One-parameter models

A one-parameter model is a class of sampling distributions that is indexed by a single unknown parameter. In this chapter we discuss Bayesian inference for two one-parameter models: the binomial model and the Poisson model. In addition to being useful statistical tools, these models also provide a simple environment within which we can learn the basics of Bayesian data analysis, including conjugate prior distributions, predictive distributions and confidence regions.

3.1 The binomial model

Happiness data

Each female of age 65 or over in the 1998 General Social Survey was asked whether or not they were generally happy. Let \( Y_i = 1 \) if respondent \( i \) reported being generally happy, and let \( Y_i = 0 \) otherwise. If we lack information distinguishing these \( n = 129 \) individuals we may treat their responses as being exchangeable. Since 129 is much smaller than the total size \( N \) of the female senior citizen population, the results of the last chapter indicate that our joint beliefs about \( Y_1, \ldots, Y_{129} \) are well approximated by

- our beliefs about \( \theta = \sum_{i=1}^{N} Y_i / N \);
- the model that, conditional on \( \theta \), the \( Y_i \)'s are i.i.d. binary random variables with expectation \( \theta \).

The last item says that the probability for any potential outcome \( \{y_1, \ldots, y_{129}\} \), conditional on \( \theta \), is given by

\[
p(y_1, \ldots, y_{129}|\theta) = \theta^{\sum_{i=1}^{129} y_i} (1 - \theta)^{129 - \sum_{i=1}^{129} y_i}.
\]

What remains to be specified is our prior distribution.

A uniform prior distribution

The parameter \( \theta \) is some unknown number between 0 and 1. Suppose our prior information is such that all subintervals of \([0, 1]\) having the same length also have the same probability. Symbolically,

\[
\Pr(a \leq \theta \leq b) = \Pr(a + c \leq \theta \leq b + c) \quad \text{for} \quad 0 \leq a < b < b + c \leq 1.
\]

This condition implies that our density for \( \theta \) must be the uniform density:

\[
p(\theta) = 1 \quad \text{for all} \quad \theta \in [0, 1].
\]

For this prior distribution and the above sampling model, Bayes’ rule gives

\[
p(\theta|y_1, \ldots, y_{129}) = \frac{p(y_1, \ldots, y_{129}|\theta)p(\theta)}{p(y_1, \ldots, y_{129})}
= p(y_1, \ldots, y_{129}|\theta) \times \frac{1}{p(y_1, \ldots, y_{129})}
\propto p(y_1, \ldots, y_{129}|\theta).
\]

The last line says that in this particular case \( p(\theta|y_1, \ldots, y_{129}) \) and \( p(y_1, \ldots, y_{129}|\theta) \) are proportional to each other as functions of \( \theta \). This is because the posterior distribution is equal to \( p(y_1, \ldots, y_{129}|\theta) \) divided by something that does not depend on \( \theta \). This means that these two functions of \( \theta \) have the same shape, but not necessarily the same scale.

Data and posterior distribution

- 129 individuals surveyed;
- 118 individuals report being generally happy (91%);
- 11 individuals do not report being generally happy (9%).

The probability of these data for a given value of \( \theta \) is

\[
p(y_1, \ldots, y_{129}|\theta) = \theta^{118}(1 - \theta)^{11}.
\]

A plot of this probability as a function of \( \theta \) is shown in the first plot of Figure 3.1. Our result above about proportionality says that the posterior distribution \( p(\theta|y_1, \ldots, y_{129}) \) will have the same shape as this function, and so we know that the true value of \( \theta \) is very likely to be near 0.91, and almost certainly above 0.80. However, we will often want to be more precise than this, and we will need to know the scale of \( p(\theta|y_1, \ldots, y_n) \) as well as the shape. From Bayes’ rule, we have

\[
p(\theta|y_1, \ldots, y_{129}) = \theta^{118}(1 - \theta)^{11} \times p(\theta)/p(y_1, \ldots, y_{129})
= \theta^{118}(1 - \theta)^{11} \times 1/p(y_1, \ldots, y_{129}).
\]