

Assignment1

Format: R.rmd and .docx/.doc files (compress into a zip file).

Deadline: 3:00am, 2019/3/2.

Submission: CEIBA

If you don't have "rethinking" package, please run the code below.

```
install.packages(c("coda","mvtnorm","devtools","loo"))
library(devtools)
devtools::install_github("rmcelreath/rethinking")
```

The below data indicate the gender (male=1, female=0) of officially reported first and second born children in 100 two-child families.

```
birth1 <- c(1,0,0,0,1,1,0,1,0,1,0,0,1,1,0,1,1,0,0,0,1,0,0,0,1,
            0,0,0,0,1,1,1,0,1,0,1,1,1,0,1,0,1,1,0,1,0,0,1,1,0,
            1,0,0,0,0,0,0,0,1,1,0,1,0,0,1,0,0,0,1,0,0,1,1,1,1,
            0,1,0,1,1,1,1,0,0,1,0,1,1,0,1,0,1,1,1,0,1,1,1,1)

birth2 <- c(0,1,0,1,0,1,1,1,0,0,1,1,1,1,0,0,1,1,1,0,0,1,1,1,
            0,1,1,1,0,1,1,1,0,1,0,0,1,1,1,1,0,0,1,0,1,1,1,1,
            1,1,1,1,1,1,1,1,1,1,0,1,1,0,1,1,0,1,1,0,0,0,0,
            0,0,1,0,0,0,1,1,0,0,1,0,0,1,1,0,0,0,1,1,0,0,0,0)
```

So for example, the first family in the data reported a boy (1) and then a girl (0). The second family reported a girl (0) and then a boy (1). The third family reported two girls.

Use these vectors as data. So for example to compute the total number of boys born across all of these births, you could use: `sum(birth1) + sum(birth2)`

Question 1

Using **grid approximation**, compute the posterior distribution for the probability of a birth being a boy. Assume a **uniform prior probability**. Which parameter value maximizes the posterior probability?

Question 2

Using the `sample()` function, draw 10,000 random parameter values from the posterior distribution you calculated above. Use these samples to estimate the 50%, 89%, and 97% highest posterior density intervals. (hint: use `mode_hdi()` in `tidybayes` packages)

Question 3

Use `rbinom` function to simulate 10,000 replicates of 200 births. You should end up with 10,000 numbers, each one a count of boys out of 200 births. Compare the distribution of predicted numbers of boys to the actual count in the data (111 boys out of 200 births). There are many good ways to visualize the simulations, but the `dens` command (part of the `rethinking` package) is probably the easiest way in this case. Does it look like the model fits the data well? That is, does the distribution of predictions include the actual observation as a central, likely outcome?

Question 4

Now compare 10,000 counts of boys from 100 simulated first borns only to the number of boys in the first births, `birth1`. How does the model look in this light?

Question 5

The model assumes that sex of first and second births are independent. To check this assumption, focus now on second births that followed female first borns. Compare 10,000 simulated counts of boys to only those second births that followed girls. To do this correctly, you need to count the number of first borns who were girls and simulate that many births, 10,000 times. Compare the counts of boys in your simulations to the actual observed count of boys following girls. How does the model look in this light? Any guesses what is going on in these data?