MATH70093– Computational Statistics
Group coursework—Autumn 2021

Submit by 1pm on Friday 3 December 2021. This work needs to be solved with your allocated group. Write a joint report and upload it as a PDF file with **at most 5 pages** (excluding appendices). Your report should include an abstract with the following information:

- Group member names and CIDs
- The contribution of each group member to the assessment. This can include the questions that a member helped to solve, which parts of the report did they write or reviewed. If a group member does not take part in the assessment, it should be noted in the assessment and reported to the Course Director/ Module Lecturer.
- Only one submission is needed for each group, so only one of you needs to submit the final report on Blackboard via Turnitin. The person who submits the report should share the proof of submission to all the students in the group. It is the responsibility of each group member to ensure that the report has been submitted properly and on a timely manner.

Considerable emphasis will be put on clarity of expression, quality of presentation and on the depth of understanding. Ensure that your answers are well written, organised and are in the form of properly written sentences that include your full statistical reasoning. Provide your `R` code in appendix – do not use code in your essay in favour of mathematical equations.

Download your group dataset $\mathcal{D} = \{Y_i, X_{1i}, X_{2i}\}_{i=1}^{30}$ available on Blackboard. Consider the following linear model

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i, \quad \epsilon_i \sim_{\text{i.i.d}} \mathcal{N}(0, \omega) \quad \text{for } i = 1, \dots 30 \tag{1}$$

where β_1, β_2 and ω are fixed unknown parameters.

1. Assuming the following prior distributions $\beta_1, \beta_2 \sim \mathcal{N}(0, 5)$, $\omega \sim$Gamma(1,1) independently, derive the posterior distribution $p(\beta_1, \beta_2, \omega|\mathcal{D})$ up to a normalizing constant. Clearly define all the terms.
2. Construct an MCMC algorithm to sample from the joint distribution of the three unknown parameters given the data $\mathcal{D}$ and the specified priors. Provide a pseudo-code describing how the MCMC algorithm is constructed in this case. Justify your choice of proposal distribution. Verify graphically that the MCMC algorithm has reached convergence and assess the mixing of the Markov chain.
3. Compute an estimate of the posterior mean for each parameter. Explain mathematically how you obtain these estimates. You might want to consider a burn-in period.

Hint: To avoid numerical errors you might need to evaluate the logarithm of the target (posterior) distribution, $f(x)$, and use the formula $f(y)/f(x) = \exp(\log(f(y)) - \log(f(x)))$ when computing the acceptance probability. Note that the R functions `dnorm` and `dgamma` allow to compute the logarithm of the densities with the option `log=T`.