For typed text, my vectors and matrices are denoted with bold font. On your homework, please explicitly denote vectors using the arrow notation!

Part 1

1. (R) Suppose an experimental machine in a lab is either fine, or comes from a bad batch of machines that should be recalled by the manufacturer. Scientists in the lab want to estimate the failure rate of their machine and decide whether or not to return it. They encode their prior uncertainty about the failure rate θ with the following density:

$$f_\theta(\theta) = \frac{1}{4}\frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}\left(3\theta(1-\theta)^7 + \theta^7(1-\theta)\right), \qquad 0 < \theta < 1.$$

   (a) Make a plot of this prior density and explain why it "makes sense" for the scientists. Based on the prior density, which do the scientists think is more likely: that their machine is fine, or bad?

   - To plot the prior, create a vector of possible θ values where increments between each value are small (like 0.001 or 0.005) and evaluate your prior at these values (just like we did in the discrete case a few weeks ago). This time, use `geom_line()` (ggplot) or `type = "l"` (base R) to get a smooth curve that indicates a continuous prior.

   (b) The scientists run the machine n times. Let $y_i = 1$ if the machine fails on the i-th run, and 0 otherwise, for $i = 1, \ldots, n$ runs. Write out the exact posterior distribution for θ, simplifying as much as possible. The posterior is a mixture (i.e. weighted average) of two distributions that you know. Identify those two distributions.

   (c) Suppose the scientists run the machine $n = 5$ times and observe $\sum y_i = 2$. Now plot the posterior distribution of θ.

   (d) Using your discretized approximation to the posterior in (c), obtain an 80% symmetric credible interval for θ. *Hint: first normalize the posterior density values so they sum to 1, then consider some sort of running/cumulative sum.*

2. (R) (Continuation from 1) Let's try to avoid discretizing by obtaining a Monte Carlo sample from the posterior distribution! But wait...how can we sample from the posterior distribution if it's not a named distribution? Well, you found in 1(b) that the posterior is a mixture (i.e. weighted average) of two named distributions. Let's call the named distributions p_1 and p_2.

   To sample a random variable Z from the mixture distribution $f_Z(z) = wp_1(z) + (1-w)p_2(z)$, we can imagine tossing a coin where the probability of Heads is w. If the coin lands Heads, sample z from p_1, and if the coin lands Tails, sample z from p_2.

(a) Using the technique described above, obtain a Monte Carlo sample of $S = 5000$ values from the posterior for θ. *Note, you can "flip the coin" using either* `rbinom()` *or* `sample()`. *Also, you should think about what your weights are!* Obtain a Monte Carlo approximation to the posterior distribution (e.g. make a normalized histogram or density estimate from your S samples), and then obtain an 80% credible interval for θ based on your Monte Carlo sample.

(b) Parts b and c coming soon! Sorry for the delay!

3. Also coming soon, sorry for delay!

4. The following will be graded as one problem!

   (a) (Probability review – useful for use later!) If $X \sim \text{Gamma}(a, b)$, find the PDF of the random variable $W = \frac{1}{X}$.

   (b) (`R`?) I've said in class that when we have Bernoulli data, a noninformative prior for the success probability θ would be $\theta \sim \text{Unif}(0, 1) \equiv \text{Beta}(1, 1)$. In other words, this might encode "I have no prior information about θ and so all values will be *a priori* likely". However, when considering binary data, it's very common to examine the log-odds; i.e. we're interested in $\gamma = \log \frac{\theta}{1-\theta}$. This is what we do in logistic regression.

   Either analytically or via Monte Carlo simulation, plot the prior for γ that is induced by the prior $\theta \sim \text{Unif}(0, 1)$. Is the prior informative about γ?

Part 2

Throughout Part 2, we will work with the sampling model $Y_1, \ldots, Y_n | \theta \overset{\text{iid}}{\sim} N(\theta, \sigma^2)$ and the prior $\theta \sim N(\mu_0, \sigma_0^2)$. **Importantly**: we will assume that σ^2 is known (not random), so the only unknown parameter is θ.

5. We will continue working with the posterior for θ that we derived in class.

   (a) Express the posterior mean μ_n as a weighted average of the prior mean and the sample mean. (In this sense, the posterior mean compromises between the prior and data means). What are the weights proportional to?

   (b) We could derive the exact posterior predictive distribution for an unobserved Y^* via integration, but there's actually an easier route!
   Let's use the following fact:

$$Y^* | \theta, \sigma^2 \sim N(\theta, \sigma^2) \iff Y^* = \theta + \epsilon^*, \text{ where } \epsilon^* | \theta, \sigma^2 \sim N(0, \sigma^2)$$

   Using properties of expectation and variance:

   i. Find the posterior predictive expectation $\mathbb{E}[Y^* | \mathbf{y}]$ (remember σ^2 is known).

ii. Find the posterior predictive variance $\text{Var}(Y^*|\mathbf{y})$.

iii. Using these previous two steps and some properties of independent Normals, name the exact posterior predictive distribution.

(c) Some intuition building: As $n \to \infty$, what values do the posterior variance σ_n^2 and the posterior predictive variance $\text{Var}(Y^*|\mathbf{y})$ approach? Why does this "make sense"?

6. (R) Do teachers' expectations impact academic development of children? To find out, researchers gave an IQ test to a group of 11 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 A); for the other five students in the group, the researchers told teachers that the test predicts them to have no potential for growth (no growth group B). At the end of school year, they gave IQ tests again to all 11 students, and the change in IQ scores of each student is recorded.

We will consider the following sampling models for IQ scores change, independent across the two groups A and B:

$$Y_{A,i}|\theta_A \overset{\text{iid}}{\sim} N(\theta_A, 25) \qquad i = 1, \ldots, n_A$$
$$Y_{B,i}|\theta_B \overset{\text{iid}}{\sim} N(\theta_B, 25) \qquad i = 1, \ldots, n_B$$

where n_A and n_B are the samples sizes of the two groups.

The following shows the IQ score change of students in the accelerated group and the no growth group:

- Accelerated group: 14, 10, 11, 15, 9, 18
- No growth group: 3, 2, 6, 10, 11

(a) Using the conjugate family of priors for Normal data, we will formulate relatively vague prior beliefs for the true mean change in IQ scores. Our two priors for θ_A and θ_B will be independent and have mean 0 and variance of 16. Describe in words what this choice of prior means in the context of the problem. Then, based on the observed data and these priors, write down the exact posteriors for θ_A and θ_B.

(b) We aim to answer the following question: Is the average improvement for the accelerated group A larger than that for the no growth group B? Let's define the parameter $\delta = \theta_A - \theta_B$ to measure the difference in means. Using either exact or simulation-based methods, answer the research question by finding the posterior probability that $\delta > 0$. (You should confirm for yourself that answering this question is equivalent to learning if $\delta > 0$).

(c) We now aim to answer the following question: What is the probability that a new child assigned to the accelerated group A would have larger improvement than a new child assigned to the no growth group B? Using either exact or simulation-based methods, answer the research question by finding the posterior predictive probability $\Pr(Y_A^* > Y_B^*|\mathbf{y}_A, \mathbf{y}_B)$.

(d) Time to perform a posterior predictive check: let's assess if the Normal model is inadequate. We will use a test function that measures skewness: $T(\mathbf{y}) = 3(m_{\mathbf{y}} - \bar{y})/s_{\mathbf{y}}$ where $m_{\mathbf{y}}$ is the sample median, $\bar{y}$ is the sample mean, and $s_{\mathbf{y}}$ is the sample standard deviation. If a distribution is symmetric, the value should be 0! For each group A and B, obtain $S = 2000$ predictive datasets ${\mathbf{y}_A^*}^{(s)}$ and ${\mathbf{y}_B^*}^{(s)}$ (remember they must be of same sample size as original data, respectively). For each dataset, record the value of the test function. Then make two histograms for the simulated values of the test functions, and compare them to the observed values from the real data. Based on what you find, assess the fit of the Normal data model for these data.

General rubric

Points	Criteria
5	The solution is correct *and* well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.
Notes:	For problems with code, well-written means only having executed lines of code that are necessary to solving the problem (you're welcome to comment out code for yourself to keep), as well as presenting the solution for the reader to easily see.