Exercise 2: Linear Regression

Jason Doll

June 8, 2018

Objective

This execise will expand on exercise 1 by introducing linear regression as an extension of the simple model of the mean using the Stan programing language. This exercise will provide practice at coding regression models in Stan.

Background

In this example, we will randomly generate data for a response variable, y, with known parameters and predictors, x. The R code to initialize workspace, generate data, and package and send the data with initial values to Stan are provided in the "Ex2_LReg.R" file located in the "Ex2_LReg" folder. The instructor will review this R file with you. Your task will be to create the Stan file using the code provided below. You are encouraged to type rather than copy/paste the code into a new .stan document. The model you will fit to individual observations i is:

$$
\begin{aligned}
\text{Model:} \quad & \mathrm{y}_i = \alpha + \beta \mathrm{X}_i + \epsilon_i \\
& \epsilon_i \sim \text{Normal}(0, \sigma) \\
\text{Priors:} \quad & \alpha \sim \text{Normal}(0, 100) \\
& \beta \sim \text{Normal}(0, 100) \\
& \sigma \sim \text{half-cauchy}(0, 5)
\end{aligned}
$$

R packages required for this exercise

1. rstan

Directions

Create a new text file and save it as "Ex2_est_lr.stan" in the "Ex1_mean" folder. Enter the Stan code below in the new .stan file.

Stan code

```
data {
  int<lower=0> n;  //number of observations
  vector[n] x;     //observed x values, predictors
  vector[n] y;     //observed y values, response
}

parameters {
  real<lower=0> sigma; //standard deviation
  real alpha;          //y-intercept
  real beta;           //slope
```

```
}

model {
  //reference priors
  alpha ~ normal(0,100);
  beta ~ normal(0,100);
  sigma ~ cauchy(0,5);

  //likelihood, loop through number of observations
  for(i in 1:n){
    y[i] ~ normal(alpha + beta * x[i], sigma);
  }
}
```

R Code

Open "Ex1_mean.R" and run lines 1 through the lines:

```
#send everything to Stan
fit2 <- stan(file = 'Ex2_est_lr.stan',
             data = dataList ,
             init = initslst,
             chains = nchains,
             iter = 1000 ,
             warmup = 500 ,
             thin = 1 )
```

Traceplots, parameter estimates and plots of results can be viewed using the code below (note this code is provided):

```
#View traceplots
traceplot(fit2)
#view results
fit2

#extract results
est_alpha=rstan::extract(fit2,"alpha")$alpha
est_beta=rstan::extract(fit2,"beta")$beta
est_sd=rstan::extract(fit2,"sigma")$sigma

#plot results
par(mfrow=c(1,3))
hist(est_alpha,breaks=50);abline(v=a,lwd=5);
hist(est_beta,breaks=50);abline(v=b,lwd=5);
hist(est_sd,breaks=50);abline(v=sd,lwd=5);
par(mfrow=c(1,1))
```