MATH 538: Bayesian Statistics
Quiz #1 (Take-Home)
Due Date: Monday September 23, 2024 by 11:59pm

Important Instructions: Please read the instructions very carefully.

- There is 1 question on this exam. You must type ALL answers to receive full marks.
- You must do the exam individually, and cannot discuss the questions and/or the course material during the exam period with any other individual except the instructor. You can use your notes and/or the textbook, but you are not allowed to use the internet or any online source related to the course material. Copying solutions from the Internet is considered forgery and cheating.
- You should hand in your solutions (including the R codes) in one pdf file. You should also hand in your R codes in one separate .txt, .R, .RMD file. Make clear breaks/comments in your R codes to separate questions and/or different sections of a question. Both files should be uploaded **onto Canvas to Quiz #1 Take Homee**. Name the two files as follows:
  Solutions: YourFirstNameYourLastName-Exam1-Solutions.pdf
  Example: ValeriePoynor-Quiz1-Solutions.pdf
  R code: YourFirstNameYourLastName-Quiz1-Rcode
  Example: ValeriePoynor-Quiz1-Rcode.txt OR ValeriePoynor-Quiz1-Rcode.R
- Upload your files by **11:59pm on Monday, September 18$^{\mathbf{th}}$**. The two files should be uploaded on Canvas.
- The instructor may schedule an interview with you to discuss your exam where you may be asked to elaborate on your solutions.
- **The deadline is firm, and late penalties will be applied if the submission is not received by the due date.**

1. **[65 marks]** [RELIABILITY APPLICATION] Consider the *Engine.csv* data set. The Raleigh distribution was named after Lord Raleigh, a renowned mathematician and physicist who received the Nobel Prize in 1904 for his discovery of Argon and related research. The Raleigh distribution, which is defined by the magnitude of two vectors arising from independent normal distributions centered at zero and having the same variance, is often utilized to model lifetime data. Here, we will apply the Raleigh distribution to a dataset describing the time (in weeks) to a valve seat replacement in 24 diesel engines.

$$y_i \overset{iid}{\sim} Raleigh(\sigma^2) \text{ where } f(y_i|\sigma^2) = \frac{y_i}{\sigma^2} e^{-y_i^2/(2\sigma^2)}$$

$$\text{with prior } \sigma^2 \sim \Gamma^{-1}(a_0, b_0) \text{ where } p(\sigma^2|a_0, b_0) = (b_0^{a_0}/\Gamma(a_0))(\sigma^2)^{-(a_0+1)} \exp\left(-b_0/\sigma^2\right)$$

for $i = 1, ..., n_j$

In class, you showed that the posterior distribution has the form:

$$\sigma^2|\boldsymbol{y} \sim \Gamma^{-1}(n + a_0, 0.5\sum_{i=1}^{n} y_i^2 + b_0)$$

Note: Please use the inverse gamma distribution in the MCMCpack library and the Raleigh distribution in the VGAM library for relevant questions in this problem.

(a) Perform an exploratory analysis on the observed valve seat replacement times. Provide plots, summary tables, and discussion.

(b) Derive the form of the NORMALIZED likelihood under this model with respect to σ^2.

(c) The parameter of the Raleigh distribution, σ, represents the scale and is also the mode. From 10 previous replacement measurements, we believe apriori that most engines will required a valve replacement around 6 months (or 24 weeks). The prior mean of the Inverse Gamma distribution and the fact that $a_0 + 1$ is the prior effective sample size to specify values for a_0 and b_0. Overlay the prior, likelihood, and posterior densities of σ^2 in one figure. Discuss your findings. In particular, discuss how informative the prior appears to be compared to the data. Is the posterior mode closer to the prior mode or the MLE?

(d) Obtain and plot 2000 Monte Carlo samples of σ^2 from its posterior distribution. Use your samples to also obtain a 95% posterior credible interval for σ^2.

(e) Analytically derive the prior and posterior predictive distributions of a valve seat replacement time, and plot these distributions (note, you may have to write your own functions for these distributions) in the same graph over replacement times ranging from 0 to 100 weeks. Provide a discussion on how they compare.

(f) Obtain samples from the posterior predictive distribution using Monte Carlo. Report a point and 95% credible interval and formally interpret the interval as a sentence in context of the problem. What is the probability that an engine will need a valve seat replacement before 6 months?