Problems labeled with `R` indicate that you should complete the problem using `R`. You should complete these problems in a `.Rmd` or `.qmd` file, knit/render to PDF, and submit the printed PDF alongside the rest of the problems. If you need to do some calculations by hand, that's fine! Just clearly indicate that on the document.

A rubric can be found on the last page.

1. (`R`) The number of particles Y emitted per hour from a substance depends on the unknown proportion θ of the substance that is radioactive. For each possible value of θ,

$$\Pr(Y = y|\theta) = f_{Y|\theta}(y|\theta) = \frac{(5\theta)^y e^{-5\theta}}{y!} \qquad y = 0, 1, \ldots$$

   Suppose that the substance is monitored and it is observed that $Y = 3$.

   (a) Make a graph of the posterior $f_{\theta|Y}(\theta|y)$ given the observed data, likelihood, and the prior $f_\theta(\theta) = \frac{1}{101}$ for each $\theta \in \{0, \frac{1}{100}, \frac{2}{100}, \ldots, 1\}$ (and 0 otherwise).

   (b) What are the mean, median, and mode of θ under our prior? Using code and/or your graph, what are the posterior mean, median, and mode?

2. I am always interested in how people affiliated with the college pronounce "Middlebury", so you are tasked with collecting some data to help me learn more!

   IMPORTANT: don't collect the data until you've done parts (a)-(c)!

   You will survey 10 people (ideally unrelated to you/each other in order to be as independent as possible). For each person, ask them to pronounce the name of the college. It is important that you don't pronounce it first! Define $Y_i = 1$ if the i-th person pronounces "middle-BERRY" and 0 if they pronounce it "middle-BURRY".

   (a) If you define $X = \sum_{i=1}^{10} Y_i$, what likelihood would you define for X?

   (b) If I instead want to work with the original data $\mathbf{Y} = (Y_1, \ldots, Y_{10})$, what likelihood would you define for $\mathbf{Y}$? i.e., what is $f_{\mathbf{Y}|\theta}(y_1, \ldots, y_{10}|\theta)$? State/notate any assumptions you must make, and simplify as much as possible.

   (c) Treat your unknown parameter θ as a discrete random variable. Define a prior distribution for θ, and justify your choice! *You almost surely have some opinions, so I better not see a discrete uniform! And remember, a distribution is not fully specified without its support!*

   (d) Collect your data, and write your observed $\mathbf{y}$ here.

   (e) (`R`) Using your likelihood from (b), plot your prior and posterior (ideally with the same y-axis). Then comment: how, if at all, did your posterior change from the prior?

   (f) For these data, whether or not you use your likelihood from (a) or (b) should not alter the resulting posterior. Can you briefly explain/demonstrate mathematically why that is?

3. Sequential Bayesian analysis/Bayesian learning. In the previous problem, we obtained our posterior using all 10 data points "at once". What would happen if we first sampled three people and obtained the posterior, and then sampled an additional seven people? How would that change our ultimate, 'final' posterior for θ based on 10 total observations? We will explore that here.

   Let's simplify things even more. Suppose we have two observations only: Y_1 and Y_2. We can either choose to obtain the posterior $f_{\theta|Y_1,Y_2}(\theta|y_1, y_2)$ by:

   - Option 1: using $\mathbf{Y} = (Y_1, Y_2)$,
   - Option 2: first obtaining Y_1 and then Y_2, or
   - Option 3: first obtaining Y_2 and then Y_1.

   Option 1 is what we've seen, where the ultimate posterior for θ assuming the prior $f_\theta(\theta)$ before observing any data is:

$$f_{\theta|\mathbf{Y}}(\theta|y_1, y_2) = \frac{f_{\mathbf{Y}|\theta}(y_1, y_2|\theta)f_\theta(\theta)}{f_{\mathbf{Y}}(y_1, y_2)}$$

   Let's see what happens if we do Options 2 (parts (a)-(c)) and 3 (part (d)).

   (a) First obtain the posterior for θ given only the *first* data point y_1.

   (b) Now, we will update the model in light of observing the new, second data point y_2. At this point, we've already changed our understanding of θ using y_1 in part (a)! So what distribution encapsulates our prior beliefs about θ, where "prior" is now used as "prior to observing Y_2"?

   (c) Based on (b), what is the new posterior for θ given first y_1 and then y_2? That is, what is $f_{\theta|Y_1,Y_2}(\theta|y_1, y_2)$ when performing this sequential analysis?

   (d) Now let's do Option 3: first sample Y_2, obtain the posterior for θ, and then sample Y_1 and obtain your ultimate posterior for θ. How does the final posterior $f_{\theta|Y_2,Y_1}(\theta|y_2, y_1)$ compare the posterior you obtained in (c)? Based on what you find, does the order of the data entering the model influence the ultimate posterior model of θ?

   (e) Under what assumptions are your ultimate posteriors from Options 2 and 3 and the posterior from Option 1 the same? (i.e. when does performing a sequential analysis yield the same ultimate posterior as "dumping all your data" in at the beginning?)

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.