STA 3431 (Monte Carlo Methods), Fall 2021

Homework #1 Assignment: worth 18% of final course grade.

Due: On Quercus by 11:00 p.m. **sharp** (Toronto time) on Friday October 8.

GENERAL NOTES:

- Homework assignments are to be solved by each student individually. You may discuss questions 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, considering in detail the accuracy and consistency of your answers, etc.
- 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. For this first assignment, you must use R for the first two questions (though if you wish then you can also solve them using a different language and then compare which is better). For the other questions, you may use another standard computer language 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!
  - If you are showing plots of very long runs, then consider only plotting e.g. every 100th value, to keep the file size manageable.
- When you are finished the assignment, then you should upload all of your solutions as one single pdf file, to the course's Quercus page under the Assignments tab. Make sure your file includes your detailed solutions to all of the questions, including full explanation, all relevant source code, and all relevant program output. (Please email the instructor if you have any problems uploading your solutions. And be sure to leave enough time for the actual upload.)
- Please also include your name and student number and department and program and year and e-mail address at the beginning of your file – thank you.
- **Late penalty:** If uploaded x hours late, then ceiling($x/12$) points off (out of 18).

THE ACTUAL ASSIGNMENT:

1. [4] Write a computer program in R to generate pseudorandom Uniform[0,1] numbers, using a method of your choice which is not identical to one already mentioned in class (either some other method besides LCG, or perhaps an LCG but with different parameters). Your program should just use simple arithmetic, and should not use any built-in randomness functions at all. Explain your reasons for your choice of method. Then, perform various random-number-generations and plots, plus several statistical tests of your own choosing, and discuss how random/uniform/independent/etc your generator seems to be.

2. [4] Write a computer program in R to compute a good "classical" (i.i.d.) Monte Carlo estimate (including standard error and 95% confidence interval) of $\mathbf{E}\left|Y^2 Z^5 \sin(Y^3 Z^2)\right|$, where $Y \sim \text{Exponential}(3)$ and $Z \sim \text{Normal}(0,1)$ are independent. (The bars mean "absolute value".) Your program should use your own pseudorandom function from the previous question, and should not use any built-in randomness functions (for any distributions). Run your program multiple times, and produce a final estimate. Then, using only the results of your Monte Carlo estimator (not numerical integration or any other comparison), discuss how accurate you think your estimate is, and why.

3. [5] For this question, let A, B, C, and D be the last four digits of your student number, in order. (So, for example, if your student number was 840245070*, then $A = 5$, $B = 0$, $C = 7$, and $D = 0$.) Then, let $g : \mathbf{R}^5 \to [0, \infty)$ be the function defined by:

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

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

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.

Identify the values of A, B, C, and D. (This should be easy!)

Then, write a computer program to get a good estimate of $\mathbf{E}_\pi[(X_1+X_2^2)/(2+X_3X_4+X_5)]$ using an importance sampler with your own choice of function "f". (You may use the computer's built-in pseudorandom functions if you wish.) Discuss the reasons for your own choice of f, and the extent to which your algorithm does or does not work well. Then, produce a final estimate, and discuss how accurate you think your estimate is.

4. [5] Repeat the previous question, but this time using a rejection sampler instead of an importance sampler, again with your choice of function "f". (And remember that rejection samplers are only valid once a certain inequality has been proven.)

[END; total points = 18]

*(Historical note: this was the instructor's actual student number when he was a UofT undergraduate student during 1984–88.)