

# Assignment 2

Julia Haaf & Nicole Cruz

## Exercise 1: Check for parameter recovery in a linear model using simulated data.

Generate some simulated independent and identically distributed data with  $n = 100$  data points as follows:

```
y <- rnorm(100, mean = 500, sd = 50)
```

Next, fit a simple linear model with a normal likelihood:

$$y_n \sim \text{Normal}(\mu, \sigma) \quad (1)$$

Specify the following priors:

$$\begin{aligned} \mu &\sim \text{Uniform}(0, 60000) \\ \sigma &\sim \text{Uniform}(0, 2000) \end{aligned} \quad (2)$$

Generate posterior predictive distributions of the parameters and check that the true values of the parameters  $\mu = 500, \sigma = 50$  are recovered by the model. What this means is that you should check whether these true values lie within the range of the posterior distributions of the two parameters. This is a good sanity check for finding out whether a model can in principle recover the true parameter values correctly.

## Exercise 2: Chocolate Bars

Imagine your favorite chocolate bar is called *Nutty Buddy Bites*. Because you buy Nutty Buddies so often, you notice that they seem to be lighter than just a few years ago even though the weight on the package hasn't changed. As a good statistician, you decide to test whether your suspicions are warranted. For the sake of research, you buy 100 Nutty Buddies and a good scale, and you weigh each of the chocolate bars. You save the final data set as `nutties.csv`.

- The packaging says that the bars are supposed to weigh 50g. Using normal distributions, choose priors that represent **your** assumptions/beliefs about chocolate bar weight. To think about a reasonable set of priors for  $\mu$  and  $\sigma$ , you should come up with your own subjective assessment about what you think a reasonable range of values can be for  $\mu$  and how much variability might happen. There is no correct answer here.
- Fit the model with just a few iterations, say 50 iterations (set warmup to the default of 25, and use four chains). Does the model converge?
- Now fit the model properly with sufficient iterations. Visualize the posteriors.

### Exercise 3: Other priors

- a. Can you come up with very informative priors that influence the posterior in a noticeable way (use normal distributions for priors, not uniform priors)? Again, there are no correct answers here; you may have to try several different priors before you can noticeably influence the posterior.
- b. Generate and plot prior predictive distributions based on this prior and plot them.
- c. Generate posterior predictive distributions based on this prior and plot them.

### Exercise 4: Your visualization set

Generate a script with some useful visualizations for your personal Bayesian workflow. Include Prior and Posterior prediction, visualization of the posterior distributions, and anything else you find useful. Also note which visualizations can be reused on a regular basis and which depend a lot on the model/data at hand.