Assignment 4.b for **STATGR6103**

submitted to Professor Andrew Gelman

Advait Rajagopal

3 October 2016

1 Question 1

The file `http://www.stat.columbia.edu/ gelman/bda.course/dogs.txt` contains the data for the first 25 trials on each of the 30 dogs in the experiment. Occurrence of shock is indicated by S, non-occurrence by a blank. (The number assignments to the dogs were made by Solomon and Wynne and do not imply that a selection has been made.)

1.1 Part A

Create a probabilistic learning model for these data. The model should reflect the increases in probability of avoidance with practice as jumping is reinforced by shock and avoidance. The model would give the probability of avoidance as it depends on trial number and previous experience. You also might want to allow the parameters to vary by dog

1.1.1 Logistic Regression Model

In the first instance while trying to model probability of avoidance, I identify it is a probability and therefore it should lie between 0 and 1. This suggests that I can use a logistic regression.

However the sequential nature of shocks in trials is very important to the model. For one it is because all dogs were *definitely* shocked in the first trial and secondly, since the goal is to identify learning we would expect to see a decline in the number of shocks or an increase in the number of avoidances as trials increase. It is important to keep in mind that a logistic regression will assign non-zero probabilities to both outcomes.[1] I set up the following logistic regression model as a starting point for the probability that dog i, $i = (1, .., 30)$ in trial t, $t = (1, .., 25)$ gets a shock.

$$Pr(y_{it} = 1|\beta) = logit^{-1}(\beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it)}$$

where $y_{it} = 1$ for a shock and 0 for an avoidance. X_{1it} counts the number of avoidances and X_{2it} counts the number of shocks received until that trial. The goal is to estimate 3 parameters $(\beta_0, \beta_1, \beta_2)$, so I use a noninformative uniform prior distribution. From the regression results in Table 1 we observe that $|\beta_1| > |\beta_2|$ meaning that past avoidances have a greater impact on predicting the probability of an avoidance.

[1] see Gelman et.al [2000], at `http://www.stat.columbia.edu/ gelman/research/published/dogs.pdf`

Table 1: Posterior estimates of β's

	mean	2.5%	25%	50%	75%	97.5%
β_0	1.82	1.37	1.67	1.82	1.97	2.26
β_1	-0.36	-0.43	-0.38	-0.36	-0.34	-0.29
β_2	-0.21	-0.30	-0.24	-0.21	-0.18	-0.13

1.1.2 Logarithmic Regression Model

Here we use a logarithmic model. This model is an improvement over the logistic model because it does not have an intercept since as the probability of a shock is 1 at the first trial.

$$Pr(y_{it} = 1|\beta) = exp(\beta_{1i}X_{1it} + \beta_{2i}X_{2it)}$$

with all the variables and parameters the same as defined in section 1.1.1. We observe that the β's have been suffixed with an i and this allows us to account for variation between dogs or in other words we allow the parameters to vary by dog. Given that I provide priors for the distribution of β's as well as estimate dog level β's I explore the results of the logarithmic regression in detail later in the assignment.

1.2 Part B

Set up a weakly informative prior distribution.

Here I create a full hierarchical model that will estimate the parameters β_{1i} and β_{2i} as having their own distributions governed by further hyperparameters. All the knowledge we have about the experiment is actually made explicit in setting up these priors. For starters we know that the values of β_{1i} and β_{2i} have to be negative as intuitively we believe that as shocks and avoidances increase the probability of avoidance should go up, meaning the dogs will learn. The following are the priors;

$$\beta_{1i} \sim N(\mu_1, \tau_1)$$

$$\beta_{2i} \sim N(\mu_2, \tau_2)$$

$$\mu_1 \sim N(0, 5)$$

$$\mu_2 \sim N(0, 5)$$

I set a wide hyperprior on μ's allowing them to have quite a high variance parameter associated with them as well as I model τ with a noninformative hyperprior so as to not constrain the model too with very strong priors. The intuition behind τ is that it captures the difference between dogs. If the variance τ is very large then all dogs are different from each other and if it is small it means the dogs are more or less similar even if they aren't identical.

1.3 Part C

Use Stan to get draws from the posterior distribution and discuss the fit of the model to the data (weaknesses and strengths).

The results of the hierarchical Bayesian regression are shown in Table 2.

Table 2: Hyperparameters

	mean	2.5%	25%	50%	75%	97.5%	R-hat
μ_1	-0.31	-0.34	-0.32	-0.32	-0.30	-0.25	1.02
μ_2	-0.09	-0.19	-0.09	-0.08	-0.07	-0.06	1.08
τ_1	0.03	0.00	0.01	0.02	0.03	0.12	1.11
τ_2	0.02	0.00	0.01	0.02	0.03	0.05	1.12

The distribution of the "avoidance" and "shock" coefficients (β_{1i}, β_{2i}) are shown in Figure 1.

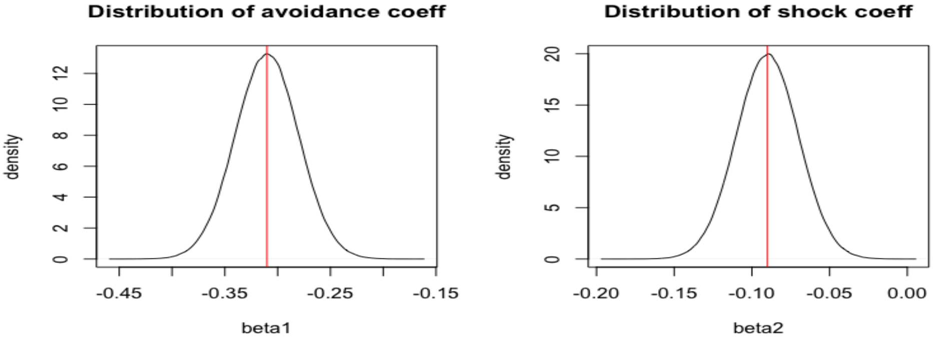


Figure 1: Distribution of parameters

The mean values of (μ_1, μ_2) are (-0.31 , -0.09) respectively with a very small associated standard deviation (τ_1, τ_2) = (0.03 , 0.02). Thus $exp(\beta_{1i})$ and $exp(\beta_{2i})$ have an average value of (0.73,0.91) indicating that an avoidance or a shock multiplies the estimated or predicted probability of a shock by a factor of 0.73 or 0.91 respectively. Again we see that the coefficient associated with avoidance is larger (in absolute terms) than the one associated with shocks. The τ parameters denoting variance are quite small and this indicates most dogs are similar to each other with small between-dog variations. Figure 2 shows the distribution of the mean and standard deviation of the predicted shocks based on draws from the posterior distribution.

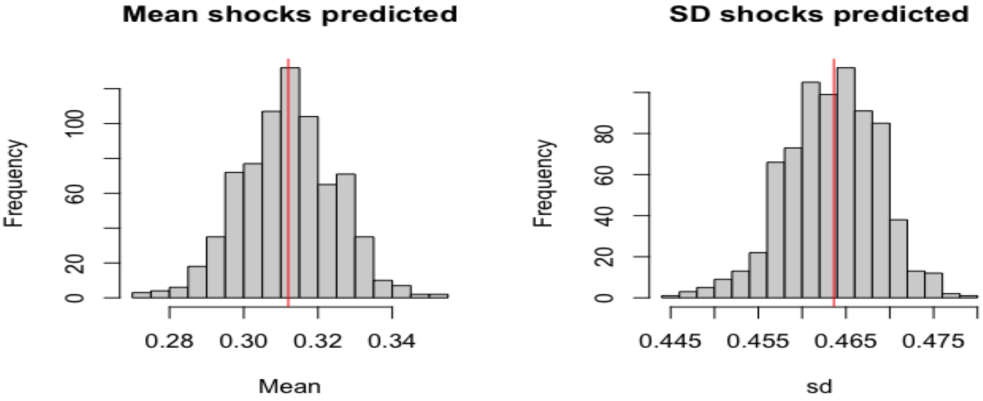


Figure 2: The red line in (a) and (b) shows the true mean and standard deviation of the boolean shock variable as predicted by the model.

In order to evaluate the fit of this model we observe if the mean and standard deviation of shocks as predicted by the model are centered around the original mean and standard deviation of the boolean shock variable.
Another interesting way to check if the predictive model fits the discrete data is to plot proportion of avoidances for incremental trials.

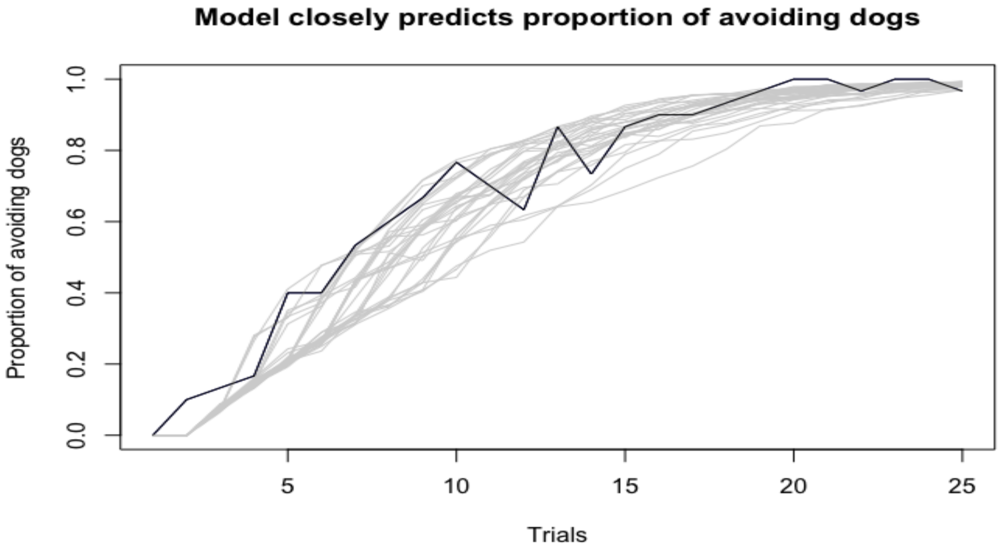


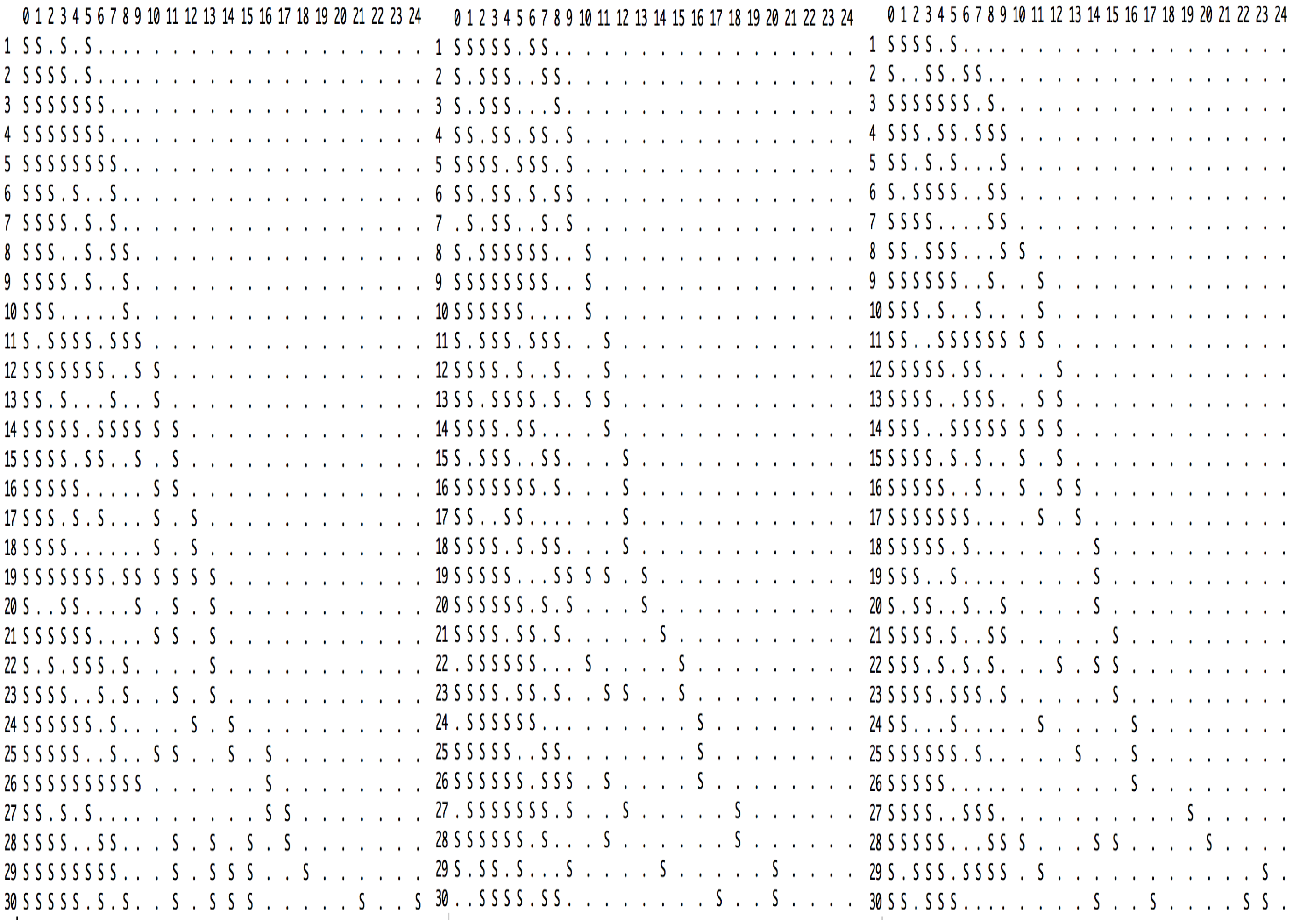
Figure 3: The black line shows the proportion of avoidance in dogs for the true data and the gray lines show the sets of replicated data.

1.4 Part D

Simulate a set of fake data from your model (that is, create a table of simulated data from 30 new dogs). Display, on a single page, the data from the real and the fake dogs. Discuss the similarities and differences between the real data and the simulated data.

I simulate a fake dataset based on the underlying parameters and display the raw data, logistic regression data and logarithmic regression data side by side. The “S” stands for a shock and the “.” stands for an avoidance by the dog. Figure 4 shows the results of 30 dogs undergoing 25 trials each.

The logarithmic model seems to do most justice to the original dataset for two reasons. The first is that in the logarithmic regression the first column has a probability of 1 that a dog was shocked. This is not the case in the logit model. The second is that the shocks (like the actual experiment) seem to be concentrated more at the beginning and start to wane towards the last few trials indicating that the dogs learn as the experiment progresses which reaffirms my understanding in Figure 3 that probability of avoidance increases as trials increase.



```
   0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
1  SS.S.S...................
2  SSSS.S...................
3  SSSSSSS..................
4  SSSSSSS..................
5  SSSSSSSS.................
6  SSS.S..S.................
7  SSSS.S.S.................
8  SSS..S.SS................
9  SSSS.S..S................
10 SSS.....S................
11 S.SSSS.SSS...............
12 SSSSSSS..SS..............
13 SS.S...S..S..............
14 SSSSS.SSSSSS.............
15 SSSS.SS..S.S.............
16 SSSSS.....SS.............
17 SSS.S.S...S.S............
18 SSSS......S.S............
19 SSSSSSS.SSSSSS...........
20 S..SS....S.S.S...........
21 SSSSSS....SS.S...........
22 S.S.SSS.S....S...........
23 SSSS..S.S..S.S...........
24 SSSSSS.S....S.S..........
25 SSSSS..S..SS..S.S........
26 SSSSSSSSSS......S........
27 SS.S.S..........SS.......
28 SSSS..SS...S.S.S.S.......
29 SSSSSSSS...S.SSS..S......
30 SSSSS.S.S..S.SSS.....S..S
```

```
   0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
1  SSSSS.SS.................
2  S.SSS..SS................
3  S.SSS...S................
4  SS.SS.SS.S...............
5  SSSS.SSS.S...............
6  SS.SS.S.SS...............
7  .S.SS..S.S...............
8  S.SSSSSS..S..............
9  SSSSSSSS..S..............
10 SSSSSS....S..............
11 S.SSS.SSS..S.............
12 SSSS.S..S..S.............
13 SS.SSSS.S.SS.............
14 SSSS.SS....S.............
15 S.SSS..SS...S............
16 SSSSSSS.S...S............
17 SS..SS......S............
18 SSSS.S.SS...S............
19 SSSSS...SSSS.S...........
20 SSSSSS.S.S...S...........
21 SSSS.SS.S.....S..........
22 .SSSSSS...S....S.........
23 SSSS.SS.S..SS..S.........
24 .SSSSSS.........S........
25 SSSSS..SS.......S........
26 SSSSSS.SSS.S....S........
27 .SSSSSSS.S..S.....S......
28 SSSSSS.S...S......S......
29 S.SS.S...S....S.....S....
30 ..SSSS.SS........S..S....
```

```
   0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
1  SSSS.S...................
2  S..SS.SS.................
3  SSSSSSS.S................
4  SSS.SS.SSS...............
5  SS.S.S...S...............
6  S.SSSS..SS...............
7  SSSS....SS...............
8  SS.SSS...SS..............
9  SSSSSS..S..S.............
10 SSS.S..S...S.............
11 SS..SSSSSSSS.............
12 SSSSS.SS....S............
13 SSSS..SSS..SS............
14 SSS..SSSSSSSS............
15 SSSS.S.S..S.S............
16 SSSSS..S..S.SS...........
17 SSSSSSS....S.S...........
18 SSSSS.S.......S..........
19 SSS..S........S..........
20 S.SS..S..S....S..........
21 SSSS.S..SS.....S.........
22 SSS.S.S.S...S.SS.........
23 SSSS.SSS.S.....S.........
24 SS...S.....S....S........
25 SSSSSS.S.....S..S........
26 SSSSS...........S........
27 SSSS..SSS..........S.....
28 SSSSS...SSS...SS....S....
29 S.SSS.SSSS.S...........S.
30 SS.SSS........S..S....SS.
```

Figure 4: (a) Shows the raw data from the original experiment. (b) Shows the predicted values of shock from a logistic regression. (c) Shows the predicted values from a logarithmic regression.

2 Code

2.1 R Code

Listing 1: R Code

```
# #Read in data
rm(list = ls())
#
library(rstan)
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
##
setwd("/Users/Advait/Desktop/New␣School/Fall16/BDA/Class7")
dogs <- read.table(
"http://www.stat.columbia.edu/˜gelman/bda.course/dogs.txt",skip = 2)
names(dogs) <- c("dog", 0:24)
str(dogs)
dogs1 <- dogs[,-c(1)]
dogs1 <- ifelse(dogs1 == "S",1,0)
dogs1
##
q <- matrix(data = NA,nrow=dim(dogs1)[1],ncol=dim(dogs1)[2])
for (i in 1:dim(dogs1)[1]){
  for(j in 1:dim(dogs1)[2]){
    q[i,j] <- sum(dogs1[i, 1:c(j - 1)])
  }
}
q[,1] <- rep(0, 30)
###
id <- matrix(data = NA, nrow = dim(dogs1)[1], ncol=dim(dogs1)[2])
for(i in 1:dim(dogs1)[1]){
  id[i,] <- rep(i,25)
}
##
prev.trials <- rep(0:24,30)
grand_snoop <- as.data.frame(
  cbind(c(t(dogs1)),
  c(t(q)),
  prev.trials,
  c(prev.trials - t(q)),
  c(t(id)))
)
```

Listing 2: R code contd.

```
##Creating the Grand Snoop Matrix for all further operations
grand_snoop[1:5,]
names(grand_snoop) <- c(”bool.shock”,”prev.shock”,”prev.trials”,
                        ”avoidance”,”dog.id”)
grand_snoop[1:5,]

y <- grand_snoop$bool.shock
x1 <- grand_snoop$avoidance
x2 <- grand_snoop$prev.shock
id <- grand_snoop$dog.id

#Logistic_fit
stanc(”logistic_4b.stan”)
fit1 <- stan(”logistic_4b.stan”,
             data = list(”y”,”x1”,”x2”,”id”,”y_rep”),
             iter = 1000,
             chains = 3)
print(fit1)

#Logarithmic_fit
stanc(”logarithmic_4b.stan”)
fit2 <- stan(”logarithmic_4b.stan”,
             data = list(”y”,”x1”,”x2”,”id”,”y_rep”),
             iter = 1000,
             chains = 3)
print(fit2)
ext2 <- extract(fit2)
ext1 <- extract(fit1)
#####################
#Plots
####################
##Histogram
y_rep2 <- ext2$y_rep
y_rep2_mean <- c()
y_rep2_sd <- c()
for (i in 1:750){
  y_rep2_mean[i] <- mean(y_rep2[i,])
  y_rep2_sd[i] <- sd(y_rep2[i,])
}
par(mfcol = c(1,2))
hist(y_rep2_mean,breaks = 15, col = ”lightgray”,
     xlab = ”Mean”, main = ”Mean␣shocks␣predicted”)
abline(v=mean(y), col = ”red”, cex = 2)
```

Listing 3: Rcode contd.

```
hist(y_rep2_sd,breaks = 15, col = ”lightgray”,
     xlab = ”sd”, main = ”SD␣shocks␣predicted”)
abline(v=sd(y), col = ”red”, cex = 1.2)

##Proportion of Avoidance
par(mfcol = c(1,1))
avoid <- 1-apply(as.matrix(dogs1), 2, mean)
avoid_prop <- matrix(colMeans(ext1$y_rep), ncol = 25, nrow = 30
                     ,byrow = T)
avoid_prop <- cbind(rep(1,30), avoid_prop)
##
par(mfcol = c(1,1))
plot(1:25, avoid, type = ”l”,col = ”darkblue”,
     ylab = ”Proportion␣of␣avoiding␣dogs”,
     xlab = ”Trials”,
     main = ”Model␣closely␣predicts␣proportion␣of␣avoiding␣dogs”
     )
for(i in 1:30){
  lines(1:25,(1 - avoid_prop[i,c(1:25)]),
        type = ”l”,
        col = ”lightgray”, cex = .6)}
lines(1:25, avoid, type = ”l”)
###

##Ordering and displaying
avoid_new <-ext2$y_rep

avoid_new_sample <- matrix(avoid_new[190,],
                           ncol = 25, nrow = 30, byrow = T)

foo <- data.frame(dogs1)
names(foo) <- 0:24

ordshock <- c()
for(i in 1:30){
  ordshock[i] <- max((1:25)[foo[i,]==1])
}
raw_order <- foo[order(ordshock),]
row.names(raw_order) <- NULL
raw_order
```

Listing 4: R code contd.

```
ordshock_pred <-c()
for(i in 1:30){
  ordshock_pred[i] <- max((1:25)[avoid_new_sample[i,]==1])
}
pred_order <- avoid_new_sample[order(ordshock_pred),]
row.names(pred_order) <- NULL
pred_order

avoid_bad <- ext1$y_rep
avoid_bad_sample <- matrix(avoid_bad[190,],
                           ncol = 25, nrow = 30, byrow = T)
ordshock_bad <-c()
for(i in 1:30){
  ordshock_bad[i] <- max((1:25)[avoid_bad_sample[i,]==1])
}
bad_order <- avoid_bad_sample[order(ordshock_bad),]
row.names(bad_order) <- NULL
bad_order

# Display
rawdata <- as.data.frame(
  ifelse(raw_order == 1, ”S”, ”.”),quote = F)
baddata <- as.data.frame(
  ifelse(bad_order == 1, ”S”, ”.”),quote = F)
names(baddata) <- c(0:24)
fakedata <- as.data.frame(
  ifelse(pred_order ==1, ”S”, ”.”)[,1:25], quote = F)
names(fakedata) <- c(0:24)
print(rawdata)
print(fakedata)
print(baddata)

##Density of coefficients
par(mfcol = c(1,2))
newb1 <- rnorm(10ˆ6, -0.31, 0.03)
newb2 <- rnorm(10ˆ6, -0.09, 0.02)
plot(density(newb1), main = ”Distribution␣of␣avoidance␣coeff”,
     xlab = ”beta1”, ylab = ”density”)
abline(v = mean(newb1),col = ”red”)
plot(density(newb2), main = ”Distribution␣of␣shock␣coeff”,
     xlab = ”beta2”, ylab = ”density”)
abline(v = mean(newb2), col = ”red”))
```

2.2 Stan Code

Listing 5: Logistic Regression [full multilevel model commented out]

```
data{
  int y[750];
  int x1[750];
  int x2[750];
  #int id[750];
}
parameters{
  real b0;
  real b1;
  real b2;
  #real<upper = 0> b1[30];
  #real<upper = 0> b2[30];
  // real mu1;
  // real mu2;
  // real<lower = 0> tau1;
  // real<lower = 0> tau2;
}
model{
  // b1 ˜ normal(mu1,tau1);
  // b2 ˜ normal(mu2,tau2);
  // mu1 ˜ normal(0,5);
  // mu2 ˜ normal(0,5);
//   for (i in 1:30){
//   for (n in 1:750){
//   y[n] ˜ bernoulli_logit(b0 + b1[id[i]]*x1[n] + b2[id[i]]*x2[n]);
for(i in 1:750){
  y[i] ˜ bernoulli_logit(b0 + b1*x1[i] + b2*x2[i]);
}
// }
// }
// }
// generated quantities{
//   real<lower = 0, upper=1> y_rep[750];
//   for (i in 1:30){
//   for (n in 1:750){
//     y_rep[n] = bernoulli_rng(inv_logit(b0 + b1[id[i]]*x1[n] + b2[id[i]]*x2[n]));
  //   }
 // }
// }
```

Listing 6: Logarithmic Regression

```
data{
  int y[750];
  int x1[750];
  int x2[750];
  int id[750];
  }
parameters{
  real<upper = 0> b1[30];
  real<upper = 0> b2[30];
  real mu1;
  real mu2;
  real<lower = 0> tau1;
  real<lower = 0> tau2;
}
model{
  b1 ~ normal(mu1,tau1);
  b2 ~ normal(mu2,tau2);
  mu1 ~ normal(0,5);
  mu2 ~ normal(0,5);
  for (i in 1:30){
  for (n in 1:750){
   target += bernoulli_lpmf(y[n]|exp(b1[id[i]]*x1[n] + b2[id[i]]*x2[n]));
}
}
}
generated quantities{
  real<lower = 0, upper=1> y_rep[750];
  for (i in 1:30){
  for (n in 1:750){
    y_rep[n] = bernoulli_rng(exp(b1[id[i]]*x1[n] + b2[id[i]]*x2[n]));
    }
    }
}
```