580.694: HW #4

Max Collard

The "mean connectome"

Write down a statistical decision theoretic framework for finding the "mean human connectome".

There are an enormous number of valid ways of interpreting this question. Here is one plausible approach:

Sample space Each time we run our experiment, we obtain a collection of n directed graphs with d labeled vertices; considering each graph as a $d \times d$ adjacency matrix, we can identify our sample space as

$$\mathcal{S} = \left(\{0, 1\}^{d \times d} \right)^n$$

Model class First, we will assume each sample in our collection of n graphs is in fact iid. Further, we will simplify the model for each individual sample by assuming that each edge is "connected" independently as a bernoulli random variable; *i.e.*,

$$A_{ij}^{(k)} \sim \text{Bernoulli}(p_{ij})$$

Hence, we index these distributions by particular values of the "edge probability" matrix, $P = (p_{ij}) \in [0, 1]^{d \times d}$.

This gives us the distribution of samples *conditioned* on a particular choice of parameter P; *i.e.*, the above discussion defines the conditional pmf of individual brains, as a function on directed graph–space:

$$f(g \mid P)$$

To be precise within a Bayesian framework, since P is clearly not known *a priori*, we need to specify some kind of *prior* on P, *i.e.*,

$$f(g) = f(g \mid P) f(P)$$

(To improve tractability, it is probably convenient—though by no means "best"—to select a non-informative prior.)

Action space Our goal, in finding the "mean connectome", is *parameter estimation*; that is, assuming that all of our obtained data were generated using the same *true* "edge probability" matrix P, we would like to obtain a rule that converts these data into an *estimate* $\hat{P}$ of that generating matrix. Hence, our action space is the home of P,

$$\mathcal{A} = [0, 1]^{d \times d}$$

Decision rule class We know that our decision rule class must be a function taking our data to a parameter estimate, *i.e.*, $\Phi \subseteq \mathcal{A}^{\mathcal{S}}$. Since no simplifications in the form of thus function seem particularly appropriate, we'll look at the whole shebang,

$$\Phi = \mathcal{A}^{\mathcal{S}}$$

and see if we can make any *global* statements about decision rule optimality.

Loss Our loss functional, $L : \mathcal{A} \times \mathcal{A} \to \mathbb{R}^+$, should tell us how "wrong" a particular choice of parameter $\hat{P}$ is relative to the truth P^*. It is convenient to let this loss be the mean squared error of the individual *entries* of P; *i.e.*,

$$L(\hat{P}, P^*) = L\left((\hat{p}_{ij}), (p^*_{ij})\right) = \sum_{i,j} \left(p^*_{ij} - \hat{p}_{ij}\right)^2$$

Risk Consider (X, P^*) jointly distributed according to the model class described above, whose realizations live in $\mathcal{S} \times \mathcal{A}$. Then, our end goal will be to select a decision rule $F \in \Phi$ that minimizes

$$R[F] = E[L(F(X), P^*)]$$

where the expectation is taken over the entire *joint* distribution. This is important—without considering the distribution over possible Ps, depending on the structure of the rest of the problem, we may obtain solutions that are very good for true edge probability matrices P in some subset of $\mathcal{A}$, but abysmally on other parts of parameter space. (If we choose to use an informative prior, the effect of this conditioning will arise due to its implicit inclusion in the joint distribution (X, P^*) which defines the expectation operator.)

Hence, a "valid rule" $F : \mathcal{S} \to \mathcal{A}$ is a function that tells us, from an observed sample of iid human brain graphs X, a "guess" of the edge probability matrix, $\hat{P} = F(X)$. We obtain the *optimal* rule F^* by taking

$$F^* = \arg\min_{F \in \Phi} R[F]$$

Our goal, in this decision-theoretic framework, would be to rigorously obtain this optimal F^*.