MATH97125 – Computational Statistics
Coursework 3 – Autumn 2020

Submit by 12pm (noon) on Friday 11 December 2020. Upload your final version only - once the report is uploaded there is no option for re-uploading. Avoid last minute uploads. **Hand-in no more than 8 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.

1. *(60% of total mark)* Consider the following density on $\mathbb{R}^2$

$$f(x, y) = k \exp\left\{-\frac{x^2}{100} - \left(y + \frac{3}{100}x^2 - 3\right)^2\right\}$$

where k is a normalising constant.

   (a) Construct an MCMC algorithm to sample from the density f. Provide a pseudo-code describing precisely the MCMC algorithm you constructed. 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 $f(x,y)/f(x',y') = \exp(\log(f(x,y)) - \log(f(x',y')))$ when computing the acceptance probability.

   (b) Compute an estimate of $E(X)$ and $E(Y)$ where (X, Y) follows the distribution with density f. Explain mathematically how you obtain these estimates. You might want to consider a burn-in period.

   (c) Read the article entitled *Adaptive proposal distribution for random walk Metropolis algorithm* by Heikki Haario, Eero Saksman and Johanna Tamminen – available on Blackboard. Describe using mathematical equations and a pseudo-code a MCMC algorithm to sample from the density f using the adaptive proposal described in section 3 of the article. Implement this MCMC sampler. Explain the advantages of this adaptive MCMC sampler compare to the MCMC sampler constructed in question 1(a).

2. *(40% of total mark)* Consider the following target density

$$g(\theta) = k\left(\frac{4}{10}\exp\{-(a-\theta)^2\} + \frac{6}{10}\exp\{-(30-\theta)^2\}\right)$$

where a is the last digit of your CID and k is a normalising constant.

   (a) Demonstrate that (simple) random-walk Metropolis-Hastings algorithms will not explore the full space well.

   (b) Implement a parallel tempering algorithm to sample from g. Provide a pseudo-code describing the algorithm you constructed. Precisely state the temperatures you are using as well as the transition kernels for each chain. Verify graphically that the sampler is exploring the full space and correctly sampling from the target distribution.

   (c) Estimate $P(\theta > 20)$ where θ follows the distribution with density g. Check your results using numerical integration.

As this is assessed work you need to work on it INDIVIDUALLY. It must be your own and unaided work. You are not allowed to discuss the assessed coursework with your fellow students or anybody else. All rules regarding academic integrity and plagiarism apply. Violations of this will be treated as an examination offence. In particular, letting somebody else copy your work constitutes an examination offence.