MATH 347 Homework 3 (Total 40 points)

Due: Thursday 10/3, at the beginning of the class

Name: ______________________________

- **Print out this cover page and staple with your homework.**
- **Show all work. Incomplete solutions will be given no credit.**
- **You may prepare either hand-written or typed solutions, but make sure that they are legible. Answers that cannot be read will be given no credit.**
- **R graphical outputs must be printed instead of hand-drawn.**

1. (*12 points; 3 points each part*)
   In 1998, the New York Times and CBS News polled 1048 randomly selected $13 - 17$ year olds to ask them if they had a television in their room. Among this group of teenagers, 692 of them said they had a television in their room. Alex and Benedict both want to use the binomial model for this dataset, but they have different prior believes about p, the proportion of teenagers having a television in their room

   (a) Alex asks 10 friends the same question, and 8 of them have a television in their room. Alex decides to use this information to construct his prior. Design a continuous beta prior reflecting Alex's belief.

   (b) Benedict thinks the 0.2 quantile is 0.3 and the 0.9 quantile is 0.4. Design a continuous beta prior reflecting Benedict's belief.

   (c) Calculate Alex's posterior and Benedict's posterior distributions. Plot the two priors on one graph, and plot the corresponding posteriors on another graph. In addition, obtain 95% credible intervals for Alex and Benedict.

   (d) Conduct prior predictive checks for Alex and Benedict. Which prior do you think is more appropriate for this teenagers and television data. Explain.

2. (*9 points; 3 points each part*)
   Continuing from Question 1. Consider the odds of having a television in the room. Recall that if p is the probability of having a television in room, then the odds is $\frac{p}{1-p}$.

   (a) Find the mean, median and 95% posterior interval of Alex's analysis of the odds of teenagers having a television in their room.

   (b) Find the mean, median and 95% posterior interval of Benedict's analysis of the odds of teenagers having a television in their room.

   (c) Compare the two posterior summaries from parts (a) and (b).

3. (*5 points*)
   Write your own R code to simulate $S = \{10, 100, 500, 1000, 5000\}$ random samples of p from the Beta(15.06, 10.56) distribution. Use the `quantile` function to find the approximated middle 90% credible interval of p for each value of S. Describe your findings and comment on the effect of simulation size S on the accuracy of the simulation results. Recall that the exact middle 90% posterior interval estimate is [0.427, 0.741].

4. (*6 points; 3 points each part*)
Do teachers' expectations impact academic development of children? To find out, researchers gave an IQ test to a group of 12 elementary school children. They randomly picked six children and told teachers that the test predicts them to have high potential for accelerated growth (accelerated group); for the other six students in the group, the researchers told teachers that the test predicts them to have no potential for growth (no growth group). At the end of school year, they gave IQ tests again to all 12 students, and the change in IQ scores of each student is recorded. Table 1 shows the IQ score change of students in the accelerated group and the no growth group.

Table 1: Data from IQ score change of 12 students; 6 are in the accelerated group, and 6 are in the no growth group.

Group	IQ score change
Accelerated	20, 10, 19, 15, 9, 18
No growth	3, 2, 6, 10, 11, 5

The sample means of the accelerated group and the no growth group are respectively $\bar{y}_A = 15.2$ and $\bar{y}_N = 6.2$. Consider independent sampling models, where the IQ scores for the accelerated group (no growth group) are assumed normal with mean μ_A (μ_N) with known standard deviation $\sigma = 4$.

$$Y_{A,i} \overset{i.i.d.}{\sim} \text{Normal}(\mu_A, 4), \text{ for } i = 1, \cdots n_A, \tag{1}$$

$$Y_{N,j} \overset{i.i.d.}{\sim} \text{Normal}(\mu_N, 4), \text{ for } i = 1, \cdots n_N, \tag{2}$$

where $n_A = n_N = 6$.

(a) Assuming independent sampling, write down the likelihood function of the means (μ_A, μ_B).

(b) Assume that one's prior beliefs about μ_A and μ_N are independent, where $\mu_A \sim N(\gamma_A, \sigma_A)$ and $\mu_N \sim N(\gamma_N, \sigma_N)$. Show that the posterior distributions for μ_A and μ_N are independent normal and find the mean and standard deviation parameters for each distribution.

5. (*8 points; 4 points each part*)
Continuing from Question 4. Assume that one has vague prior beliefs and $\mu_A \sim N(0, 20)$ and $\mu_N \sim N(0, 20)$.

(a) Is the average improvement for the accelerated group larger than that for the no growth group? Consider the parameter $\delta = \mu_A - \mu_N$ to measure the difference in means. The question now becomes finding the posterior probability of $\delta > 0$, i.e. $p(\mu_A - \mu_N > 0 \mid \mathbf{y}_A, \mathbf{y}_N)$, where $\mathbf{y}_A$ and $\mathbf{y}_N$ are the vectors of recorded IQ score change. (Hint: simulate a vector s_A of posterior samples of μ_A and another vector s_N of posterior samples of μ_N (make sure to use the same number of samples) and subtract s_N from s_A, which gives us a vector of posterior differences between s_N and s_A. This vector of posterior differences serves as an approximation to the posterior distribution of δ.)

(b) What is the probability that a randomly selected child assigned to the accelerated group will have larger improvement than a randomly selected child assigned to the no growth group? Consider $\tilde{Y}_A$ and $\tilde{Y}_N$ to be random variables for predicted IQ score change for the accelerated group and the no growth group, respectively. The question now becomes finding the posterior predictive probability of $\tilde{Y}_A > \tilde{Y}_N$, i.e. $p(\tilde{Y}_A > \tilde{Y}_N \mid \mathbf{y}_A, \mathbf{y}_N)$, where $\mathbf{y}_A$ and $\mathbf{y}_N$ are the vectors of recorded IQ score change, each of length 6. (Hint: Show that the posterior predictive distributions of $\tilde{Y}_A$ and $\tilde{Y}_N$ are independent. Simulate predicted IQ score changes from the posterior predictive distributions for the two.)