University of Toronto Mississauga

STA312 H5S: Computational Statistics - Winter 2020

Instructor: Dr. Luai Al Labadi

Assignment # 3

Due Date: Wed. March 25th, 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 3
  # 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.

Note: In all questions where the MC estimation is required, set $m = 10^4$ in your `R` code.

Question 1 Let $X_1, X_2, \ldots, X_m$ be a random sample from a (cumulative) distribution function $F(x)$. Define

$$I(X_i \le x) = \begin{cases} 1 & X_i \le x \\ 0 & X_i > x \end{cases}.$$

The *empirical cumulative distribution function* (ECDF) is defined as

$$F_m(x) = \frac{1}{m}\sum_{i=1}^{m} I(X_i \le x).$$

Prove for any x and y,

(a) $F_m(x)$ is unbiased estimator of $F(x)$.

(b) $\mathrm{Var}(F_m(x)) = F(x)(1-F(x))/m$.

(c) $E\left[(F_m(x) - F(x))(F_m(y) - F(y))\right] = (F(\min(x,y)) - F(x)F(y))/m$.

(d) Determine whether $F_m(x)$ and $F_m(y)$ are positively or negatively correlated. **Justify you answer**.

(e) Using R, generate a $Beta(1,1)$ variable with sample size 100. Make plot of the ECDF. Superpose on this plot the $Beta(1,1)$ cumulative distribution function. Make a title. Comment on the quality of the fit.

Question 2 Suppose that we wish to calculate $\theta = \int_0^1 g(x)dx$. The Monte Carlo method works in the following way. Generate independent uniform random variables $U_1, \ldots, U_m$ on $(0,1)$ and compute

$$\widehat{\theta}_1 = \frac{1}{m}\sum_{i=1}^{m} g(U_i).$$

(a) Show that $E(\widehat{\theta}_1) = \theta$ and that $\widehat{\theta}_1$ converges to θ in probability as $m \to \infty$.

(b) Use the central limit theorem to find m such that $P(|\widehat{\theta}_1 - \theta| > 0.05) \le 0.05$ if $g(x) = \cos(2\pi x)$.

(c) let f be a density on $(0,1)$. Generate a sample $X_1, \ldots, X_m$ from f and estimate θ by

$$\widehat{\theta}_2 = \frac{1}{m}\sum_{i=1}^{m} \frac{g(X_i)}{f(X_i)}.$$

Show that here again that $E(\widehat{\theta}_2) = \theta$ and that $\widehat{\theta}_2$ converges to θ in probability as $m \to \infty$.

(d) Find an expression for $\sigma^2 = Var_f(\widehat{\theta}_2)$.

Question 3 Let $X \sim \chi^2(8)$.

(a) Use Hit-or-Miss algorithm to find an estimate of $P(X > 20.5)$.

(b) Using `R`, provide the numerical value of the estimator in (a), its variance and the associated 95% confidence interval.

(c) Now estimate $P(X > 20.5)$ using importance sampling with an exponential density Eponential with mean 8.

(d) Using R, provide the numerical value of the estimator in (c), its variance and the associated 95% confidence interval. **Hint:** $rexp(m, rate = 1/8)$.

For Part (b) and Part (d), consult the example in Slide # 11 of Unit 4, Part II.

Good Luck