MATH4007 COMPUTATIONAL STATISTICS

Assessed Coursework 1 — 2022/2023

Your work should be submitted electronically via the module's Moodle page by **15:00 Thursday 15th December 2022**. Since this work is assessed, your submission must be entirely your own work (see the University's policy on Academic Misconduct). Submissions up to five working days late will be subject to a penalty of 5% of the maximum mark per working day.

Submission requirements

The submission should be uploaded electronically via the submission box on Moodle, and contain:

1. A pdf file containing any computational results (plots/relevant output) and discussion. This can be produced using e.g. R Markdown, or by copying output into a Word document. Please convert any documents **to pdf** for uploading.

2. A pdf of your theoretical working. A scan of handwritten work is fine, but you could also typeset using Latex if you prefer. If it's more convenient, you can combine this and the above part into one, e.g. if you wish to put everything in one Latex document, but this is not required.

3. An R script file, i.e. **with a .r extension** containing your R code. This should be clearly formatted, and include *brief* comments so that a reader can understand what it is doing. The code should also be ready to run without any further modification by the user, and should reproduce your results (approximately, for simulation-based results).

Please make sure that all required working, results, details of implementation and discussion are contained in **components 1 and 2** of the above list and not in the script file. The work will be assessed based on the working, output and discussion in these components, and the script file will only be used for verification of results. The exception is for the R code itself, whereby it is sufficient to say "refer to script file" where a question asks you to write R code.

A complete submission consists of all the files in your final submission. The submission time of the work will be based on the time at which the submission is complete, i.e. all files are uploaded. **Please carefully check after uploading your work that the files you upload are the correct ones.** Updates to any part of the submission after the deadline will be considered a new submission and late penalties will be applicable.

Questions

1. Consider a random variable X with probability density function (pdf)

$$f(x; r, R) = \frac{1}{2\pi}\left[1 + \frac{r}{R}\cos(x)\right],$$

where $0 \leq x \leq 2\pi$ and r and R are parameters satisfying $0 < r < R$.

   (a) Describe the rejection algorithm using a uniform proposal distribution.

   (b) Write an R function to implement your rejection method. The function should take as inputs the parameters r and R, and also store the number of attempts needed to successfully produce one accepted sample of X.

   (c) For the case $r = 0.9$, $R = 1$, use your function to sample 10000 values of X. Plot a histogram of your samples and overlay the true pdf, to check the sampler is working.

   (d) How does the theoretical efficiency (in terms of expected number of attempts needed to successfully produce one accepted sample of X) depend on the parameters r and R?

   (e) For various values of r and R, use your R function to illustrate empirically your answer in (d).

[16]

2. Data are available on the amount of a medicine in the blood (y) at $n = 10$ different times, in hours, after administration (x):

x	1	2	3	4	5	6	7	8	9	10
y	1.3	1.9	2.4	2.5	2.4	2.5	2.6	2.7	2.7	2.7

Consider the nonlinear regression model

$$y_i = b_1 \exp\left\{1 - \exp\left(-\frac{x_i}{b_2}\right)\right\} + \epsilon_i, \quad i = 1, \ldots, n,$$

where $b_1 > 0$ and $b_2 > 0$ are parameters and the $\epsilon_i \sim N(0, 0.01)$, $i = 1, \ldots, n$, are independent and identically distributed random errors. The times (x) were fixed and known without error, so the data are viewed as a random sample of y for x fixed. Inference for the unknown parameters b_1 and b_2 is required.

Here, we'll perform a Bayesian analysis. To complete the model, we give b_1 and b_2 independent exponential prior distributions, each with parameter λ, where the pdf of the exponential distribution is $f(z) = \lambda\exp(-\lambda z)$. We'll use $\lambda = 0.01$ for a diffuse prior on b_1 and b_2.

   (a) Derive the posterior distribution $\pi(b_1, b_2|\boldsymbol{x}, \boldsymbol{y})$, up to proportionality in b_1 and b_2, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the vectors of data.

   (b) Give details of, and write R code to implement, the 2-d steepest ascent algorithm to find the mode of the posterior distribution $\pi(b_1, b_2|\boldsymbol{x}, \boldsymbol{y})$. HINT: Work with the log of the posterior!

   NOTE: You are not expected to reparameterize to ensure b_1 and b_2 are positive — you should start suitably close to the optimum and use a small enough step size so that the algorithm converges to the optimum of interest.

(c) The marginal posterior distribution of b_2 is

$$\pi(b_2|\boldsymbol{x},\boldsymbol{y}) = \int_0^\infty \pi(b_1, b_2|\boldsymbol{x},\boldsymbol{y})db_1.$$

Give full details of Laplace's method to compute $\pi(b_2|\boldsymbol{x},\boldsymbol{y})$ at a particular point b_2.

(d) Write a function in `R` to compute $\pi(b_2|\boldsymbol{x},\boldsymbol{y})$ at a particular point b_2 using Laplace's method derived in (c).

(e) Plot $\pi(b_2|\boldsymbol{x},\boldsymbol{y})$ using your function from (d).

(f) Write a function in `R` to perform the Golden-ratio method to find the mode of $p(b_2|\boldsymbol{x},\boldsymbol{y})$, using your `R` function from part (d) as the function to optimize.

(g) Hence, find the mode of $p(b_2|\boldsymbol{y})$ to an accuracy of 1 decimal place.

[24]