STA 440: Case 1 Report #2

Jeffrey Ho, Amanda, Levenberg, Lucy Lu

September 19, 2016

Background

Through network modeling and analysis, this study aims to elucidate whether a subjects brain connectivity is associated with their level of openness, while controlling for confounding variables such as age and sex of the subject.

Model Introduction

In order to model the probability of a connection between two brain regions of a subject, we fit a generalized linear mixed effects model.

$$\begin{aligned} Pr(y_{i,u,v} = 1) &= \text{probit}(\eta_{i,u,v}) \\ \eta_{i,u,v} &= \underbrace{b_{i0} + b_{i1} age_i + b_{i2} open_i}_{\text{random effects}} + \underbrace{\beta_1 sex_i + \beta_2 \mathbb{1}(u,v)}_{\text{fixed effects}} \end{aligned}$$

The fixed effects consist of sex and an indicator for whether or not the two regions are in the same hemisphere. β_1 and β_2, therefore, are population parameters that apply to every subject and every pair of regions. The random effects consist of an intercept and the subject's age and level of openness. They are random because on one hand, we expect them to have different effects on brain connectivity for different subjects; on the other hand unlike sex and the indicator for same hemisphere, age and openness can never exhaust the population of interest. In order to infer about a new subject whose age or openness is far beyond the range represented in our data, we allow b_{i0}, b_{i1}, b_{i2}'s to be random variables drawn iid from the same distributions (population distributions) for every subject.

Data Augmentation and Gibbs Sampler

The traditional approach to fitting a binary response model uses maximum likelihood, and inferences about the model are based on asymptotic theory. However, it has been found that the MLE has significant bias for small sample sizes. Consequently, we use exact Bayesian methods to model the data via data augmentation. In essence, a latent variable $z_{i,u,v}$ is introduced for every binary response $y_{i,u,v}$, and defined as following:

$$y_{i,u,v} = \begin{cases} 1, z_{i,u,v} > 0 \\ 0, z_{i,u,v} \leq 0 \end{cases}$$
$$z_{i,u,v} \sim N(\eta_{i,u,v}, 1)$$

By doing so and applying conjugate priors on model parameters, we arrive at the following joint posterior and full conditionals:

Priors:

$$b_{ij} \overset{iid}{\sim} N(\mu_j, \Sigma_j) \ j = 0, 1, 2$$
$$\mu_j \sim N(\lambda_j, \phi_j)$$
$$\Sigma_j \sim IG(a_j, b_j)$$
$$\beta_k \sim N(\theta_k, \psi_k) \ k = 1, 2$$

Joint Posterior:

$$\pi(\underbrace{\beta_1, \beta_2, \mu_j, \Sigma_j, \sigma^2}_{\Theta}, b_{ij}, \tilde{Z}|\tilde{Y}, \tilde{X})$$
$$\propto \pi(\Theta) \prod_{i=1}^{n} \prod_{j=0}^{2} N(b_{ij}; \mu_j, \Sigma_j) \prod_{m=1}^{M} \{\mathbb{1}(z_m > 0)\mathbb{1}(y_m = 1) + \mathbb{1}(z_m < 0)\mathbb{1}(y_m = 0)\} \times N(z_m; \eta_m, 1)$$

n : total number of subjects

M : total number of (i, u, v) triads, where $u < v$

Full Conditionals:

$$\pi(\beta_k|-) \propto \pi(\beta_1) \prod_{m=1}^{M} N(D_m; \beta_k x_{mk}, 1)$$

$$\sim N(\frac{\theta_k + \psi_k \sum_{m=1}^{M} x_{mk} \cdot D_m}{1 + \psi_k \sum_{m=1}^{M} x_{mk}^2}, \frac{\psi_k}{1 + \psi_k \sum_{m=1}^{M} x_{mk}^2})$$

$$D_m = z_m - (\eta_m - x_{mk})$$

$$\pi(b_{ij}|-) \sim N(\frac{\mu_j + \Sigma_0 \sum_{l=1}^{L} D_l}{1 + \mu_j \sum_{l=1}^{L} x_{lj}^2}, \frac{\mu_j}{1 + \mu_j \sum_{l=1}^{L} x_{lj}^2})$$

$$D_l = z_l - (\eta_l - x_{lj})$$

L : total number of (u, v) dyads, $u < v$, when i is fixed

$$\pi(\mu_j|-) \propto \pi(\mu_j) \prod_{i=1}^{n} N(b_{ij}; \mu_j, \Sigma_j)$$

$$\sim N(\frac{\lambda_j \Sigma_j + \phi_j \sum_{i=1}^{n} b_{ij}}{\Sigma_j + n\phi_j}, \frac{\phi_j \Sigma_j}{\Sigma_j + n\phi_j})$$

$$\pi(\Sigma_j|-) \propto \pi(\Sigma_j) \prod_{i=1}^{n} N(b_{ij}; \mu_j, \Sigma_j)$$

$$\sim IG(a_j + \frac{n}{2}, b_j + \frac{1}{2} \sum_{i=1}^{n} (b_{ij} - \mu_j)^2)$$

$$\pi(z_{i,u,v}|-) \sim N(\eta_{i,u,v}, 1) \begin{cases} \text{truncated at the left by 0 if } y_{i,u,v} = 1 \\ \text{truncated at the right by 0 if } y_{i,u,v} = 0 \end{cases}$$

Gibbs Sampler:

1. Sample $\mu_j^{(0)}, \Sigma_j^{(0)}$ from their respective priors.
2. Sample $\beta_k^{(0)}, b_{ij}^{(0)}$ from their respective priors.
3. Calculate $\eta_{i,u,v}^{(0)}$ and draw $z_{i,u,v}^{(0)}$ from $N(\eta_{i,u,v}^{(0)}, 1)$
4. For $t = 1 : T$, iteratively sample $\mu_j^{(t)}, \Sigma_j^{(t)}, \beta_k^{(t)}, b_{ij}^{(t)}, z_{i,u,v}^{(t)}$ from their respective full conditionals.