Computational Statistics 732A90 – Fall 2023
Computer Lab 3

Krzysztof Bartoszek, Bayu Brahmantio, Frank Miller, Héctor Rodriguez Déniz
Department of Computer and Information Science (IDA), Linköpings University

November 15, 2023

This computer laboratory is part of the examination for the Computational Statistics course. Create a group report, (that is directly presentable, if you are a presenting group), on the solutions to the lab as a .PDF file. Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments.
All R code should be included as an appendix into your report.
A typical lab report should 2-4 pages of text plus some amount of figures plus appendix with codes. In the report reference all consulted sources and disclose all collaborations.
The report should be handed in via LISAM (or alternatively in case of problems by email), by **23:59 22 November 2023** at latest. Notice there is a deadline for corrections **23:59 21 January 2024** and a final deadline of **23:59 11 February 2024** after which no submissions nor corrections will be considered and you will have to redo the missing labs next year. The seminar for this lab will take place **29 November 2023**.
The report has to be written in English.

Question 1: Sampling algorithms for a triangle distribution

Consider the following density with a triangle-shape (another triangle distribution than considered in Lecture 3):

$$f(x) = \begin{cases} 0 & \text{if } x < -1 \text{ or } x > 1, \\ x+1 & \text{if } -1 \le x \le 0, \\ 1-x & \text{if } 0 < x \le 1. \end{cases}$$

We are interested to generate draws of a random variable X with this density.

a. Choose an appropriate and simple envelope $e(x)$ for the density and program a random generator for X using rejection sampling.

b. In Lecture 3, another triangle distribution was generated using the inverse cumulative distribution function method, see page 9-10 of the lecture notes. Let Y be a random variable following this distribution. A random variable $-Y$ has a triangle distribution on the interval $[-1, 0]$. Program a random generator for X using composition sampling based on Y and $-Y$. You can use the code from the lecture to generate Y.

c. Sums or differences of two independent uniformly distributed variables can also have some triangle distribution. When U_1, U_2 are two independent $Unif[0, 1]$ random variables, $U_1 - U_2$ has the same distribution as X. Use this result to program a generator for X.

d. Check your random generators in each of a. to c. by generating 10000 random variables and plotting a histogram. Which of the three methods do you prefer if you had to generate samples of X? Use the data from one method to determine the variance of X.

Question 2: Laplace distribution

The double exponential (Laplace) distribution is given by formula:

$$DE(\mu, \lambda) = \frac{\lambda}{2} \exp(-\lambda|x - \mu|)$$

a. Write a code generating double exponential distribution $DE(0, 1)$ from $\mathrm{Unif}(0, 1)$ by using the inverse CDF method. Explain how you obtained that code step by step. Generate 10000 random numbers from this distribution, plot the histogram and comment whether the result looks reasonable.

b. Use rejection sampling with $DE(0, 1)$ as envelope to generate $\mathcal{N}(0, 1)$ variables. Explain step by step how this was done. How did you choose constant a in this method? Generate 2000 random numbers $\mathcal{N}(0, 1)$ using your code and plot the histogram. Compute the average rejection rate R in the rejection sampling procedure. What is the expected rejection rate ER and how close is it to R? Generate 2000 numbers from $\mathcal{N}(0, 1)$ using standard `rnorm()` procedure, plot the histogram and compare the obtained two histograms.