

# IQ brain data analysis

Chanwoo Lee, October 4, 2020

## 1 Cross validation result

I performed 5-folded cross validation and assess log-likelihood on test dataset.

$$L(cost, rank) = \sum_{i=1}^n y_i \log \left( \hat{\mathbb{P}}(y_i = 1 | \mathbf{X}_i, cost, rank) \right) + (1 - y_i) \log \left( 1 - \hat{\mathbb{P}}(y_i = 1 | \mathbf{X}_i, cost, rank) \right).$$

The following table is the top 5 best combination of (rank, sparsity). Figure 1 shows the averaged

sparsity	rank	log_likelihood
67	1	-15.808
68	1	-15.917
67	2	-16.067
66	3	-16.215
66	2	-16.280

Table 1: Top 5 best combination of (rank, sparsity) that maximizes the log-likelihood.

log-likelihood of 5-folded cross validation across all possible combinations of (rank, sparsity). We can check cross validation is getting bad as sparsity and rank increase. From previous result, I think that rank 1 might be too small to have reasonable performance on classification. In this reason, I choose rank 2 and sparsity 60 where the log likelihood value is -23.414.

## 2 Brain IQ analysis

We already analyze the relationship between estimated probability and IQ. One another question we might have is to identify the brain connections that have influence on IQ.

First, I want to check difference of brain connections of two IQ groups. I compare the brain connections of individuals from top 5% and 95% of IQ. The procedure is as follows. I calculated the connection rates of each brain edge among top 5% and 95%.

$$\begin{aligned} \mathbf{X}_{\text{high}} &= \text{mean}(\{\mathbf{X}_i\}_{i \in G_1}), \\ \mathbf{X}_{\text{low}} &= \text{mean}(\{\mathbf{X}_i\}_{i \in G_2}), \end{aligned}$$

where  $G_1$  and  $G_2$  are index sets of individuals whose IQ are top 5% and top 95% respectively. (i,j)-th entry of  $\mathbf{X}_{\text{high}}$  means connection rates of i-th brain node and j-th brain node of individuals in high IQ group and vice versa for  $\mathbf{X}_{\text{low}}$ . Then, I checked the difference between two group.

$$\mathbf{X}_{\text{IQdiff}} = \mathbf{X}_{\text{high}} - \mathbf{X}_{\text{low}}.$$

In addition, I made two group. One group has top 5% of estimated conditional probability  $\hat{\mathbb{P}}(y_i | \mathbf{X}_i)$  and the other has top 95%. By the same way, I calculated top 5% and 95% difference of brain connections based on probability estimations. i.e.

$$\mathbf{X}_{\text{PBdiff}} = \mathbf{X}_{\text{high}} - \mathbf{X}_{\text{low}},$$

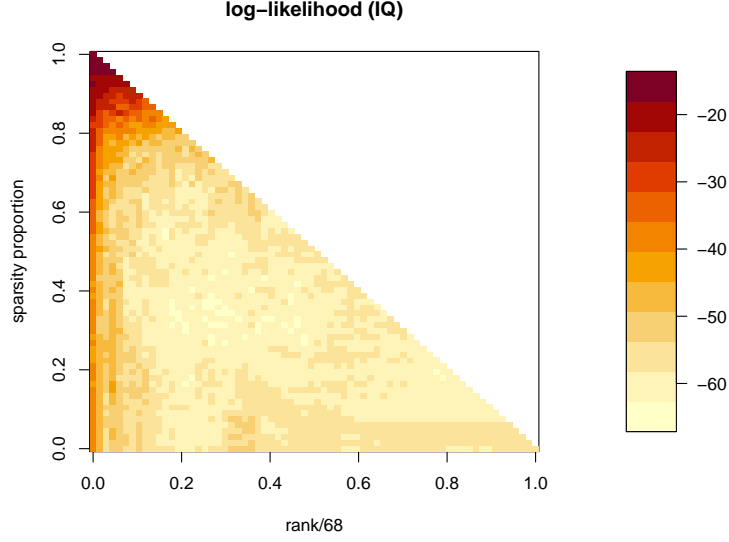


Figure 1: Averaged log-likelihood on 5-folded cross validation according to the sparsity and rank.

where  $\mathbf{X}_{\text{high}} = \text{mean}(\{\mathbf{X}_i\}_{i \in G_3})$ ,  $\mathbf{X}_{\text{low}} = \text{mean}(\{\mathbf{X}_i\}_{i \in G_4})$  and  $G_3, G_4$  are index sets of individuals whose probability is top 5% and top 95% respectively.

The following figure plots the adjacency matrices of the brain connections where  $\mathbf{X}_{\text{IQdiff}}$  ( $\mathbf{X}_{\text{PBdiff}}$ ) is greater than 0.4. We can see that high IQ group tends to have more connections between right and left hemispheres. In addition, the group difference is less noisy when individuals are divided by estimated probability estimation rather than IQ values. Figure 3 plots brain edges which have top 10 greatest entries in  $\mathbf{X}_{\text{IQdiff}}$  and  $\mathbf{X}_{\text{PBdiff}}$  respectively.

However, those analysis are just based on group differences which does not have good interpretation. To be specific, we can not interpret that those edges are important to determine high IQ of individuals. Those edges are just one trait shared among individuals in high IQ group.

I suggest two ways to identify most influential edges on IQ. First method is based on classification while the second one on probability estimation.

## 2.1 A method based on classification

When we use classification results, we can easily use the coefficient  $\mathbf{B}$  to identify the influential edges. Notice that we predict if individual is in high IQ group or not based on the following rule.

$$y = \text{sign}(\langle \mathbf{B}, \mathbf{X} \rangle + b).$$

Therefore, we can think that edges whose corresponding entries of  $\mathbf{B}$  has big magnitude are influential on IQ. Figure 4 shows matrix  $\mathbf{B}$  when used equal weight and plots top 10 edges on brain.

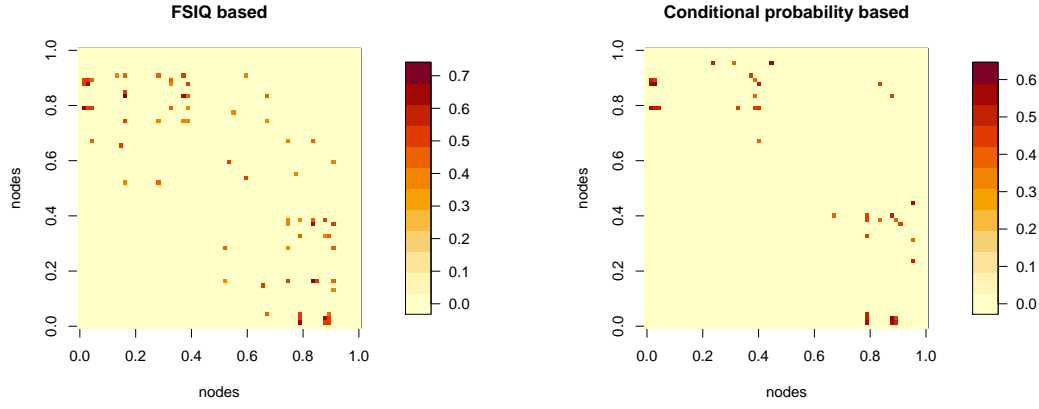


Figure 2: Figure plots  $\mathbf{X}_{\text{IQdiff}}$  (left) and  $\mathbf{X}_{\text{PBdiff}}$  (right) whose entries are greater than 0.4

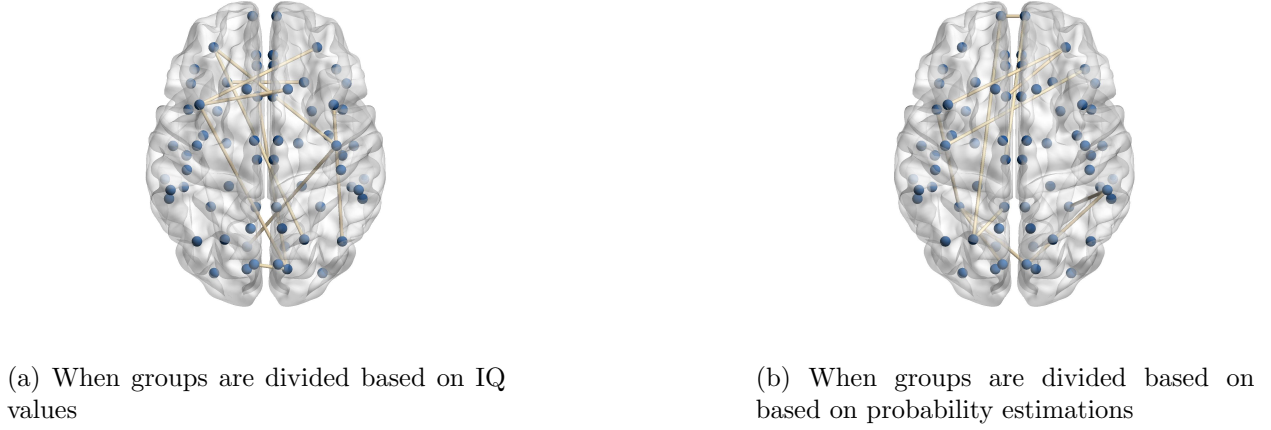


Figure 3: Top 10 brain edges that displays big difference between low and high IQ groups.

## 2.2 A method based on probability estimation

When we use probability estimation result, I suggest using elements in the basis of adjacency matrices. To be specific, define the basis of adjacency matrices as

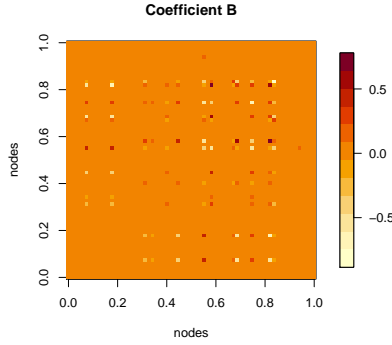
$$\mathcal{X} = \{\mathbf{A}_{[ij]}[i, j] : i \in [68], j \in [68] - [i]\}, \text{ where}$$

$$[\mathbf{A}_{[ij]}]_{pq} = \begin{cases} 1 & \text{if } (p, q) = (i, j) \text{ or } (j, i), \\ 0. & \text{otherwise} \end{cases} \text{ for all } (p, q) \in [68] \times [68].$$

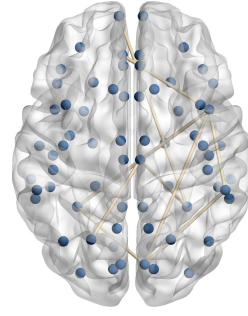
An element  $\mathbf{A}_{[ij]}$  can be interpreted as the connection between  $i$ -th brain node and  $j$ -th brain node. We calculate the conditional probability given the elements of the basis

$$\mathbb{P}(y = 1 | \mathbf{A}_{[ij]}) \text{ where } \mathbf{A}_{[ij]} \in \mathcal{X}.$$

We choose edges  $(i, j)$  as an influential edge if  $\mathbb{P}(y = 1 | \mathbf{A}_{[ij]})$  is high. Figure 5 shows (a) a probability matrix whose entry is  $\mathbb{P}(y = 1 | \mathbf{A}_{[ij]})$  and (b) the brain edges whose corresponding probability is



(a) Coefficient matrix  $\mathbf{B}$  when weight is 0.5

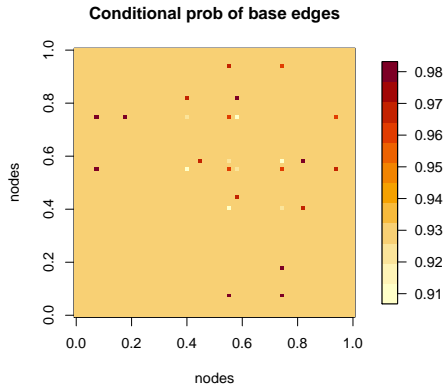


(b) Brain edges whose corresponding entry of  $\mathbf{B}$  is top 10 largest values.

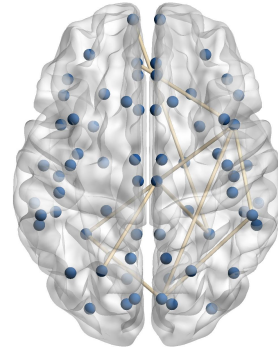
Figure 4: Identification of influential edges based on the coefficient matrix  $\mathbf{B}$ .

greater than 0.96. One numerical fact to notice is that each element of the basis  $\mathcal{X}$  has conditional probability greater than 0.91 and most of elements has the probability 0.93. One possible reason is that  $\mathbf{A}_{[ij]}$  is too sparse compared to adjacency matrix of each individual. On average, adjacency matrix of each individual is around 1600 linear combinations of the basis  $\mathcal{X}$ . Since the signal  $\mathbf{A}_{[ij]}$  is too small, it takes more weight to change predicted sign from  $y = 1$  to  $y = -1$ . But I think this issue is okay as long as our main purpose is to sort out the influential edges among elements in  $\mathcal{X}$ .

One disadvantage of this method is that it is based on the case where only one edge is considered. So we cannot explain interactions between edges.



(a) Probability matrix whose entry is  $\mathbb{P}(y = 1 | \mathbf{A}_{[ij]})$ .



(b) Brain edges whose corresponding  $\mathbb{P}(y = 1 | \mathbf{A}_{[ij]})$  is greater than 0.96.

Figure 5: Identification of influential edges based on the coefficient matrix  $\mathbf{B}$ .