UNIVERSITY OF EDINBURGH
SCHOOL OF MATHEMATICS
BAYESIAN DATA ANALYSIS, 2017/2018, SEMESTER 2
VANDA INÁCIO & KEN NEWMAN

Practical 1: Conjugate Bayesian inference in R

1. **Analysis of binomial data: drug data**. Consider the example from lecture 1 where a new drug is being considered for relief of chronic pain, with the success rate θ being the proportion of patients experiencing pain relief. In the past, drugs of this type have shown variable pain relief rates, with a mean of 40% and a standard deviation of 10%. We have seen that these could be translated into a Beta$(9.2, 13.8)$ distribution. This drug had 15 successes out of 20 patients.

   (i) Calculate the posterior distribution of the success rate θ.

   (ii) What is the posterior mean and 95% highest posterior density (HPD) interval for the response rate.
   Hint: For computing the HPD interval you can use, for instance, the function `hpd` from the `R` package `TeachingDemos`.

   (iii) Compute a symmetric 95% credible interval. Compare this to the 95% HPD interval.

   (iv) What is the probability that the true success rate is greater than 0.6.

   (v) How is this value affected if a uniform prior is adopted? And how is it affected in case of adopting Jeffreys' prior?

   (vi) Using the original Beta$(9.2, 13.8)$ prior, suppose 40 more patients were entered into the study. What is the chance that at least 25 of them experience pain relief? *Hint*: You might want to use the `beta` and `gamma` functions implemented in `R`.

   (vii) We might ask whether the observed data is 'compatible' with the expressed prior distribution. One method is to calculate the predictive probability of observing such an extreme number of successes under this prior: this is a standard p-value but where the null hypothesis is a distribution. Use the predictive distribution for 20 future patients to find the probability of getting at least 15 successes (i.e., at least 15 patients experiencing pain relief). Do you think this suggests the data are incompatible with the prior?

   (viii) Check for prior/data conflict by making the prior/likelihood/posterior plot.

2. Suppose that most drugs (95%) are assumed to come from the stated Beta$(9.2, 13.8)$ prior, but there is a small chance that the drug might be a 'winner'. 'Winners' are assumed to have a prior distribution with mean 0.8 and standard deviation 0.1.

   (i) What Beta distribution might represent the 'winners' prior? Remember that a Beta(a, b) distribution has mean $a/(a+b)$ and variance $ab/\{(a+b)^2(a+b+1)\}$.

   (ii) Plot the mixture prior.

(iii) What is now the chance that the response rate is greater than 0.6?
Hint: You might start by showing that if

$$\theta \sim \pi\text{Beta}(a_1, b_1) + (1-\pi)\text{Beta}(a_2, b_2),$$

then

$$\theta \mid y \sim \omega_1\text{Beta}(a_1 + y, b_1 + n - y) + (1-\omega_1)\text{Beta}(a_2 + y, b_2 + n - y),$$

where

$$\omega_1 = \pi \frac{B(a_1 + y, b_1 + n - y)}{B(a_1, b_1)} \left(\pi \frac{B(a_1 + y, b_1 + n - y)}{B(a_1, b_1)} + (1-\pi) \frac{B(a_2 + y, b_2 + n - y)}{B(a_2, b_2)} \right)^{-1}.$$

Here y denotes the number of successes.

(iv) For this mixture prior, repeat the prior/data compatibility test performed previously. Are the data more compatible with this mixture prior?

(v) Check for prior/data conflict by making the prior/likelihood/posterior plot.

3. **Analysis of normal data: systolic blood pressure**. Suppose we are interested in the long-term systolic blood pressure (SBP), in mmHg, of a particular 60-year old female. We take two independent readings of her SBP 6 weeks apart, giving values of 127 and 133. Each measurement is assumed to be normally distributed around her underlying long-term SBP θ with standard deviation $\sigma = 5$.
We have additional information: a population survey revealed females aged 60 had a mean long-term SBP of 120 with standard deviation 10.

(i) Use the information from the population survey to specify a normal prior for the woman's mean SBP.

(ii) What is the posterior mean and 95% symmetric credible interval for the woman's SBP? Compare this with the maximum likelihood estimate and 95% confidence interval.

(iii) Suppose 2 other readings were taken, both of 130. What would be the 95% credible interval now?