SCI 2025: Homework 8

Setup

In this homework, we will once again work with data from:

Nettle, D. (1998). Explaining global patterns of language diversity. *Journal of Anthropological Archaeology*, 17:354–74.

First, load the data into your R session.

```
library(rethinking)
data(nettle)
head(nettle)
```

```
     country num.lang    area  k.pop num.stations mean.growing.season
1    Algeria       18 2381741  25660          102                6.60
2     Angola       42 1246700  10303           50                6.22
3  Australia      234 7713364  17336          134                6.00
4 Bangladesh       37  143998 118745           20                7.40
5      Benin       52  112622   4889            7                7.14
6    Bolivia       38 1098581   7612           48                6.92
  sd.growing.season
1              2.29
2              1.87
3              4.17
```

```
4               0.73
5               0.99
6               2.50
```

The meaning of each column in the dataset is given below:

(1) country: Name of the country
(2) num.lang: Number of recognized languages spoken
(3) area: Area in square kilometers
(4) k.pop: Population, in thousands
(5) num.stations: Number of weather stations that provided data for the next two columns
(6) mean.growing.season: Average length of growing season,in months
(7) sd.growing.season: Standard deviation of length of growing season,in months

You should use quadratic approximation via `rethinking::ulam()` for all model fitting.

Question 1:

Revisit the models from Homework 2, where we tried to predict `num.lang` as a function of `k.pop`. But this time, use something other than a Gaussian distribution for the likelihood. Write down the format definition of the model, including the likelihood, priors, and any link functions (if you are using one). The choice is yours, but please explain your reasoning and assumptions.

Question 2:

Perform a prior predictive check for the model you defined in Question 1. Assess both the prior predictive distribution of the outcome, as well as the prior function relating `k.pop` to `num.lang`.

Question 3:

Fit the model you have defined in Question 1 (and perhaps refined after performing the prior predictive check) using `ulam()`. Evaluate the fit model using standard MCMC diagnostics. Do you notice any problems? How might you address them?

Question 4

Perform a posterior predictive check for the model you fit in Question 3, like the prior check in Question 2. Assess the predictive adequacy of your model and the implied relationship between `k.pop` and `num.lang`.

Question 5

Fit a new model that uses a different likelihood function, link funciton, or both. You may find it makes sense to change your priors as well.

Perform posterior predictive checks for the new model. How does it compare to your initial model? If appropriate, use WAIC or PSIS to compare the models. Can you explain the results?