7 Introduction to Bayesian statistics

This chapter introduces Bayesian statistics, Markov chain Monte Carlo (MCMC) techniques, and integrated nested Laplace approximations (INLA). We will keep the explanations simple and conceptual. However, some sections do contain some mathematics. We marked them with an asterisk '*'. If you are not interested in the underlying mathematics, then you can skip these sections, or just read the summaries in the 'owl notes'.

Prerequisite for this chapter: A working knowledge of R and linear regression is required.

7.1 Why go Bayesian?

The use of Bayesian techniques can be motivated in different ways. Some scientists start by arguing that they have prior knowledge, and that this prior knowledge should be incorporated into the models. In Chapter 3 we analysed osprey data; eggshell thickness was modelled as a function of DDD. You don't need to be a scientist in order to know that a breakdown product of a pesticide is bad for ospreys. Why would you not use this knowledge (which essentially translates as a negative slope for DDD) in the models? Using prior knowledge during the analysis then immediately leads to other scientists criticising Bayesian-based approaches because (according to them) using prior knowledge means that the models are not objective anymore.

Another angle of motivating Bayesian approaches is to first criticise frequentist approaches and then show how useful the output from Bayesian techniques is. The criticism is about the interpretation of frequentist confidence intervals and *p*-values. Their interpretation goes via a statement like 'if we were to repeat the experiment a large number of times, then in 5% of the cases we would expect to find an even larger *t*-value'. This is a statement based on fictive data. In reality we are not repeating an experiment. What we would like to say from our results is that there is a 95% probability that a regression parameter is between *a* and *b*. This is wishful thinking for a frequentist scientist, but it is reality for a Bayesian analyst.

Figure 7.1 contains the so-called posterior distribution of a regression parameter. Pay special attention to the word 'distribution' in the previous sentence. In a Bayesian analysis we assume that the parameters are unknown stochastic quantities, and we estimate their distribution using the data (resulting in a picture like Figure 7.1). In a Bayesian analysis we end up with the probability of a parameter given the data (written as $P(\beta \mid \text{data})$), whereas in a frequentist analysis we look at the probability of the data given the parameters (written as $P(\text{data} \mid \beta)$). In other words, in a frequentist analysis we assume there is only one β and the analysis gives us an estimate plus a 95% confidence interval. The $P(\text{data} \mid \beta)$ tells us how likely (or unlikely) the data are, given the betas. In a Bayesian analysis there is no fixed value for β.

Both arguments for using Bayesian statistics will probably not convince the scientist with a frequentist background, who has limited time available, doesn't like programming, and sees colleagues publishing papers with *p*-values smaller than 0.05. And Bayesian analysis doesn't even give you *p*-values!

We decided to base our motivation to convince the reader to adopt Bayesian techniques on another argument: You have no choice. The reason that you are reading this book is most likely because you have data with a spatial and / or temporal dependency structure. The packages in R that can cope with such data are rather limited. To take full advantage of spatial and temporal models we need tools that allow us to fit such models. At the time of this writing the majority of these tools require Bayesian statistics.

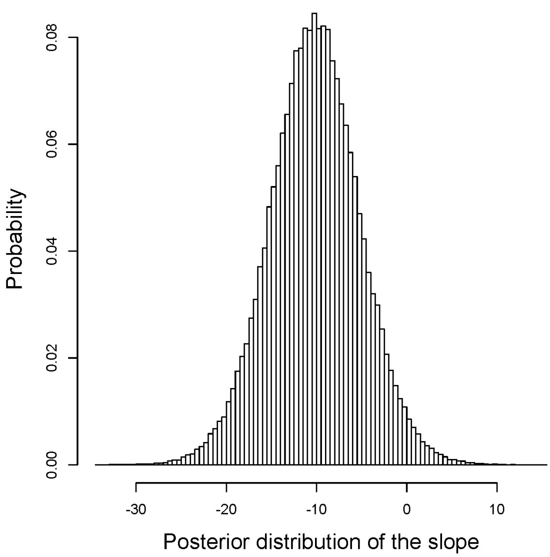

Figure 7.1. Posterior distribution of the regression parameter β.

7.2 General probability rules

We begin by reviewing some basic probability rules. Let $P(A)$ be the probability of an event A and $P(B)$ the probability of an event B. Define the joint probability $P(A \text{ and } B)$ as the probability that events A and B both occur. The following basic probability rule holds.