Maggie O'Shea
Bayesian Statistical Modeling and Computation
Professor Klaus Keller
Problem Set 3
February 14, 2025

1. **Review the key sources already assigned as reading with a special focus on:**
a) Qian, S. S., Stow, C. A., & Borsuk, M. E. (2003). On Monte Carlo methods for Bayesian inference. Ecological Modelling, 159(2-3), 269–277.
b) D'Agostini, G. (2003). Bayesian reasoning in data analysis: A critical introduction. Singapore: World Scientific Publishing. (Chapter 6 only).
c) Ruckert, K. L., Guan, Y., Bakker, A. M. R., Forest, C. E., & Keller, K. (2017). The effects of time-varying observation errors on semi-empirical sea-level projections. Climatic Change, 140(3-4), 349–360. https://doi.org/10.1007/s10584-016-1858-z

Also reviewed:
d) Kim, Y., Bang, H., Kim, Y. & Bang, H. Introduction to Kalman Filter and Its Applications. in Introduction and Implementations of the Kalman Filter (IntechOpen, 2018). doi:10.5772/intechopen.80600.

2. Draw the probability density function for the usable fuel in the tank without any other information besides the fuel gauge reading. Determine the expected value of available fuel, the most likely value of available fuel, and probability of negative fuel in the tank. Do these estimates make sense to you? (1 point)

The expected value of available fuel is 33.99 . According to the PDF (Figure 1), there is a 0.045 probability of negative fuel in the tank. The most probable value is 32.288 which is the value with the highest probability. Given the initial gauge reading of 34, the expected value and most likely value of available fuel are possible and make sense. However, it does not make sense that there is a non-zero possibility that there could be negative fuel.
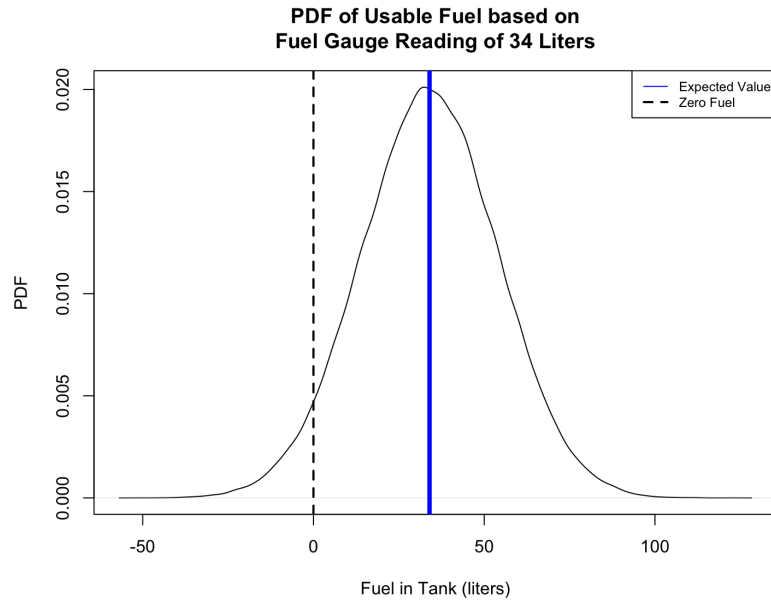
Figure 1: Probability Density Function (PDF) of Usable Fuel based on Fuel Gauge Reading of 34 Liters. The expected value is indicated in blue and the dashed line is at the point of zero fuel.

In terms of uncertainties, in this case the PDF was made by randomly sampling a normal distribution and then calculating the PDF - because I used just one seed there is uncertainty around the seed and if results would differ with differing seeds. Without performing the analysis it isn't clear how this would impact the conclusions, but it might result in a slightly different PDF, as well as help quantify uncertainty around the most probable value and expected value. The choices made here included estimating that the fuel in the tank was gaussian around the mean (observed) 34 liters with a standard deviation of 20, as well as a selected approach to making the PDF through sampling and then estimating the PDF. This is a defensible choice as it is based on observed data (the reading) as well as knowledge about the uncertainty in the sensor reading. This analysis was made reproducible by setting a seed as well as by including comments explaining the steps.

3. How can you use a proper prior to address the issue of unrealistic fuel estimates in the tank. Define this physically based prior for you (meaning this is your subjective prior). (3 points)

A new physically based prior is a uniform distribution from 0 to 182. This prior can be used to address the unrealistic fuel estimates in the tank because it sets bounds around the posterior such that they won't surpass the minimum/maximum fuel in the tank. Uncertainty would be relevant here because it's my subjective prior, which can introduce deep uncertainty, and perhaps also if there is any uncertainty around the exact size of the tank.

4. Use a grid-based method to determine your Bayesian update from your prior and the likelihood function. Add this posterior to the plot produced above. Determine now the probability of negative fuel. Has this fixed the issue? If so, how?

The probability of negative fuel is zero (Figure 2, Figure 3). This has fixed the issue because the prior was a uniform distribution from 0 to 182 such that when combined with the observed fuel, the resulting posterior was a truncated normal distribution limited by the limits of the uniform distribution, ie cut off at 0. The resulting CDF and PDF of the posterior distribution is shown in Figure 2 and 3.
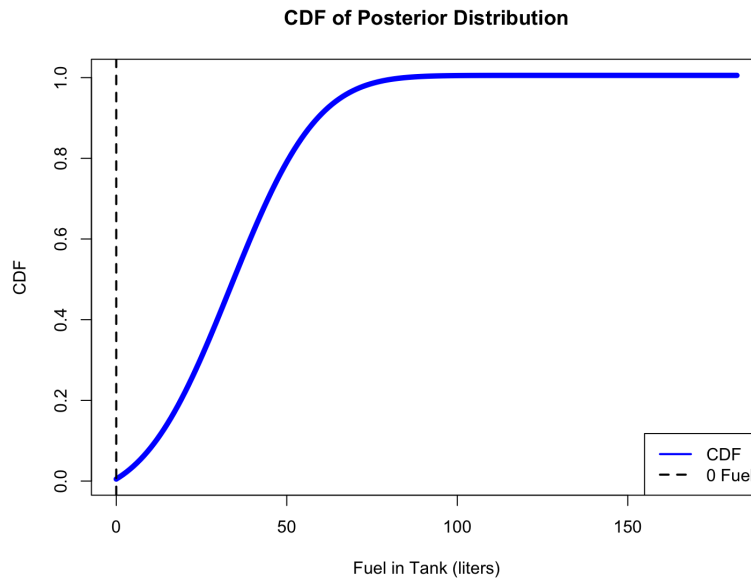

Figure 2: Cumulative Distribution Function (CDF) of Posterior Distribution estimated with physically informed prior based on fuel capacity of tank. Red lines point to the probability of 0 fuel (0).

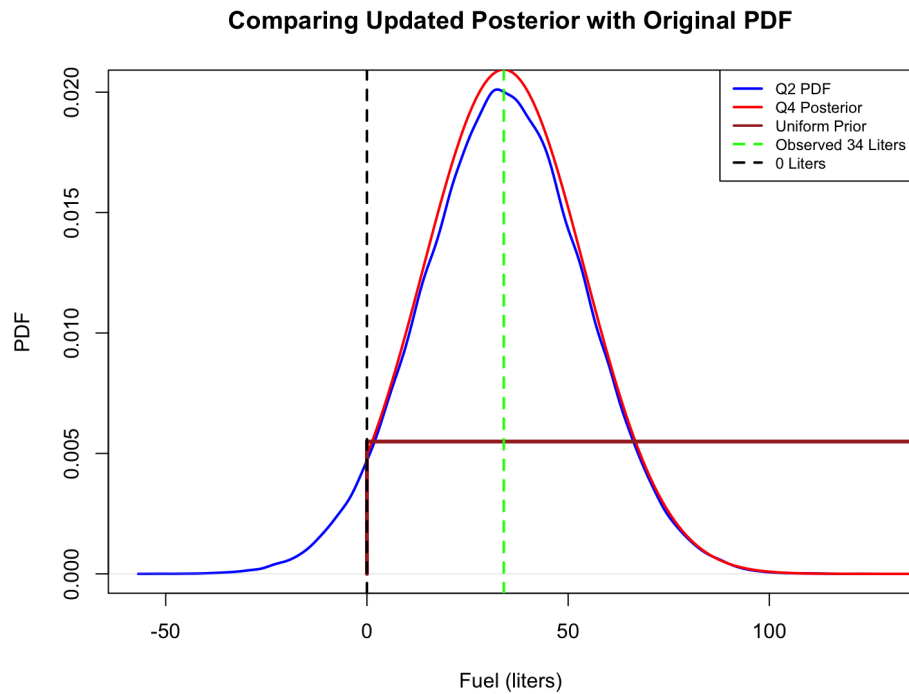**Comparing Updated Posterior with Original PDF**

Figure 3: Comparing Updated Posterior with Original PDF from Question 2 (Q2). Brown line indicates uniform prior. Green dashed lines indicates the observed 34 liters. Black Dashed line indicates 0 liters.

In this case there are a few sources of uncertainty and choices that I made. First, I chose the grid values used in the estimation of the posterior were 182 values from 0 to 182 - it's possible that if I used a different number of values ie more or less grid values the result may vary slightly. This could introduce a type of uncertainty from the data used. In this case, I think the choice that I made using182 values is defensible as 182 allows for all whole number of liters to be represented in the dataset without using so many grid values that the computational processing time is slowed. There is also prior uncertainty and model structure uncertainty because I did not try multiple models or priors to show the range of possible outcomes. The prior selected is my subjective prior. It is defensible in that it reflects the true capacity of the tank from 0 to 182 liters. Finally, this analysis was made reproducible by commenting the code to explain each piece of it such that one could follow along.

5. Repeat the step above using a Bayes Monte Carlo method

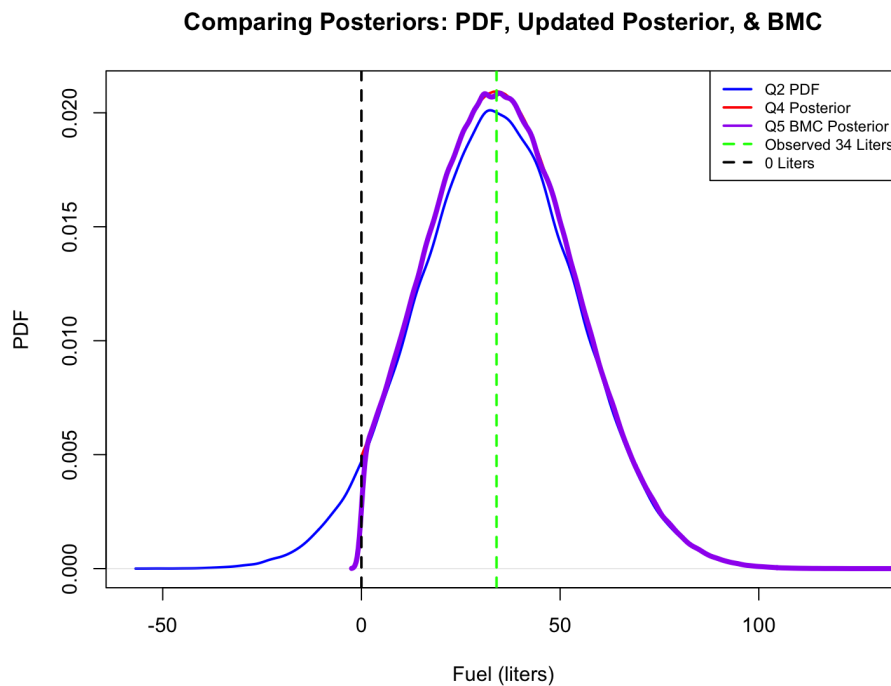**Comparing Posteriors: PDF, Updated Posterior, & BMC**

Figure 4: Comparing Results from Question 2 (Q2) PDF, Question 4 (Q4) Bayesian Posterior, and Question 5 (Q5) Bayes Monte Carlo Posterior with original PDF for fuel in the tank. Green dashed line indicates the observed 34 liters and black dashed line at 0 liters.

The steps above were repeated using Bayes Monte Carlo (BMC) method (Figure 4). Again the probability of negative fuel is 0. In this analysis, compared to that of Question 4 (Q4), there are additional sources of uncertainty including but beyond those expressed in Q4. The new sources of uncertainty primarily come from the random sampling required to perform the BMC approach particularly because we randomly sample the prior to start and then randomly sample the posterior as well. This introduces a source of uncertainty around the chosen seed (in this case 930) to perform the analysis. While it's not possible to know completely how this would impact the results if I vary the seeds, it likely would show variability around the resulting expected values or most probable values. Another choice that was made that could introduce uncertainty is the number of trials or samples taken for the random sampling. 3,000,000 samples were used to address the concern with having too few samples leading to biased results. This sample size was selected based on the passing of a convergence test. I began with 10,000 samples and tested if the result passed the convergence test based on the credible interval width, and increased the sample size until this test was passed. The results around the expected value did not vary widely even between the 10,000 sample BMC posterior and 3,000,000 expected value, however it took that many samples for convergence. Finally, this was made reproducible, again, by commenting the code as well as setting a seed to reproduce the random sampling results.

6. What assumptions would you need to make for a simple analytical Kalman filter solution to this problem? Are these assumptions realistic?

The Kalman filter requires that the process it is modeling is linear, as well as that the data's noise is gaussian. In this case, I think it is realistic that this is a linear problem, as well as the fact that the data is normally distributed given the one observation of the fuel in the tank. Without that observation, however, the data is uniformly distributed where it is equally likely to have 34 liters as it is to have 59 liters. In this case, this compromises the ability to use the Kalman filter solution here.

7. Produce a plot of the estimated available flight time. Use this plot to address these
a. What is the probability that you make an airport that is 100 minutes flight time away with at least 30 min reserve fuel required by regulations?
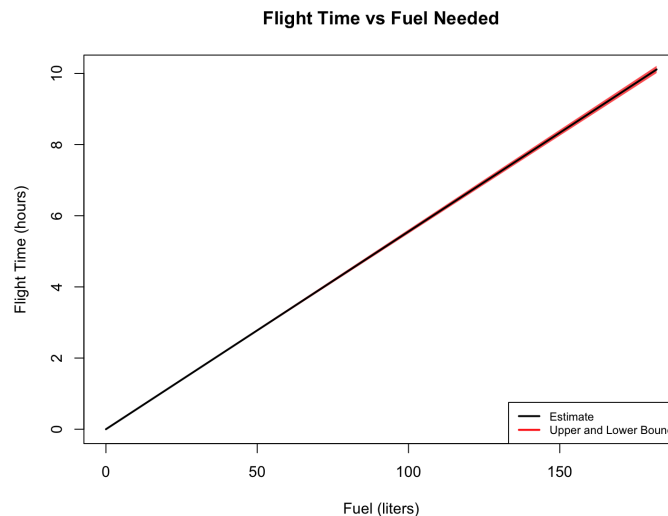**b. What is the probability that you run out of fuel trying to make it to this airport?**

Figure 5: Flight time and necessary fuel with uncertainty bounds. Flight time based on 18 liters per hour estimate of needed fuel, and uncertainty based on 2 liters per hour standard deviation.

Comparing Posteriors: PDF, Updated Posterior, & BMC
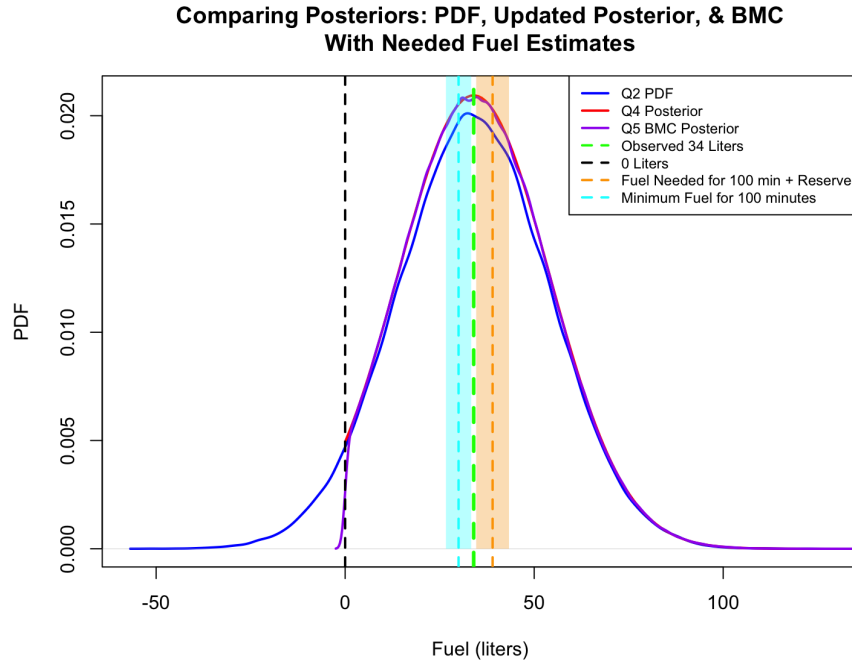**With Needed Fuel Estimates**

Figure 6: Examining the posteriors relative to the needed fuel for 100 minute flight time with 30 minute reserve (orange) and minimum fuel needed for just 100 minute flight time (blue). Shaded orange and blue represent uncertainty around the estimated probability of enough fuel. Green dashed line indicates observed 34 liters and black dashed line indicates 0 liters. Q2-5 refer to Questions 2, 4, and 5 posterior PDFs.

a) The probability of having enough fuel to make it to the airport with a 100 minute flight time and 30 minutes of fuel reserve is 0.59. However there are uncertainties in this estimate given that there are uncertainties around the fuel needed per hour. Given this, the range of probability of having enough fuel to do this with the 30 minute reserve is 0.49 to 0.66. This result was made reproducible by, again, commenting the code thoroughly.

b) The probability of running out of fuel to make it to this airport 100 minutes away is 0.40, however, again there is uncertainty in this estimate due to uncertainties around the amount of fuel needed per hour. The range of probabilities of running out of fuel range from 0.33 to 0.48. Again, in this case this result was made reproducible by commenting the code thoroughly.

Works Cited:

Qian, S. S., Stow, C. A., & Borsuk, M. E. (2003). On Monte Carlo methods for Bayesian inference. Ecological Modelling, 159(2-3), 269–277.

D'Agostini, G. (2003). Bayesian reasoning in data analysis: A critical introduction. Singapore: World Scientific Publishing. (Chapter 6 only).

Ruckert, K. L., Guan, Y., Bakker, A. M. R., Forest, C. E., & Keller, K. (2017). The effects of time-varying observation errors on semi-empirical sea-level projections. Climatic Change, 140(3-4), 349–360. https://doi.org/10.1007/s10584-016-1858-z

Kim, Y., Bang, H., Kim, Y. & Bang, H. Introduction to Kalman Filter and Its Applications. in Introduction and Implementations of the Kalman Filter (IntechOpen, 2018). doi:10.5772/intechopen.80600.

Appendix I:

Code written in software R (Version 2024.12.0+467).

- Downloaded from: https://www.r-project.org/

Saved to Github Repository:
https://github.com/maggieoshea/BayesianStatisticalModelingandComputation/tree/main/Problem Set3

Full R Script below.