F79BI: Bayesian Inference and Computational Methods
Assignment – Semester 2 Jan 2021
Total marks: 40; Assessment weight: 15%

Submission deadline: 1st April 2021

Purpose of the assignment

The purpose of the assignment is to:

- demonstrate good knowledge, understanding and application of MCMC methodology implemented in R, related to Bayesian inference;
- find creative solutions/arguments based on specialist knowledge in Bayesian Statistics;
- relate findings in a coherent well-argued text that is professionally presented.

Instructions

- **Marking.** This assignment will be marked out of 40 and will carry a 15% weight in the final mark for the course.
- **Submission deadline.** The deadline for submission of the assignment is **5.30pm (local time in UK or Malaysia) on the 1st of April, 2021**. Please make regular back-ups of your work and do not leave finalising it until the last minute as no allowance will be made for computer problems.

  The mark for coursework submitted late, but within 5 working days of the coursework deadline, will be reduced by 30%. Coursework submitted more than 5 working days after the deadline will not be marked. In a case where a student submits coursework up to five working days late, and the student has valid mitigating circumstances, the mitigating circumstances policy will apply. Students should be advised in such cases to submit a Mitigating Circumstances form for consideration by the Mitigating Circumstances Committee.
- **Form of solution.** Your solution to the assignment should take the form of a report. It should be prepared using LaTex, MS Word or other word-processing software. Relevant computer code must be included in an Appendix.
- **Length.** Reports should not exceed **8** pages in length. This limit excludes the R code, which should be placed in an Appendix.

- **How to submit.** Assignments must be submitted online on Vision through Turnitin. Online submissions will be subjected to a plagiarism check.

- **Information to be provided.** The cover page of your assignment should include the following information (see final page of this Assignment for example):

  - the text 'F79BI: Assignment';
  - your full name;
  - your HWU registration number (Person ID);
  - the names of other students that you discussed the assignment with;
  - a signed plagiarism declaration following the wording given at the end of this document (you can either add an electronic signature, or just print your name).

  Assignments which omit a signed plagiarism declaration will automatically be awarded zero marks.

- **Use of English.** Answers that are difficult to follow because of poor use of English might lose marks, so it is important to read through your work carefully before handing it in.

- **Figures and tables.** These are usually necessary to present the results of empirical work clearly. If you include figures and tables in your report, the discussion of their contents should be contained in the main text and not in captions. Please ensure that figures are easy to interpret and use clear labelling and legends.

- **Group discussions.** You may discuss this project with your classmates. However, you must conduct your own independent analyses and write your report independently of other students in your class. You will need to form your own conclusions in answering all parts of the assignment. The lecturer naturally expects that you will employ methodology that has been presented or discussed in the course but you are also encouraged to exercise curiosity and show originality.

- **Software problems.** You should be able to conduct the empirical part of this question by starting with the code examples we have discussed in the course and then adapting them. It is your own responsibility to get R code to run and to learn how to interpret the results of different functions and packages.

- **Plagiarism and collusion.** Your report must be written independently of other students in your class and all programming work must be your own individual work.

Failure to reference work that has been obtained from other sources or to copy the words and/or code of another student is plagiarism and if detected, this will be reported to the School's Discipline Committee. If a student is found guilty of plagiarism, the penalty could involve voiding the course.

Students must never give hard or soft copies of their coursework reports or code to another student. Students must always refuse any request from another student for a copy of their report and/or code.

Sharing a coursework report and/or code with another student is collusion, and if detected, this will be reported to the School's Discipline Committee. If found guilty of collusion, the penalty could involve voiding the course.

- **Feedback.** We will endeavour to provide feedback (both verbally in sessions and in writing on VISION) within 21 days of submission. Please contact the lecturers for individual feedback on particular questions.

Assessment criteria

The following aspects will be considered for marking:

- R coding, calculations, suitability of plots;
- derivation of results and relevant comments, discussion and interpretation;
- clarity of writing and general mathematical exposition;
- giving detailed answers, with appropriate explanations and good structure.

Guidelines:

- Write up your results in clear English.
- Use clear mathematical formulae to define any distribution, statistical quantity or property that you are analysing. If these are unclear you will lose marks.
- Reread everything you write to look for vagueness, ambiguity, and illogical reasoning. This means taking a step back from the detailed analysis.
- Define clearly all of the assumptions that you are using.

Broadly, the assessment criteria are as follows:

70% or higher (A grade)

Structure

- structures assignment effectively to facilitate development of argument.

Content

- displays extensive, detailed and secure knowledge and understanding of the subject
- applies mathematical methods fully accurately to support and develop argument
- demonstrates clear knowledge and understanding of qualitative and quantitative aspects of the question.

Argument

- engages directly with the question and appreciates wider implications and context
- presents a clear, coherent and persuasive argument based on correct mathematics
- displays independence or originality of judgment.

Expression

- uses fluent and accurate prose
- very good standard of presentation.

60-69% (B grade)

Structure

- structures assignment to facilitate development of argument.

Content

- displays extensive and secure knowledge and understanding of the subject

- applies mathematical methods mostly accurately to support and develop argument
- demonstrates sound understanding and knowledge of qualitative and quantitative aspects of the question.

Argument

- engages critically with the question and displays appreciation of the wider implications and context
- presents and develops ideas logically and persuasively
- demonstrates some independence of judgment and initiative.

Expression

- uses clear and generally accurate prose
- good standard of presentation.

50-59% (C grade)

Structure

- broadly structures assignment but organisation of ideas and evidence is sometimes determined by material rather than by the need to develop and support a logical argument.

Content

- displays sound and largely accurate knowledge and understanding of subject
- applies mathematical methods to support and develop argument but there are issues concerning the accuracy of results and the applicability of ideas
- demonstrates limited understanding and knowledge of qualitative and quantitative aspects of the question.

Argument

- displays understanding of the questions set but may lack sustained focus and appreciation of the wider context

- states ideas but may not develop them sufficiently or order them in a logical sequence.

Expression

- prose conveys meaning but lacks sophistication needed to present ideas persuasively
- expression may be clumsy with narrow vocabulary and spelling or grammar errors
- adequate standard of presentation.

1 Problem Description

Daily sales of a store are known to be affected by whether the store has announced super discounts on that day or not. From past experience, it is known that the fluctuation of the amount of daily sales around its mean value is normally distributed; if Y denotes the amount of daily sales, then $Y \sim N(\mu, \psi)$ where μ and ψ denote the unknown mean and precision parameter respectively. Super discount days may change the mean level, but leave the precision unaffected. To investigate the extent to which super discount days influence μ, a random sample of $n = 100$ daily sales amounts is obtained. The sample is represented as $\underline{y} = (y_1, y_2, \cdots, y_n)$ where $y_i \sim N(\mu_i, \psi)$, independently for $i = 1, 2, \cdots, n$. An indicator variable d_i, for $i = 1, 2, \cdots, n$, denotes whether the i-th observation was taken on a day when super discount was announced (corresponding to $d_i = 1$). The following equation gives a model for μ_i that captures the effect of d_i:

$$\mu_i = \alpha + \gamma d_i \tag{1}$$

for $i = 1, 2, \cdots, n$. The store provides you with data on $\underline{y}$ but is unable to provide the data on $\underline{d} = (d_1, d_2, \cdots, d_n)$. You gather information from them that super discounts days are held on $100p\%$ $(0 < p < 1)$ of days in a year, independently of each other. The departmetal store believes that p is around 0.4. The data for $\underline{y}$ (in units of ten thousand dollars) are given below and can be copied and pasted in R:

```
y <- c(
12.21,  9.57, 12.50, 10.55, 10.22, 10.67, 10.24, 11.43,  9.86, 12.25,
11.95,  9.41,  9.48,  9.48, 10.83, 11.84, 12.60, 10.04, 11.97, 10.87,
12.58, 12.01,  9.31, 10.38, 11.49, 12.32,  9.62, 12.42, 10.19, 11.78,
10.02, 12.40, 11.59, 10.05, 12.15, 10.76,  8.92,  9.75, 12.53,  9.41,
10.13, 10.24, 10.85, 12.05,  9.95, 12.00,  9.80, 11.41, 10.01, 11.88,
 9.92, 11.50,  9.31, 12.22, 12.17, 11.40, 11.87, 10.47, 12.43, 10.37,
11.76, 10.58, 10.29,  9.02, 10.68, 10.29, 10.04, 12.20,  9.30, 12.13,
12.73, 11.68, 10.43,  9.78, 10.70, 12.33, 11.52, 10.42, 10.16,  9.64,
12.33, 11.97, 11.34, 12.78,  9.59,  9.19, 11.07, 12.24, 12.90,  9.61,
 9.80,  9.45, 10.14,  9.66, 10.44, 12.09,  9.78, 11.94, 11.86, 10.15
)
```

In order to carry out a Bayesian inference procedure, you assume the following priors on the unknown parameters: $\alpha \sim N(\mu_\alpha, \sigma^2_\alpha)$, $\gamma \sim N(\mu_\gamma, \sigma^2_\gamma)$, $p \sim \text{Beta}(\alpha_p, \beta_p)$ and $\psi \sim \text{Gamma}(\alpha_\psi, \beta_\psi)$, independently of each other, where the hyperparameters μ_α, σ^2_α, μ_γ, σ^2_γ, α_p,β_p, α_ψ and β_ψ are pre-specified fixed and known values.

2 Your Task

Your tasks are the following:

1. Present appropriate graphical representations and summary statistics of the data. What is the interpretation of α and γ in terms of daily sales for the store? Find approximate values for α and γ based on your graphical representations.

   (4)

2. Identify the distribution corresponding to y_i for each $i = 1, 2, \cdots, n$ and write down the likelihood of the observed data $\underline{y}$, given α, γ, ψ and p. **(3)**

   HINT: Obtain the marginal distribution of y_i by averaging over the distribution of d_i.

3. For each $\theta \in \{\alpha, \gamma\}$, choose a prior specification on θ given by $\mu_\theta = 0$ and $\sigma_\theta^2 = 100$. Justify these choices. Choose a prior specification for p based on the text in Section 1 with prior variance 0.12. Choose a prior specification for ψ with prior mean 1 and prior variance 100. **(4)**

4. Describe how you will investigate the joint posterior distribution of $(\alpha, \gamma, \psi, p)$ using the techniques of data augmentation and Gibbs sampling. Give a clear description of how to update each component of your augmented parameter vector in your algorithm. **(12)**

5. Implement your approach in task 4 using R and use your code to explore the posterior distribution of $(\alpha, \gamma, \psi, p)$. Your answer should include: how you would choose starting values for the Markov Chain; graphical evidence that the Markov Chain has mixed satisfactorily; posterior inferences for parameters $(\alpha, \gamma, \psi, p)$ (e.g. with appropriate graphical or numerical summaries); conclusions for the daily sales problem. **(17)**

END OF ASSIGNMENT

F79BI Bayesian Inference and Computational Methods: Assignment

Full name:

HWU registration number (Person ID):

Plagiarism declaration:

I confirm that I have read and understood: (a) the note on *Plagiarism and collusion* in the assignment handout; (b) the Heriot-Watt University regulations concerning plagiarism.

I confirm that the submitted work is my own and is in my own words.

I confirm that any source (aside from course notes and lecture material) from which I obtained information to complete this assignment is listed in the assignment. Any sources not listed in the assignment are listed here: ...

Apart from the lecturer, I discussed the assignment and shared ideas with the following people: ...

Signature

Date