STAT40380/STAT40390/STAT40850
Bayesian Analysis

Dr Niamh Russell

School of Mathematics and Statistics
University College Dublin

niamh.russell@ucd.ie

February 2016

Prior distributions

- If we are Bayesian statisticians, we must always specify a prior distribution.
- Today we will discuss which types of prior distribution might be appropriate in different situations.
- Recall that for the $N(\theta, \phi)$ case with fixed ϕ and prior $\theta \sim N(\theta_0, \phi_0)$, we had $\theta | x \sim N(\theta_2, \phi_2)$ where:

$$\phi_2 = \left(\frac{1}{\phi_0} + \frac{n}{\phi} \right)^{-1}, \; \theta_2 = \frac{\phi / n}{\phi_0 + \phi / n} \theta_0 + \frac{\phi_0}{\phi_0 + \phi / n} \bar{x}$$

- Notice that as n increases, the variance of the posterior decreases and the posterior mean is influenced more by the observed mean.
- In such situations, we say that the likelihood is dominant. When information on θ is weak, this may be a desirable property.

Improper densities

- Suppose we wanted to design a prior distribution that had no effect on the posterior.
- In the above situation, we would require $\phi_0 = \infty$, which would lead to a prior distribution of $N(\theta_0, \infty)$.
- If we can envisage such a distribution, it would be uniform over the whole real line, ie.

$$p(\theta) = \kappa, \quad (-\infty < \theta < \infty)$$

- In fact, we can allow such cases because of the proportionality constraint in Bayes' theorem.
- We call these *improper prior distributions*.
- An improper prior distribution does not have the usual property $\int p(x)dx = 1$.

More on improper priors

- It turns out that we can still get a valid posterior distribution even when the prior is improper.
- This property holds provided we have a valid likelihood function, because the likelihood will dominate the prior.
- (Consider $\lim_{\phi_0 \to \infty} \theta_2$ and $\lim_{\phi_0 \to \infty} \phi_2$ in the examples we have used.)
- Not all improper prior distributions have to be flat. Another improper density we will find useful later on is

$$p(\theta) = \frac{\kappa}{\theta} \quad (0 < \theta < \infty)$$

- This prior is particularly useful on precision or variance parameters of the normal distribution

Example 1: an improper prior on the variance

Example

Suppose that $x_i \sim N(\theta, \phi)$ with θ known. Find the posterior of $\phi|\mathbf{x}$ when $p(\phi) \propto \frac{1}{\phi}$.

Example 2: Rat's weight again

Example

Find the posterior distribution of the variance of the rats' weight data under the improper prior of Example 1. Summary statistics from the data are:

$$n = 20, \ \bar{x} = 21, \ \sum(x_i - \bar{x})^2 = 664.$$

Even more on improper priors

- We often want to try models with different priors (ie different shapes, proper or improper) to determine the sensitivity to the prior assumptions. If two people have strongly differing views about a parameter, we might like to try a neutral prior which is dominated by the likelihood.
- In many cases, we are conducting an experiment to significantly increase our knowledge about a parameter, and so it makes sense to let the likelihood dominate.
- *Warning: an improper flat prior on a parameter θ suggests that we genuinely believe that θ is equally likely across a large range of values. This is often a poor assumption.*

Priors and transformation

- Suppose θ is a random variable between 0 and 1, and we have no prior information about its value
- If the prior is flat, proper $U(0,1)$ (ie $p(\theta) = 1$ for $0 < \theta < 1$), then we have:
$$p(\theta|x) \propto p(x|\theta)$$
- so the posterior is equivalent to the likelihood, after normalisation.
- Suppose we now define $\phi = 1/\theta$.
- By the change of variable rule we know
$$p(\phi) = p(\theta)\frac{d\theta}{d\phi} = \frac{1}{\phi^2}$$
(note: still a proper density)
- But we now have $p(\phi|x) \propto p(x|\phi)p(\phi)$, with the prior distribution providing some information that ϕ is small.
- We need to be careful when assuming flat prior distributions, as they may not be flat for transformed values of the parameters

Sufficiency in Bayesian statistics

Some definitions

- A *statistic* $t(\mathbf{x})$ is a function of the *data only*
- (Note: $t(\mathbf{x})$ may be a scalar or a vector)
- A statistic is a *sufficient statistic* if it provides all the information we need to learn about any parameters in which we are interested
- More fully, suppose we have:

$$p(\mathbf{x}|\theta) = p(t|\theta)p(\mathbf{x}|t, \theta)$$

- If it is the case that $p(\mathbf{x}|t, \theta) = p(\mathbf{x}|t)$, ie that $p(\mathbf{x}|t)$ does not depend on θ, we say that t is sufficient for θ.
- If t^* is found to be a function of all other sufficient statistics then it is termed *minimally sufficient*
- Proofs of sufficiency in the Lee book.

Example 3: sufficiency for the normal distribution

Example

Show that $\sum x_i$ is sufficient for θ when the data x_i are distributed as $N(\theta, \phi)$ with ϕ known.

Bayesian statistics and sufficiency

Some other examples

- If $x_i \sim N(\theta, \phi)$ with θ known then $\sum(x_i - \theta)^2$ is minimally sufficient for ϕ.
- If $x_i \sim P(\lambda)$ then $\sum x_i$ is minimally sufficient for λ
- If $x_i \sim Ga(\alpha, \beta)$ then both $\prod x_i$ and $\sum x_i$ are jointly sufficient for α and β

- In a Bayesian situation, we only need to observe the sufficient statistics to form our posterior, not the entire dataset
- This is because all the information in the likelihood is stored in the sufficient statistics