Lab 10: Linear Regression
STA 360/602

Rebecca C. Steorts

March 19, 2018

Introduction

The topic of this lab is linear regression. It follows exercise 9.1 in Hoff very closely. Open the data file with the following code:

```
swim <- read.table("~/.../swim.dat", sep=' ')
```

replacing the ellipses with the appropriate directory name. The file contains times (in seconds) of four high school swimmers swimming 50 yards. There are 6 times for each student, taken every two weeks. Each row corresponds to a swimmer and a higher column index indicates a later date.

Task 1

We will fit a separate linear regression model for each swimmer, with swimming time as the response and week as the explanatory variable. Let $Y_i \in \mathbb{R}^6$ be the 6 recorded times for swimmer i. Let

$$X_i = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ \dots & \\ 1 & 9 \\ 1 & 11 \end{bmatrix}$$

be the design matrix for swimmer i. Then we use the following linear regression model:

$$Y_i \sim \mathcal{N}_6 \left(X\beta_i, \tau_i^{-1}\mathcal{I}_6 \right)$$
$$\beta_i \sim \mathcal{N}_2 \left(\beta_0, \Sigma_0 \right)$$
$$\tau_i \sim \text{Gamma}(a, b).$$

Derive full conditionals for β_i and τ_i.

Task 2

Complete the prior specification by choosing a, b, β_0, and Σ_0. Let your choices be informed by the fact that times for this age group tend to be between 22 and 24 seconds.

Task 3

Code a Gibbs sampler to fit each of the models. For each swimmer i, obtain draws from the posterior predictive distribution for y_i^*, the time of swimmer i if they were to swim two weeks from the last recorded time.

Task 4

The coach has to decide which swimmer should compete in a meet two weeks from the last recorded time. Using the posterior predictive distributions, estimate $\Pr\left\{y_i^* = \max\left(y_1^*, y_2^*, y_3^*, y_4^*\right)\right\}$ for each swimmer i and use these probabilities to make a recommendation to the coach.