University of Toronto Mississauga

STA312 H5S: Computational Statistics - Winter 2020

Instructor: Dr. Luai Al Labadi

Assignment # 2

Due Date: Wed. February 26th, 2020 at beginning of lecture. **Late assignments will not be accepted.** If you cannot make it to the lecture, you can submit through a colleague.
Electronic submissions will not be accepted.
Solve all questions. However, **not all questions on the assignments are marked**. The questions to be marked are not announced.

Instructions:

- In this assignment, you should use R only when it is required.
- Start all your R codes with the following information: Question Number, Assignment 1, Your Last Name, First Name, Student Number. For Example,
  # Question 1, Assignment 1
  # Last Name: ABCD, First Name: XYZW
  # St. #: 0123456789
- You may not alter the output (by typing or handwriting anything). Output should be directly copied and pasted to the assignment where needed. If you do not follow these rules, your assignment will not be accepted.

- Assignments should presently neatly. **Use the following format:**

  - Theoretical questions can be **hand-written** or **typed**.
  - Attach the cover page (provided) at the front of your assignment. (Be sure to include an appropriate title for your assignment and fill in/circle all the information required).

  - For marking purposes, include the question numbers/part letters (1a,b,c, etc.) in your answers.
  - Do not simply hand in pages of R output without further explanation. Only include the relevant tables or plots that are asked in each question. Make sure to interpret the results in plain English in terms of the problem, quote relevant numbers from the output, and give justifications as a part of your solutions.
  - Do not include unnecessary code or output in the body of the assignment. At the end, include an appendix with ALL your R code and output.
  - The appendix will be checked to make sure the work was completed individually by the student and that the instructions were followed.

Assignments are individual work. Only general discussion is permitted between students. You must hand in solutions in your own words. Do not let others see your solutions or your selected article. It is plagiarism (a serious academic offence) to submit solutions in other people's words (including but not limited to other students, the instructor's, solutions from previous years or courses, websites, etc). You are responsible for knowing and adhering to the University of Toronto's Code of Behaviour on Academic Matters (see course outline).

Answer All the following questions.
Follow the form of densities given in the formula sheet.

Question 1 Let X be a random variable from $Poisson(\lambda)$, where $\lambda > 0$. That is, $P(X = i) = p_i = \frac{e^{-\lambda}\lambda^i}{i!}$, $i = 0, 1, 2, \ldots$.

(a) Show that $p_{i+1} = \frac{\lambda}{i+1} p_i$.

(b) Use part (a) to write an algorithm to generate a sample from X.

(c) Based on your algorithm in part (b), write an **R** code to generate a sample of size 1000 from $Poisson(\lambda = 2)$. Check that the mean and the variance are of the generated sample is close to the theoretical value $\lambda = 2$. Report the values.

Question 2 Let $X_1, X_2, \ldots, X_n$ be an independent sample, where each X_i has a cdf F_{X_i}. Let $Y = \max(X_1, \ldots, X_n)$.

(a) Show that $F_Y(x) = \prod_{i=1}^n F_{X_i}(x)$.

(b) Suppose it is easy to generate a random variable from any of F_{X_i}. How can we generate a sample from $F_Y(x) = \prod_{i=1}^n F_{X_i}(x)$?

(c) Now assume that $n = 20$ and $F_{X_i}(x) = 1 - e^{-x}$, $x > 0$ (i.e. $X_i \sim Exponential(1)$). Based on part (b), write an **R** code to generate a sample of size 1000 from $F_Y(x) = \prod_{i=1}^n F_{X_i}(x)$. In your code, you may use the built-in R function "rexp" to sample X_i. For the generated sample, plot the relative frequency histogram. Make appropriate titles. Report the mean and the variance for the generated sample.

Question 3 In class, we have discussed Acceptance-Rejection algorithm and Box-Muller approach to generate a sample form $Z \sim N(0, 1)$. The following algorithm can also be used to generate $Z \sim N(0, 1)$:

- Generate $U_1, \ldots, U_{12}$ from $Uniform[-0.5, 0.5]$.
- Set $Z = \sum_{i=1}^{12} U_i$.

(a) Show that $E(Z) = 0$ and $Var(Z) = 1$. Here you need to find the exact values.

(b) Write an **R** code to generate a sample of size 1000 from $Z \sim N(0, 1)$ based on the new algorithm. For the generated sample, plot the relative frequency histogram. Make appropriate titles. Report the mean and the variance for the generated sample.

Question 4 The Acceptance-Rejection (AR) algorithm may applies to densities known up to normalizing costants. Specifically, let $f \propto \tilde{f}$. That is, $f(x) = \tilde{f}(x) / \int_{-\infty}^{\infty} \tilde{f}(x)dx = \tilde{f}/k$. Let $g(x)$ be the candidate density. Then

$$\frac{f(x)}{g(x)} \leq c \iff \frac{\tilde{f}(x)}{kg(x)} \leq c \iff \frac{\tilde{f}(x)}{g(x)} \leq ck \iff \frac{\tilde{f}(x)}{g(x)} < c',$$

where $c' = ck$. Thus, the AR algorithm becomes:

1. Generate $u \sim Uniform[0, 1]$.
2. Generate $y \sim g$.
3. If $u < \frac{\tilde{f}(y)}{c' g(y)}$, accept y and set $x = y$. Otherwise, return to setp (1).

Note, the probability of accepting one value of X is $1/c'$. Assume that we would like to generate a sample from the random variable X with density

$$f(x) \propto \tilde{f}(x) = \exp(-x^2/2)\left[\sin^2(6x) + 3\cos^2(x)\sin^2(4x) + 1\right].$$

Here, for example, $\sin^2(6x)$ means $(\sin(6x))^2$. Let the candidate density $g(x)$ be the $N(0,1)$. That is,

$$g(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2), \quad -\infty < x < \infty.$$

(a) Show that $\frac{\tilde{f}(x)}{g(x)} \leq 5\sqrt{2\pi}$. Hint: $|\sin(x)| \leq 1$.

(b) With $c' = 5\sqrt{2\pi}$, write the AR algorithm to sample from X with density $f(x)$.

(c) Write an **R** code to generate a sample of size 1000 from $f(x)$. In your code, you may use the built-in R function "rnorm" to sample g. For the generated sample, plot the relative frequency histogram. Make appropriate titles. Report the mean and the variance for the generated sample.

Good Luck