STA365 Assignment 2

Daniel Simpson

Assignment instructions and marking information

This should be submitted as a *single PDF* via Quercus by 12pm (Midday) on 16 April, 2021. The assignment should be produced using RMarkdown, R, and Stan.

- Each task is worth 10%
- All code for performing each task *must* be present. If there is no code you will lose most marks.
- Each task requires some reflective writing. Make sure you write things! In particular, there is no single solution to this task, so the justification and interpretation of the steps is important! Two otherwise identical solutions with different justifications and interpretations could receive very different marks.
- Each task requires the production of several plots and figures. These plots should be cleanly laid out with adequate captions, markings, labels, and legends.

This is an assignment that uses real data. As such, you should be prepared for the model to possibly be a poor fit for the data and you should discuss this if it occurs.

Modelling animal movement

A key task in ecology is understanding how environmental conditions affect animal behaviour. In this assignment we will look at how environmental factors change how raptors move.

Raptors are large birds that cover a lot of ground by gliding. This requires some specific environmental conditions, so it's plausible to imagine that a raptor would have two types of movement state:

- A low energy state, where the bird is preening or resting and not moving around a lot; and
- A high engery state, where the bird is moving around a lot.

There is no way to know whether or not a raptor is in a high or low energy state at any given period of time.

An appropriate measure of movement is the minimum specific acceleration (MSA), which is given for a single raptor in the attached data. The correct scale to consider MSA on is the log-scale.

The data

The data can be read in as

```r
dat <- read.csv("Verreauxs.accel.txt", sep = "\t")
head(dat)
```

```
##             date_time        msa wind_speed saws_temp hrseg
## 1 2013-04-16 13:03:44 0.23787533        3.6      16.5     1
## 2 2013-04-16 13:05:38 0.09222497        3.6      16.5     1
## 3 2013-04-16 13:07:27 0.27394695        3.6      16.5     1
## 4 2013-04-16 13:37:51 0.28776082        3.6      16.5     5
## 5 2013-04-16 13:39:44 0.59848213        3.6      16.5     5
## 6 2013-04-16 13:41:36 0.12466199        3.6      16.5     5
```

There are severaral variables in the data but the relevant variables are:

- MSA: our observation
- wind_speed: The wind speed (m/s)
- saws_temp: Temperature (celsius)

The model

Given an indicatior of state Z, where $Z = 1$ corresponds to the low energy state, and $Z = 2$ corresponds to the high energy state, we model the log-MSA as

$$y_i \mid Z, \mu, \sigma \sim N(\mu_Z, \sigma_Z^2).$$

We further assume that for each observation, the corresponding state Z_i is an iid draw from a binomial distribution with $\Pr(Z = 2) = p_i$, which might vary by individual.

The three tasks in this assignment will help you explore the task of fitting this model.

Task 1: Fitting a simple model to simulated data

Simulate data from a mixture of two normal distributions with $p_i = p$ (a constant, but unknown, parameter) and write a Stan model to fit the data.

- You can use any experimental set up you want, but you should justify your choices in writing.
- You should do some basic prior predictive checks.
- You should do some posterior predictive checks and comment on how the model fit, including any problems you had with the sampler.

Task 2: Fitting the data

Fit two different 2 component mixture model to the given data. The first model should have a fixed p, while the second model should have the probability of being in state 2 as

$$\text{logit}(p_i) = \beta_0 + \beta_1 \text{wind} + \beta_2 \text{temp}.$$

- Write Stan code to fit these model
- Use posterior predictive checks to comment on how well both models fit.
- Use leave-one-out crossvalidation to select the best model and interpret the choice.

Task 3: Spurious components

For this task, keep p constant as in Task 1.

Imagine that we are uncertain if there are 2 or 3 different energetic states. In this case, one option would be to fit a 3 component model with a prior that puts a lot of mass on 1 and 2 component models.

Design a prior for a 3 component model that has support on lower-dimensional models. (You might want to think about how we built the horseshoe prior). Explain and justify why this prior should work.

Fit the model to two *simulated* data sets:

- One where the data is simulated from a model with 2 components
- One where the data is simulated from a model with 3 component

Comment on how the model fits in both scenarios. Are there computational problems? Is the spurious component ignored?