Estimating a Mean Human Connectome

William Hockeimer
Statistical Connectomics Assignment 4: March 3, 2015

Overview

The pattern of connectivity within a human brain is variable between individuals and over time within the same individual. Yet beneath this variety in connectivity there is clearly some commonality that allows each of us to have similar experiences and remain, in some sense, the same person over time. Furthermore, the way in which an individual's physiology departs from that of the 'average' person is of great clinical interest and much can be inferred about the present and future health of a person from simple metrics like weight and height distributions. Therefore it is proposed here that a statistical decision theoretic be developed to estimate the average connectome of a human being. This information will be of enormous benefit to both the basic science and medical communities. Newly discovered connectivity patterns will surely aid our understanding of how distributed neural ensembles compute and store information while having a readout of neural structure as a benchmark will serve as an invaluable diagnostic tool for clinicians in the areas of psychiatric disease, development, aging, and neurodegenerative disorders.

Sample Space

The sample space will be *G(V,E,W)* which is the set of all graphs with a fixed number of nodes across all sampled graphs, *V*, as determined by the neuroimaging technique used. The technique will be one that can assess anatomical connectivity and in particular myelin tracts, such as DTI. The edges*, E*, will be distributed according to a decision rule explained below. Each edge will have an associated weight $w \in W\ ,\ 0 \leq w \leq 1$ whose value is also set by the decision rule.

Model

The initial portion of this framework is model free in that the parameter distributions only depend on the imaging data and not any *a priori* decision about the structure of the graph. Each voxel constitutes a node in the graph-to-be and the edges are set according to a computer vision analysis of the image. Once the edges have been distributed then the nodes are clustered using an unsupervised clustering algorithm that does not need the number of clusters as an argument (see below). After clustering, each cluster will constitute a block in an stochastic block model (SBM) graph.

Action Space

The action space in this model is assignation of each edge weight $\boldsymbol{A} = [0, 1]$. The value of $\alpha_{i,j} \in \mathbf{A}$ reflects the strength of the connection assigned to the edge between nodes i and j. A strength of 0 indicates no connection whereas a strength of 1 indicates saturation of connection strength.

Decision Rule

The decision of what the value of $\alpha_{i,j}$ will be is made according to the neuroimaging data. The density of pixels corresponding to myelin between nodes i and j is proportional to the strength of the connection between them. Some threshold θ determines how few pixels will be counted as effectively 'zero' and the dynamic range of the pixel density will be scaled such that, again, α= [0,1].

Loss Function

The loss function maps a cost onto an incorrect decision made on an edge with respect to some ground truth, $\mathscr{L}$= θ x A where θ is the parameter sample space and A is the action space. The loss will be constructed so that it is an inverted Gaussian centered on the expected value of θ, i.e. the expected actual value of the edge weight.

Risk Function

The risk function from a frequentist perspective is the expected value of the loss R = $\mathbf{E}(||\mathscr{L}||)$.