Assignment #3

1. In the study of water quality in New Zealand streams, documented in McBride et al. (2002), a high level of *Campylobacter* was defined as a level of greater than 100 per 100 ml of stream water. $n = 116$ samples were taken from stream having a high environmental impact from birds. Out of these, $y = 11$ had a high Campylobacter level. Let p be the true probability that a sample of water from this type of stream has a high Campylobacter level.

   (a) Find the frequentist estimator for p

   (b) Use a beta (1,10) prior for p. Calculate the posterior distribution $f(p \mid y)$.

   (c) Find the posterior mean and variance. What is the Bayesian estimator for p?

   (d) Find a 95% credible interval for p

2. The number of defects per 10 meters of cloth produced by a weaving machine has the Poisson distribution with mean λ. You examine 100 meters of cloth produced by the machine and observe 71 defects.

   (a) Your prior belief about λ is that it has mean 6 and standard deviation 2. Find a gamma(a, b) prior that matches your prior belief.

   (b) Find the posterior distribution of λ given that you observed 71 defects in 100 meters of cloth.

   (c) Find the posterior mean and variance. What is the Bayesian estimator for λ?

   (d) Calculate a 95 Bayesian credible interval for λ.

3. The data is the number of successes in N independent trials, $Y \in 0, 1, 2, ..., N$, and the unknown parameter is the success probability, $\theta \in [0, 1]$. Assume the Bayesian model with likelihood $Y|\theta \sim Binomial(N, \theta)$, prior $\theta \sim Beta(a, b)$,

   (a) Write an R function that takes Y , N, a, b, $prob$ (credible level) as inputs. The function should produce a plot (clearly labeled!) that overlays the prior and posterior density functions (both using the dbeta function), and it should return a list with the posterior mean, posterior standard deviation, posterior median, posterior mode, the endpoints of equal tail credible interval and high probability density credible interval.

   (b) What values of a and b would make good default values to represent a prior that carries little information about θ? Make these the default values in your function.

   (c) What values of a and b give prior mean 0.7 and prior standard deviation 0.2?

   (d) Now we observe $Y = 20$ events in $N = 30$ trials. Use your code from 3a to conduct a Bayesian analysis of these data. Perform the analysis twice, once with the uninformative prior from 3b and once with the informative prior in 3c.

   (e) Summarize the point and interval estimates in a table. In particular, how does this analysis compare to a frequentist analysis and how much are the results affected by the prior? Be sure all plots are labeled and code is commented!