Assignment 2 (2 pages)

1. Suppose we observe y successes in n trials where the probability of success in each trial is θ and suppose we use a $Beta(\alpha, \beta)$ prior for θ. Show that the posterior mean can be written as a weighted average of the prior mean and the observed proportion of successes, meaning that the posterior mean will always fall between those two values.

2. Suppose we observe y successes in n trials where the probability of success in each trial is θ.

   (a) Prove that if we choose a $Beta(1, 1)$ (Uniform) prior then the posterior variance will be smaller than the prior variance.

   (b) Show that the above isn't necessarily the case if we choose a general $Beta(\alpha, \beta)$ prior. That is, find set of values for α, β, n, y where the above is not true.

3. Suppose we wish to estimate the proportion of a voting population that support a particular ballot initiative. We choose to use a Uniform prior for the proportion of voters who support the initiative. A random sample of 100 voters is polled and 55 are in favor of the ballot initiative.

   (a) Find the posterior distribution of θ.

   (b) What is the posterior mean and variance?

   (c) The `binobp` command in the Bolstad package in R will calculate the posterior for binomial data and a beta prior. it requires 4 inputs (in order): y, n, α, β. The output includes a graph of the prior and posterior distributions. Include this graph in your assignment.

   (d) The command `abline(v=location, col="colour")` adds a vertical line to a plot, where location should be replaced by the x co-ordinate of the vertical line, and colour should be replaced by the actual color. Add 3 vertical lines to your plot from the previous part: a black line representing the observed proportion of voters who support the initiative, a red line representing the prior mean, and a blue line representing the posterior mean.

   (e) Also included in the output from binobp is a table of posterior quantiles. A 95% credible interval for the posterior distribution can be found by using the 0.025 and 0.975 quantiles. What is this 95% credible interval for your posterior distribution?

(Note this interval is exactly what people wrongly assume the classical confidence interval is, that is there is a 95% chance that θ will take a value inside this interval).

(f) What is of interest to us is whether or not the initiative will pass (that is, receive a majority of Yes votes). The R command `pbeta` computes the CDF of a beta distribution and requires 3 inputs (in order): The value where you wish to evaluate the CDF, α, β. Use this to calculate our posterior probability that the initiate will pass.

4. Consider the previous question. Suppose we wish to use an informative prior instead. We would like to use a Beta prior will a mean of 0.4 and a prior standard deviation of 0.1. What are the corresponding hyper-parameters of the prior distribution? Repeat all the steps of the previous question, using the new prior distribution.

5. Each city bus in Omaha is numbered. Suppose that they are numbered sequentially $1, 2, \ldots, M$.

   (a) If M were known, and Y represents the number of the next bus you see, find an expression for $P(Y = y|M)$. For what values of y is this valid?

   (b) Now suppose that M is unknown, as we assume a geometric prior distribution on M, that is that

   $$p(M) = \frac{1}{150}\left(\frac{149}{150}\right)^{M-1} \quad \text{for } M = 1, 2, \ldots$$

   Furthermore, suppose we observe a single bus, numbered 200. Find the posterior distribution of M (up to a constant of proportionality).

   (c) Use software (for example Wolfram-Alpha or Maple) to find the constant of proportionality for the posterior, and thus find the posterior mean and variance.

   (d) If we had decided to use the improper uniform prior $p(M) \propto 1$, would this have produced a proper or improper posterior distribution? Show your work.