STA 3431 (Monte Carlo Methods), Fall 2020

Homework #2 Assignment: worth 25% of final course grade.

Due: On Quercus by 1:00 p.m. **sharp** (Toronto time) on Friday November 6.

GENERAL NOTES (reminder):

- **Late homeworks, even by one minute, will be penalised!**
- Include at the top of the first page: Your name and student number and department and program and year and e-mail address.
- Homework assignments are to be solved by each student individually. You may discuss assignments in general terms with other students, but you must solve them on your own, including doing all of your own computing and writing.
- You should provide very complete solutions, including explaining all of your reasoning very clearly, performing detailed Monte Carlo investigations including multiple runs and error estimates and alternative approaches, justifying the choices you make, etc.
- You should always consider in detail the accuracy and consistency of your answers.
- You may use results from the lectures, but clearly indicate when you do so.
- When writing computer programs for homework assignments:
  - R is the "default" computer programming language and should normally be used. You may use other standard computer languages like C or C++ or Java or Python if you wish, but please explain that.
  - You should include your complete source code and your program output.
  - Programs should be clearly explained, with comments, so they are easy to follow.
  - Even if you have explained your code well in the comments, you also need to explain your algorithm and ideas clearly in the main text. Explanations are very important!
- Upload one separate single pdf file for each question, to the course's Quercus page under the Assignments tab. Make sure your file includes your detailed solution with full explanation, plus your source code, plus your program output.

THE ACTUAL ASSIGNMENT:

1. [6] For this question, again let A, B, C, and D be the last four digits of your student number, and again let $g : \mathbf{R}^5 \to [0, \infty)$ be the function defined by:

$g(x_1, x_2, x_3, x_4, x_5)$

$$= \; x_1^{A+6}\, 2^{x_2+3} \Big(1 + \cos\Big[x_1 + 2x_2 + 3x_3 + 4x_4 + (B+3)x_5\Big]\Big)\, e^{(C-12)x_4^2}\, e^{-(D+2)(x_4-3x_5)^2} \prod_{i=1}^{5} \mathbf{1}_{0<x_i<1}\,.$$

Let $\pi(x_1, x_2, x_3, x_4, x_5) = c\, g(x_1, x_2, x_3, x_4, x_5)$ be the corresponding five-dimensional probability density function, with unknown normalising constant c. Estimate $\mathbf{E}_\pi[(X_1 - X_2)/(2 + X_3 + X_4 X_5)]$ using an MCMC algorithm of your choice, and obtain the best estimate you can. Include discussion of the reasons for your choices, and your results' accuracy, uncertainty, standard errors, confidence intervals, etc. Also, discuss the advantages and disadvantages of your approach compared to the methods that you used for this problem on Homework #1.

2. [8] Consider the standard variance components model described in lecture, with $K = 6$ and $J_i \equiv 5$, and $\{Y_{ij}\}$ the famous "dyestuff" data (from the file "Rdye"). Consider two sets of prior values: (i) $a_1 = b_1 = a_2 = b_2 = 6$, $a_3 = b_3 = 1600$; and (ii) $a_1 = a_2 = a_3 = b_1 = b_2 = b_3 = 50$. For each of the two sets of prior values, estimate (as best as you can, together with a discussion of accuracy etc.) the posterior mean of W/V, a random-walk Metropolis (RWM) algorithm. [Hint: You may wish to work with $\log \pi$ instead of π.]

3. [8] Repeat the previous question, but this time using a Componentwise Metropolis algorithm. Does it perform better or worse than the method in the previous question?

4. [8] Repeat the previous question, but this time using a Gibbs sampler. Does it perform better or worse than the methods in the previous two questions? [Note: You should first derive from scratch all of the required conditional distributions, whether or not they were already described in lecture.]

[**END; total points = 30**]

Reminders: No class on Nov 9 (Reading Week). Final projects are due on Friday Nov 27 at 1:00 PM sharp, with student presentations in class on Nov 30 and Dec 7.