04] Scott A Huettel, Allen W Song, and Gregory McCarthy. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, 2004.

[PBM+13] Russell A Poldrack, Deanna M Barch, Jason Mitchell, Tor Wager, Anthony D Wagner, Joseph T Devlin, Chad Cumba, Oluwasanmi Koyejo, and Michael Milham. Toward open sharing of task-based fmri data: the openfmri project. *Frontiers in neuroinformatics*, 7:12, 2013.

[Ste00] Matthew Stephens. Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. *Annals of statistics*, pages 40–74, 2000.