Assignment #2

1. The bi-variate normal distribution has the joint PDF as

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{ -\frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)}{2\left(1-\rho^2\right)} \right\}$$

   (a) Find the marginal distribution of $f_X(x)$ and $f_Y(y)$

   (b) Find the conditional distribution of $f_{X|y}(x)$

2. For a Calculus I class, the final exam score Y and the average of the four earlier tests X are bi-variate normal with mean $\mu_1 = 73$, standard deviation $\sigma_1 = 12$ and mean $\mu_2 = 70$, standard deviation $\sigma_2 = 15$. The correlation is $\rho = 0.71$. Find

   (a) $E(Y|X) = \mu_{Y|X}$

   (b) $V(Y|X) = \sigma^2_{Y|X}$

   (c) $P(Y > 90|X = 80)$ that is, the probability that the nal exam score exceeds 90 given that the average of the four earlier tests is 80

3. Suppose it is known that 1% of the population suffers from a particular disease. A blood test has a 97% chance of identifying the disease for diseased individuals, but also has a 6% chance of falsely indicating that a healthy person has the disease.

   (a) What is the probability that a person will have a positive blood test?

   (b) If your blood test is positive, what is the chance that you have the disease?

   (c) If your blood test is negative, what is the chance that you do not have the disease?

4. An island has three species of bird. Species 1 accounts for 45% of the birds, of which 10% have been tagged. Species 2 accounts for 38% of the birds, of which 15% have been tagged. Species 3 accounts for 17% of the birds, of which 50% have been tagged. If a tagged bird is observed, what are the probabilities that it is of species 1, of species 2, and of species 3?

5. Let $f(x) = \exp\{-\frac{Ax^2-Bx}{C}\}$ and both $A > 0$, B and $C > 0$ be positive constants.

   (a) Find D in term of $A,\ B,\ C$ such that $Df(x)$ is a pdf of a normal distribution.

   (b) Find the mean and the variance for the normal distribution above.

6. 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, and b 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 and posterior standard deviation.

   (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 5a to conduct a Bayesian analysis of these data. Perform the analysis twice, once with the uninformative prior from 5b and once with the informative prior in 5c.

(e) Summarize the results. 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!