Projects - Quantile estimation

- R is the only programming language allowed in answers. Questions requiring R code are indicated by the symbol ♠
- The report (file name: groupe_rapport_noms, e.g., 001_rapport_robert_duche)
  - to be returned in .pdf format and must contain your answers and comments. Careful drafting is expected. It is important to justify/comment on the theoretical and numerical results
  - To integrate all or part of your code and outputs into your report, you can use the dedicated tools Notebook, Rmarkdown or LATEX+ knitr. However, it is forbidden to copy and paste raw code into the body of the text. Graphics must be carefully annotated and presented (title, colour, captions, etc.).
- A version of the code that can be tested must be provided (same file name). This code must:
  - run without errors and reproduce all the results presented in the report. You specify the seed used for the results obtained.
  - be well commented. You may be asked for an oral explanation.
  - use as much as possible the specificities of the language (bonus for the most efficient codes).
- The project will be emailed at the address stanislasduche@gmail.com with the object **groupe_name1_name2 - Monte Carlo Project Part 1**
- The project will be released in 3 times, with blocks of 3/4 exercises. **This is the first block of the project** to be returned for **Sunday October 27th 2024 at 23:59:59. Every hour late starting from October 28th, 00:00:01, will be penalized by a point.**
- The project can be carried out in groups of two or three. For groups of 3, a requirement for detail and initiative will be included in the marking of the project.
- I would remind you that plagiarism is strictly forbidden and any suspicion will result in a 0 for the exercise. The Large Language Models chatbots can be used as an aid; but any code suspected to be written by an assistant will not be corrected.

Exercise 1: Negative weighted mixture

Consider a random variable of a negative weighted mixture X following a density law proportional to

$$\forall x \in \mathbb{R}, \quad f(x) \propto f_1(x) - a f_2(x)$$

with $f_1(x) = \mathcal{N}(x; \mu_1, \sigma_1^2)$ and $f_2(x) = \mathcal{N}(x; \mu_2, \sigma_2^2)$ density laws of two normal distributions; and $a > 0$. The objective of this exercise is to understand the behavior of the random variable and to compute statistics.

PREPARED BY Stanislas du Ché

Definition

Question 1: Recall the conditions for a function f to be a probability density. Considering the tail behavior of f, derive necessary conditions on (σ_1^2, σ_2^2) for f to be a density.
Question 2: For given parameters $\theta_1 = (\mu_1, \sigma_1^2)$ and $\theta_2 = (\mu_2, \sigma_2^2)$, determine a bound a^* on a for $f(x) \propto f_1(x) - af_2(x)$ to be a well-defined density. Provide its normalization constant.
Question 3: (♠) Create an R function `f(a, mu_1, mu_2, s_1, s_2, x)` that produces the pdf of f as a function of x and of the parameters $a, \mu_1, \mu_2, \sigma_1, \sigma_2$. Plot on the same graph the pdf of f for different values of a, especially for $a = a^*$. In another graph, do the same for different values of σ_2, for σ_1 fixed. Examine the impact of a and σ_2 on the shape of f.
From now on, we set the following numerical values: $\mu_1 = 0, \mu_2 = 1, \sigma_1^2 = 3, \sigma_2^2 = 1$, and $a = 0.2$. Check they are compatible with the constraint derived above.

Inverse c.d.f Random Variable simulation

Question 4: (♠) Show that the cumulative density function associated with f is available in closed form. Create an R function `F(a, mu_1, mu_2, s_1, s_2, x)` that produces the cdf of f as a function of x and of the parameters $a, \mu_1, \mu_2, \sigma_1, \sigma_2$. Construct an algorithm that returns the value of the inverse function method as a function of $u \in (0, 1)$, of the parameters $a, \mu_1, \mu_2, \sigma_1, \sigma_2$, and of an approximation precision ε. Deduce an algorithm that implements the inverse function method for the generation of random variables from F.
Question 5: (♠) Write an R function `inv_cdf(n)` that generates n samples from f using the inverse function method. Generate `n = 10000` samples, and graphically check that `inv_cdf()` is correct.

Accept-Reject Random Variable simulation.

Question 6: Describe a method to simulate under f using the accept-reject algorithm and give the expression theoretical acceptance rate according to the parameters of f.
Question 7: (♠) Write a function `accept_reject(n)` that generates n samples from f using the accept-reject method. Generate `n=10000` samples, and graphically check that `accept_reject()` is correct. Compute the empirical acceptance rate and check whether it agrees with its theoretical value.
Question 8: (♠) Vary the value of a and plot the acceptance rate for different values of a. Describe its impact when $a \to a^*$.

PREPARED BY Stanislas du Ché