Questions for the Course on Bayesian Modeling

EJ Wagenmakers

Below are sets of questions that you'll be graded on. In order to answer these questions correctly and pass the course you will find it necessary to work through the chapters that the questions pertain to.

Exercises Chapter 4

Print out your answers to these exercises and bring them with you this Wednesday. Before you concentrate on the exercises first work through chapter 4.

1. Suppose you test a single participant in a speeded response time task such as lexical decision. In lexical decision, participants have to indicate whether visually presented letter strings are words (e.g., "chair") or not (e.g., "drapa"). Let's assume we want to model the observed response times. The problem is that response time (RT) distributions are not Gaussian distributed. Instead, RT distributions are skewed to the right. These are some example data:

```
rt = c(0.4784271, 0.2202520, 0.3707721, 0.3464157, 0.4314385,
       0.4666542, 0.3745648, 0.4288913, 0.3332206, 0.2950294,
       0.5386618, 0.4706629, 0.9419108, 0.3068890, 0.5461369,
       0.4929777, 0.4978749, 0.4867891, 0.4194809, 0.5215591,
       0.3739309, 0.4520722, 0.4226023, 0.4249337, 0.3152922,
       0.4905963, 0.2404994, 0.3914402, 0.3963014, 0.3150337,
       0.4365492, 0.7323715, 0.4522957, 0.5313994, 0.6953040,
       0.3795496, 0.5265442, 0.5015609, 0.5327790, 0.4710306,
       0.4494928, 0.2243310, 0.4472178, 0.3922798, 0.4826039,
       0.3616635, 0.4348760, 0.6843270, 0.3025242, 0.2939238)

par(cex.main = 1.5, mar = c(5, 6, 4, 5) + 0.1, mgp = c(3.5, 1, 0),
    cex.lab = 1.5, font.lab = 2, cex.axis = 1.3, bty = "n", las=1)
xlow  = 0
xhigh = 2
hist(rt, xlim=c(xlow,xhigh),lwd=2, ylab="Density", xlab="RT", main="")
rug(rt)
```

There are two ways in which you can address the problem of data that are clearly not appropriate for the model that you want to use. Method 1 is to change something about the data, and method 2 is to use a different model. Choose one method and apply it to the RT data.

2. Check your modeling results by generating posterior predictive values from your model and comparing these to the RT data.

3. For bonus points: I need an interesting, funny, or relevant example of data that **cannot** be analyzed by classical methods. What I am looking for is a situation that meets the following requirements: (a) data come in over time, without a sampling plan or an experimenter being able to control the process (many natural processes are like this); (b) the relevant statistic is a correlation, so it involves two continuous variables; (c) the null hypothesis of zero correlation needs to be somewhat believable, and so does the hypothesis of the presence of a correlation. An example is the correlation between the relative height of the US presidential candidates and their percentage of the popular vote (there is some evidence that taller candidates attract more votes). Critically, the election results some in over time, without an experimenter; the relative height and the proportion of the votes are continuous variables; and both the presence and the absence of a correlation between these variables are believable a priori.