# STA 3431 (Monte Carlo Methods), Fall 2019

## Homework #2 and Mini-Project: worth 30% of final course grade.

**Due:** In class at 10:10 a.m. **sharp** on Monday November 25.

**Or:** By e-mail to j.rosenthal@math.toronto.edu as a single pdf file (of reasonable size) by 9:30 a.m. on Monday November 25.

**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 it on your own, including doing all of your own computing and writing.
- For full points, you should provide very complete solutions, including explaining all of your reasoning clearly and neatly, performing detailed Monte Carlo investigations including multiple runs and error estimates as appropriate, justifying the choices you make, etc.
- You may use results from lecture, 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 for homework (and tests). You may perhaps use other standard computer languages like C and C++ and Java and Python with prior permission from the instructor.
  - You should include your complete source code and your program output.
  - Programs should be clearly explained, with comments, so they are easy to follow.
  - You should always consider such issues as the accuracy and consistency of the answers you obtain.

**THE ACTUAL ASSIGNMENT:**

**1.** [12] 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:

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

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

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 $c$. Write programs (with explanation) to estimate $\mathbf{E}_\pi[(X_1 - X_2)/(2 + X_3 + X_4X_5)]$ using two different MCMC algorithms of your choice, and obtain the best estimate you can with each of them. Include discussion of the reasons for your choices, and their accuracy, uncertainty, standard errors, confidence intervals, etc. Also, discuss the advantages and disadvantages of your two approaches compared to each other and to the methods that you used for this problem on Homework #1.

**2.** 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) $a_1 = b_1 = a_2 = b_2 = 5$, $a_3 = b_3 = 1500$; and (ii) $a_1 = a_2 = a_3 = b_1 = b_2 = b_3 = 100$. For each set of prior values, estimate (as best as you can, together with a discussion of accuracy etc.) the posterior mean of $W/V$, in each of three ways:

**(a)** [8] With a random-walk Metropolis algorithm. [Hint: Perhaps work with $\log \pi$.]

**(b)** [8] With a Metropolis-within-Gibbs algorithm.

**(c)** [8] With a Gibbs sampler. [Note: first derive from scratch all of the conditional distributions, whether or not they were already described in lecture.]

**(d)** [4] Discuss the relative merits of all three algorithms for this example, for each of the two sets of prior values.

**3.** [20] **Mini-Project:** Find an interesting and challenging quantity to compute, and conduct a Monte Carlo investigation to compute it. The quantity could be inspired by a research paper that you read, or an application related to your own field of research, or a topic of general interest to you, related to anything from statistical inference to bioinformatics applications to artificial intelligence to card shuffling to game playing to astronomy – try to be creative. You should attempt to compute this quantity using various Monte Carlo methods (perhaps those developed in class, perhaps others), and investigate its success or failure, with computational evidence together with whatever theoretical analysis (e.g. standard errors, or geometric ergodicity, or ...) you can manage. (It is okay if some or all of the methods fail.) Finally, you should state your conclusions, regarding the value(s) you were trying to compute, and also regarding which Monte Carlo methods did or did not work well.

The main part of your project should be a **maximum of ten double-spaced pages**. However, it can also include an **Appendix** of any length, which may or may not be read by the grader. Full source code and program output for all software you write should be included in the Appendix (with appropriate summaries in the main part), with programs well commented and easy to follow. Your topic, motivation, methodology, and results, should all be **very clearly explained** within the main part of the project.

Your project should not directly repeat material from another course or project, though it could be related. If you do make any use of results or programs or ideas from other sources or other courses, then that is okay but this fact should be clearly explained.

It is intended that students will complete this assignment individually. However, if you wish, with advance permission from the instructor (at least two weeks before the due date), you may work in a group of 2–3 students on a correspondingly larger project.

Although the emphasis of this project is on computation, if you wish to focus more on some theortical or methodological issue(s) related to the course content then that could be okay too; contact the instructor if you are considering this.

This assignment is rather open-ended, and the amount of investigating that you do is up to you. However, your investigation and write-up should be quite substantial, involving significant amounts of discussion, analysis, and computer programming. You may contact the instructor with any questions. Good luck!

**[END; total points = 60]**

**Reminders:** There will be a sit-down test on Monday Nov 18 in class, worth 30% of your final grade, with no aids allowed. And, student presentations will be on Nov 25 and Dec 2.