STAT40380/STAT40390/STAT40850 Bayesian Analysis

Dr Niamh Russell

School of Mathematics and Statistics
University College Dublin

niamh.russell@ucd.ie

January 2016

Last time we introduced Bayes' Theorem for events A and B:

Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

However, this is not a Bayesian Analysis yet. Bayesian statistics uses an interpretation of Bayes' theorem.

- Suppose we replace event A with a quantity about which we want to learn. We call these quantities *parameters*.
- Furthermore, suppose we replace the event B with quantities that we already know. We call these quantities *data*.

The left hand side of the equation will now read as *the probability distribution of the parameters given the data*.

Bayes' Theorem for parameters and data

If we let the parameters be denoted by θ and the data by x, Bayes' Theorem now reads as:

Updated version

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

Notes:

- Each of these items can be considered to be a probability distribution (rather than the probability of an event as we had before).
- $p(x)$ is often difficult to calculate (as we shall see), and relies only on the data we've already observed so the theorem is often written more simply as a proportionality statement:

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

Prior, posterior and likelihood

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

Each of these elements has an important name:

- The term on the left hand side is the probability distribution of the parameters given the data and is known as the posterior distribution.
- The term $p(x|\theta)$ is the probability distribution of the data given the parameters. You should have seen this before - it is known as the *likelihood*.
- The term $p(\theta)$ is the probability distribution of the parameters θ. It is known as the *prior distribution*. It represents knowledge about the parameters before the data are observed.

Bayes' theorem is thus often written as:

$$\textit{posterior} \propto \textit{likelihood} \times \textit{prior}$$

Example 1: The Normal distribution

Example

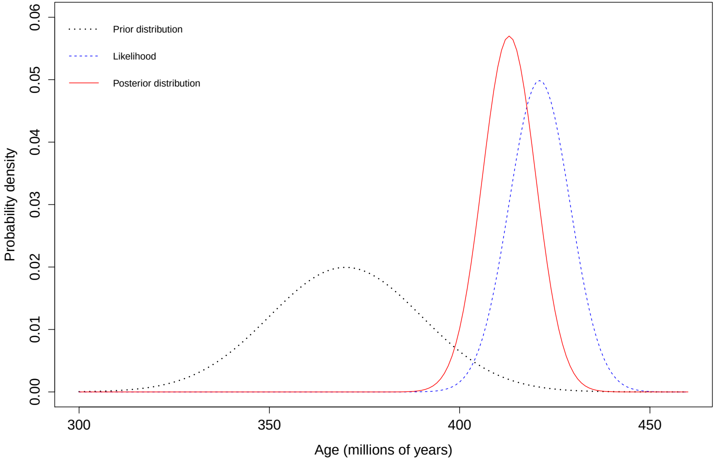
Suppose we observe a data point as x and assume that it follows a normal distribution with mean θ and known variance ϕ. Suppose also that we have some prior information that θ is normally distributed with mean θ_0 and variance ϕ_0 both known. Find the posterior of θ given x.

Example 2: Dating volcanic rock

Example

In the early 1960s, the first dates were obtained of a volcanic rock via the K/Ar method. The results suggested an age of 370 ± 20 million years. In the 1970s, measurements were made on the same rock via the Rb/Sr method, giving an estimated age of 421 ± 8 million years. If the errors are represented as standard deviations, and the distributions can be assumed to be normal, what is the estimated posterior age of the rock?

Example 2: Dating volcanic rock (cont.)



Notes 1

- If we observe two independent data points x_1 and x_2, and require $p(\theta|x_1, x_2)$, we have:

$$\begin{aligned} p(\theta|x_1, x_2) &\propto p(x_1, x_2|\theta)p(\theta) \\ &= p(x_1|\theta)p(x_2|\theta)p(\theta) \\ &\propto p(\theta|x_1)p(\theta|x_2) \end{aligned}$$

  so the posterior is sequentially updated, first by observation x_1, and then by observation x_2.

- For this reason, Bayes' theorem is often known as a *learning algorithm* as it allows us to update our knowledge with each new data point.
- The sequential updating can be summarised by the phrase *"today's posterior is tomorrow's prior"*.
- Alternatively, if x_1 and x_2 are observed simultaneously we can simply update in terms of the product of the likelihoods.

Notes 2

- If we have multiple observations $x_1, \ldots, x_n$, our posterior becomes:

$$
\begin{aligned}
p(\theta|x_1, \ldots, x_n) &\propto p(x_1, \ldots, x_n|\theta)p(\theta) \\
&= \prod_{i=1}^{n} p(x_i|\theta)p(\theta).
\end{aligned}
$$

- When dealing with lots of observations we will write $\mathbf{x}$ (or $\vec{x}$ on the board) to signify $x_1, \ldots, x_n$.
- Finally, when working with Bayes' theorem and the normal distribution, it is convenient to work in terms of the precision rather than the variance. Remember that precision = 1/variance.
- BUGS always works with the precision.

Example

Suppose (as in example 1) that we have a prior distribution for θ as $\theta \sim N(\theta_0, \phi_0)$ with θ_0 and ϕ_0 known, but that this time we have n observations where $x_i \sim N(\theta, \phi)$ (with ϕ known). What is the posterior distribution for θ?

Example 4:Chest measurements

Example

A body of data suggests that the chest measurements of men (measured in inches) are normally distributed with mean 38 and variance 9. A large study of size 10,000 was then undertaken giving a mean of 39.8 and a standard deviation of 2. What is the posterior estimate for mean chest measurement?