MATH70093 – Computational Statistics
Coursework 3 – Autumn 2021

Submit by 5pm on Friday 10 December 2021. Upload your report on Turnitin. **Hand-in no more than 10 pages** (excluding appendices). **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 an appendix – do not use code in your essay in favour of mathematical equations.

State the two last digits of your CID at the beginning of the report. Download your individual dataset $\mathcal{D} = \{y_t\}_{t=1}^{50}$ available on Blackboard. Consider the following model

$$Y_t = \alpha\, t + \beta \cos(\omega\, t) + \epsilon_t, \quad \epsilon_t \sim_{\text{i.i.d}} \mathcal{N}(0, \sigma^2) \quad \text{for } t = 1, \ldots 50 \tag{1}$$

where α and β are fixed unknown parameters; and $\omega = \pi/10$ and $\sigma^2 = 0.1$ are known. Assume the following prior distributions: α and β follow a Student's t-distribution with 5 degrees of freedom independently.

1. State the posterior distribution $p(\alpha, \beta|\mathcal{D})$ up to a normalizing constant. Clearly define all the terms.
2. Construct an MCMC algorithm to sample from the joint distribution of the two 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.

   Hint: To avoid numerical errors you might need to use the formula $p(\alpha, \beta|\mathcal{D})/p(\alpha', \beta'|\mathcal{D}) = \exp(\log(p(\alpha, \beta|\mathcal{D})) - \log(p(\alpha', \beta'|\mathcal{D})))$ when computing the acceptance probability.
3. Produce a scatter plot of the approximate sample from the joint distribution $p(\alpha, \beta|\mathcal{D})$.
4. 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.
5. The posterior predictive distribution of y_t^* for a given t is defined as follows

   $$p(y_t^*|\mathcal{D}, t) = \iint p(y_t^*|\alpha, \beta, t) p(\alpha, \beta|\mathcal{D}) d\alpha\, d\beta\ .$$

   (a) Let $t = 55$. Estimate and plot $p(y_{55}^*|\mathcal{D}, t = 55)$ as a function of y_{55}^*.

   (b) Let Y_t^* be a random variable following the distribution with p.d.f. $p(y_t^*|\mathcal{D}, t)$. Estimate $E(Y_t^*)$ as a function of t. Produce a plot comparing $E(Y_t^*)$ and the data as a function of t for t varying from 0 to 60.
6. Suppose that we would like to compare the model $\mathcal{M}_1$ described by equation (1) to the following model, $\mathcal{M}_2$:

   $$y_t = \beta \cos(\omega\, t) + \epsilon_t, \quad \epsilon_t \sim_{\text{i.i.d}} \mathcal{N}(0, \sigma^2) \quad \text{for } t = 1, \ldots 50 \tag{2}$$

   where $\omega = \pi/10$ and $\sigma^2 = 0.1$ are known. Consider the same prior distribution for β as in model $\mathcal{M}_1$.

   (a) Using the output of the MCMC sampler from question 2, compute a 95% credible interval for α in model $\mathcal{M}_1$. Discuss the relative evidence in favour of models $\mathcal{M}_1$ and $\mathcal{M}_2$ based on this credible interval.

   (b) Devise a Reversible Jump MCMC to estimate $p(\mathcal{M}_1|\mathcal{D}) = 1 - p(\mathcal{M}_2|\mathcal{D})$. Precisely describe how the Reversible Jump MCMC algorithm is constructed in this case, specifying all the terms and functions involved in the algorithm. Comment on the acceptance rate of the proposed Reversible Jump MCMC and verify graphically that the MCMC algorithm has reached convergence. Assuming equal prior probability for both models, estimate $p(\mathcal{M}_1|\mathcal{D})$ using the output of the sampler. Discuss the result and compare it to the result from question 6(a).
7. Suppose now that we are interested in inferring the parameter ω in model $\mathcal{M}_1$, assuming a uniform prior distribution between 0 and π. Assume that the parameters α and β are fixed; you can use the values obtained in question 4. Discuss the performance of a simple random-walk Metropolis Hasting algorithm to sample from $p(\omega|\alpha, \beta, \mathcal{D})$ including the impact of the choice of the variance of the proposal distribution.
