**Problem Sets for ENGG 107: Bayesian Statistical Modeling and Computation**

**(Winter 2023)**

Overview of Assignments

1. PS1: Review of Academic Integrity and Course Syllabus
2. PS2: Monte Carlo / Convergence / Seeds / Scripts / Plotting
3. PS3: Analytical Methods / Bayes Monte Carlo / Grid Methods
4. PS4: Markov Chain Monte Carlo
5. PS5: Project Pitch
6. PS6: Prior Predictive Checks / Probabilistic Inversion
7. PS7: Decision Analyses Links / Deep Uncertainty
8. Assignment: Final Project Summary

**Problem Set #1: Review of Academic Integrity and Course Syllabus**

Assigned: Friday, January 13 2023 via Canvas

Due: Friday, January 20, 2023 via Canvas

**Tasks:**

1. Review the entire course syllabus on CANVAS.
2. Review the documents linked under academic integrity
3. Review the statement on “Students’ Consent to recording of course meetings and office hours that are open to multiple students”
4. Please address whether (and why) the following cases are consistent with the academic integrity rules this course abides by:
   1. You collaborate with a few fellow students after class on how to tackle a problem set by discussing the background papers and the example codes provided in class. Then you write your code by yourself, cite all sources, and acknowledge the discussions with your fellow students by name.
   2. You download a code that does a similar task and analyze the flow, logic, method choices, and the syntax. You then put this code away and write your own code and analysis summary from scratch and cite the sources that served as inspiration.
   3. You copy code from a fellow student in this course and submit it as your own work without citing the source.
   4. You are unsure about syntax for a function call in the code. You look up an example from the help function and from the web. You copy the expression from the example and change and input names. You do not cite the source for the function call.
   5. You take code from a chat session from openai.com.
5. Submit your signed syllabus as well as the answers to questions 4a to 4e as a pdf via CANVAS.

**Problem Set #2: Monte Carlo / Convergence / Seeds / Scripts / Plotting**

Assigned: Friday, January 20 2023 via Canvas

Due: Friday, January 27, 2023 via Canvas

**Tasks:**

Notes: Tasks 1 to 4 are you doing something. Task 5 is you documenting this. You get points for task 5.

1. Review an example script we discussed in class.
2. Review at least two other example scripts (for example from code replication repositories from papers in your application area)
3. Review the key sources already assigned as reading with a special focus on:
   1. Labs 0 to 3 in: Applegate, P. J., & Keller, K. (Eds.). (2016). Risk analysis in the Earth Sciences: A Lab manual. 2nd edition. Leanpub. Retrieved from https://leanpub.com/raes
4. Use a Monte Carlo simulation method to:
   1. determine the mean and the 95 percentile of sample from a known univariate normal distribution with a mean of zero and a standard deviation of one with your estimated uncertainties and
   2. determine the value of pi with your estimated uncertainties
5. For each task:
   1. Produce a pdf file summary that documents the task, your approach, any assumptions, your results, and includes at least one figure that illustrates your main finding(s) (5 points)
   2. Please discuss your choices (for example about how to select samples and how to determine convergence). For each choice, provide a brief overview of plausible choices and how you made your specified choice. (5 points)
   3. Address whether (and if so how) these analyses are reproducible (1 points)
   4. Include the code as an appendix in the pdf file (1 points)
   5. Check whether you have all required citations (1 points)
   6. Check whether you have assigned a copyright and a license to your codes (1 points)
   7. Check whether the figure follows standard design praxis (see, for example, [this guide](https://www.nature.com/documents/nature-final-artwork.pdf)) (1 points)

Please submit your summary pdf as well as your code to Canvas. The points are for each task.

**Problem Set #3: Analytical Methods / Bayes Monte Carlo / Grid Methods**

Assigned: Friday, January 27 2023 via Canvas

Due: Friday, February 3, 2023 via Canvas

**The Problem:**

You are part of a team that is refining a safety system for an airplane that can land an airplane automatically at the next available and safe airport in case of an emergency (see <https://www.youtube.com/watch?v=PiGkzgfR_c0> for an example). The computer system makes a decision on which airport to choose to land based on factors such as wind, runway length, legal requirements, fuel use, and available fuel. The team asks you to design, test, and document a draft computer program that takes as input a reading from a fuel sensor (together with some other information) and produces an estimate of the usable fuel in the tank with an estimate of the associated uncertainties and chooses a location to land. You also need to understand the interface of your tasks to the overall problem.

**Inputs:**

* Total fuel tank capacity 182 liters
* Digital fuel sensor has an error following a Gaussian distribution with a standard deviation of 20 liters.
* Fuel sensor reading is 34 liters
* Your airplane requires 18 gallons per hour (uncertain with a Gaussian standard deviation of 2 gallons per hour).

**Tasks:**

1. Review the key sources already assigned as reading with a special focus on:
   1. Qian, S. S., Stow, C. A., & Borsuk, M. E. (2003). On Monte Carlo methods for Bayesian inference. Ecological Modelling, 159(2-3), 269–277.
   2. D’Agostini, G. (2003). Bayesian reasoning in data analysis: A critical introduction. Singapore: World Scientific Publishing. (**Chapter 6 only**).
   3. 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>
2. Draw the probability density function for the useable 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?
3. How can you use a proper prior to address the issue of negative fuel in the tank. Define this physically based prior for you (meaning this is *your subjective prior*).
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 in task 1. Determine now the probability of negative fuel. Has this fixed the issue? If so, how?
5. Repeat the step above using a Bayes Monte Carlo method.
6. What assumptions would you need to make for a simple analytical Kalman filter solution to this problem? Are these assumptions realistic?
7. Produce a plot of the estimated available flight time. What is the probability that you make an airport that is 100 minutes flight time away with 30 min reserve fuel required by regulations?
8. For documentation
   1. Produce a pdf file summary that documents the task, your approach, any assumptions, your results, and includes at least one figure that illustrates your main finding(s).
   2. Please discuss your choices. For each choice, provide a brief overview of plausible choices and how you made your specified choice.
   3. Address whether (and if so how) your analysis is reproducible
   4. Include the code as an appendix in the pdf file
   5. Check whether you have all required citations
   6. Check whether you have assigned a copyright and a license to your codes
   7. Check whether the figure follows standard design praxis (see, for example, [this guide](https://www.nature.com/documents/nature-final-artwork.pdf))
   8. Submit your summary pdf as well as the codes to Canvas

**Problem Set #4: Markov Chain Monte Carlo**

**Assigned:**  Friday, January **x** 2023 via Canvas

**Due:** Friday, January x, 2023 via Canvas

**Tasks:**

1. Revisit the decision problem from the previous problem set. Repeat the inference and the decision analysis using a Markov Chain Monte Carlo method.
   1. Assess the convergence of your inference.
   2. Define and use a positive control to assess the accuracy of the numerical method.
2. For documentation
   1. Produce a pdf file summary that documents the task, your approach, any assumptions, your results, and includes at least one figure that illustrates your main finding(s).
   2. Please discuss your choices. For each choice, provide a brief overview of plausible choices and how you made your specified choice.
   3. Address whether (and if so how) your analysis is reproducible
   4. Include the code as an appendix in the pdf file
   5. Check whether you have all required citations
   6. Check whether you have assigned a copyright and a license to your codes
   7. Check whether the figure follows standard design praxis (see, for example, [this guide](https://www.nature.com/documents/nature-final-artwork.pdf))
   8. Submit your summary pdf as well as the codes to Canvas

**Problem Set #7: Link to Decision Analyses / Deep Uncertainties**

**Assigned:**  Friday, January **x** 2023 via Canvas

**Due:** Friday, January x, 2023 via Canvas

**Problem:**

Revisit the aircraft fuel problem you analyzed in problem ser # 3 with a few additional twists.

There is a field you can land right away with an expected risk of damage to the airplane of 5% of total value. The risk for landing on the airfield 100 flight minutes away is 1 *per mil* of total value. For simplicity assume that as soon as the fuel runs out, you need to land and running out of fuel is the only failure mode. You are an experienced pilot and have a smart passenger.

You and the passenger remember that the person who flew before you in the aircraft did add gas to the tank. The problem is that your passenger is certain that the prior pilot was in a hurry and only filled up the tank to ¼ of the capacity. You, on the other hand, are certain that the previous pilot filled up the tank to the maximum of the capacity.

**Tasks:**

1. Redo the analysis for problem set #3 with these two deeply uncertain priors.
2. Consider the following objectives outlined below. Specify for each objective what is your decision?
   1. Your only objective is to minimize the risk of damage to the aircraft (interpreted here as the expected value of damages without discounting and without a utility function). You apply Laplace’s principle of insufficient reason.
   2. Now add the objective that you want to arrive with 90% reliability with 30 min of fuel left in the tank. You apply Laplace’s principle of insufficient reason.
   3. Now add the objective that you want to “robustly” make the landing with 30 min fuel to spare. Robustly is here interpreted as achieving the objective in all deeply uncertain probabilistic scenarios.
3. For documentation
   1. Produce a pdf file summary that documents the task, your approach, any assumptions, your results, and includes at least one figure that illustrates your main finding(s).
   2. Please discuss your choices. For each choice, provide a brief overview of plausible choices and how you made your specified choice.
   3. Address whether (and if so how) your analysis is reproducible
   4. Include the code as an appendix in the pdf file
   5. Check whether you have all required citations
   6. Check whether you have assigned a copyright and a license to your codes
   7. Check whether the figure follows standard design praxis (see, for example, [this guide](https://www.nature.com/documents/nature-final-artwork.pdf))
   8. Submit your summary pdf as well as the codes to Canvas