732A90: Computational Statistics

Computer lab3 - Group11

Sofie Jörgensen, Oriol Garrobé Guilera, David Hrabovszki

07 February 2020

Question 1: Cluster sampling

1.

2.

Question 2: Different distributions

In this question we are given the double exponential (Laplace) distribution with the location parameter μ and the scale parameter α, where the formula is given by

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

1.

The aim of this task is to generate random numbers from a double exponential distribution with $\mu = 0$ and $\alpha = 1$, that is $DE(0, 1)$, from a uniform distribution Unif$(0, 1)$ by using the inverse CDF method. For this, we will generate 10000 random numbers, and a plot histogram of the results.

Let,

$$X \sim DE(0, 1)$$

Then,

$$f_x(x) = \frac{1}{2}exp(-|x|)$$

Also,

$$F_x(x) = \int_{-\infty}^{x} f_x(s)ds = \frac{1}{2}exp(-|s|)ds$$

Therefore,

$$F_x(x) = \begin{cases} \frac{exp(x)}{2}, x < 0 \\ 1 - \frac{exp(-x)}{2}, x \geq 0 \end{cases}$$

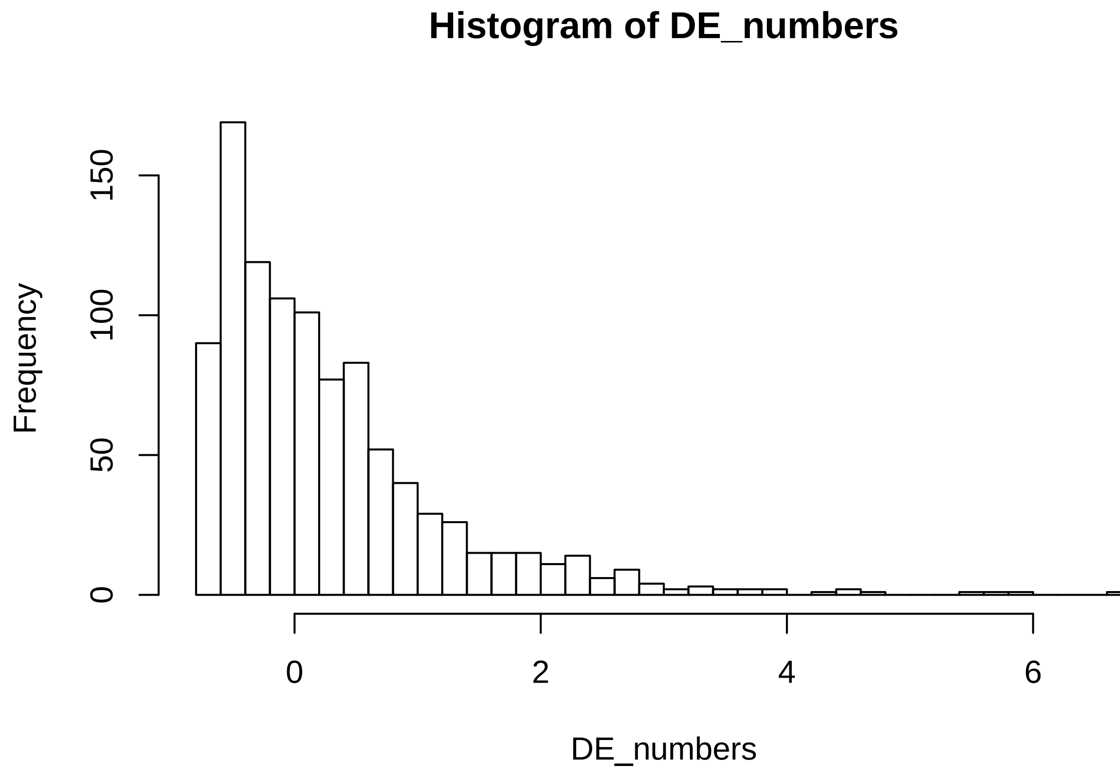
!!!QUESTION: Here I only use the function for x>0. Should we do it for bigger and smaller than 0????? !!!!

Using the CDF method we look for F_x^{-1}, which is:

$$F_x^{-1}(y) = -ln(2 - 2y)$$

Hence, if $U \sim U(0, 1)$, then

$$-ln(2 - 2U) = X \sim DE(0, 1)$$



Using this formula bla bla...

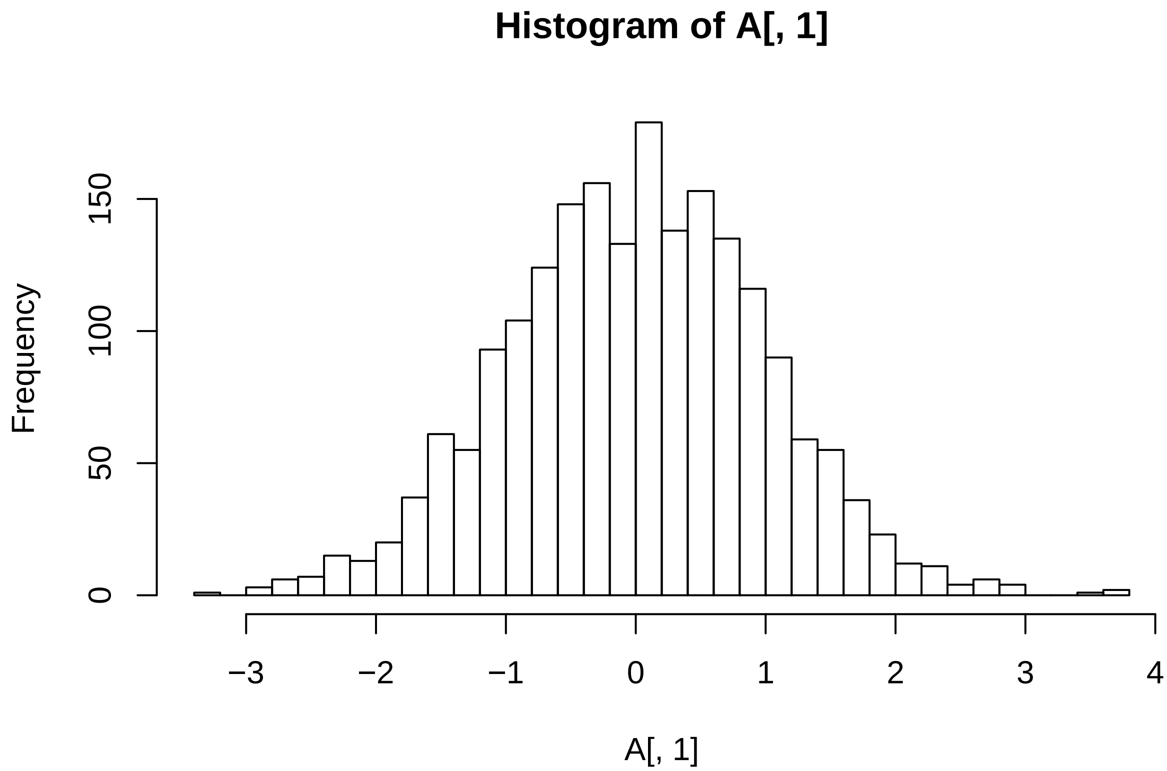
The histogram looks like an exponential???

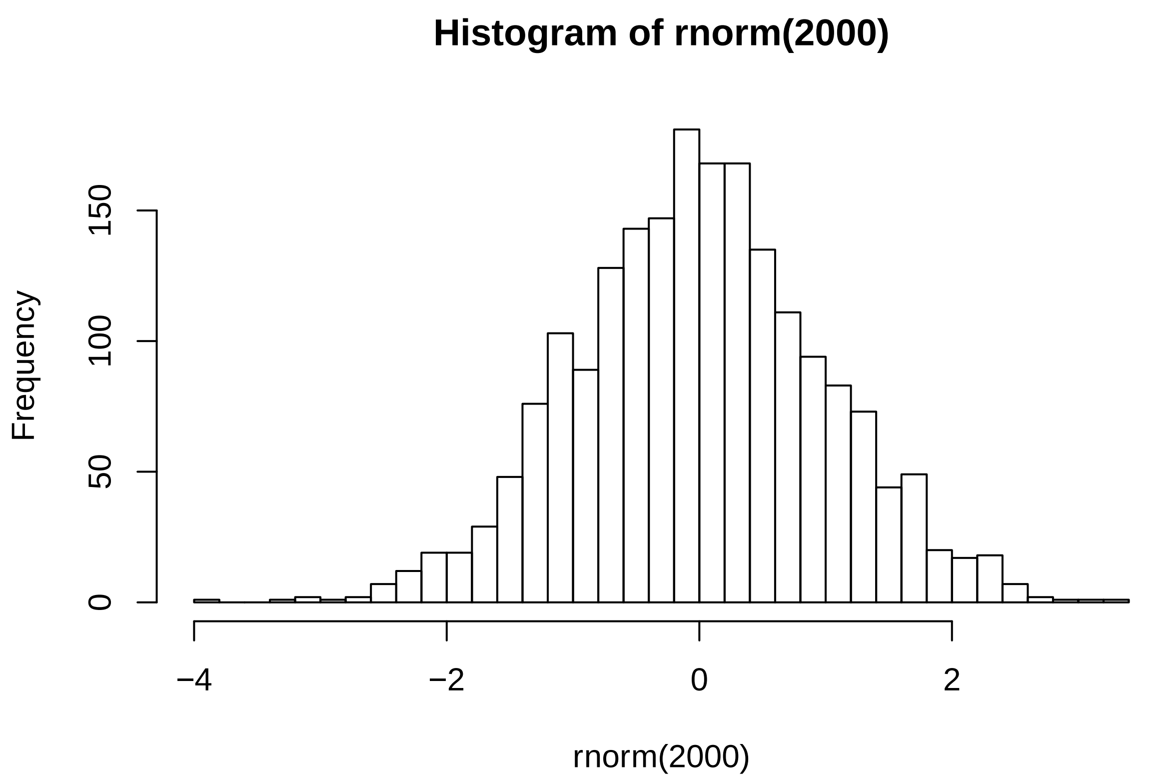
2.

Now we are going to apply the Acceptance/Rejection method by using the double exponential distribution obtained from Question 2.1, in order to generate standard normal variables. In other words, $DE(0,1)$ is the majorizing density f_Y and $N(0,1)$ is the target density f_X. The implementation of the Acceptance/Rejection method is taken from the example in the course material provided in the lecture. From Figure X (refer to histogram in Question 2.1) we can observe that the shape of the distribution looks quite similar to a standard normal distribution since it is centered around zero and have a similar spread, thus the choice of the $DE(0,1)$ can be considered as reasonable.

The idea is to generate a random number from $DE(0,1)$, i.e. $Y \sim f_Y$ and a random number from $U \sim \text{Unif}(0,1)$. Then if $U \leq \frac{f_X(Y)}{cf_Y(Y)}$ holds, then we accept the generated number Y, otherwise it is rejected. The rejection is controlled by the majorizing constant c, and a value of c closer to 1 will imply fewer rejections. This means that if $c = 1$, then f_Y and f_X are the same density function. Thus we wish to pick a majorizing constant c, close to 1 such that it fulfills the requirement that $c \cdot f_Y(x) \geq f_X(x)$ for all x. We test a sequence of numbers of c from 1 to 5, and obtained 1.32 as the smallest c such that the requirement is still fulfilled for any x.

Thereafter we generate 2000 standard normal random numbers with this setting. The result is presented in a histogram in Figure X. In the Acceptance/Rejection method, we compute the number of rejections R with `num.reject` and we obtain $R = 0.317$ for $c = 1.32$. The number of rejections plus the single draw, when the random number is accepted, must equal the total number of draws. Since the total number of draws is Geometric distributed with mean c, the expected rejection rate ER is given by $c - 1 = 1.32 - 1 = 0.32$. By computing the difference $ER - R = 0.003$ we can conclude that the expected rejection rate ER and the average rejection rate R are very close to each other. Also, the two histograms looks similar when comparing Figure X with Figure X for 2000 random numbers. Hence we are satisfied with our implemented random generator.


Histogram of A[, 1]
Frequency
0
50
100
150
−3
−2
−1
0
1
2
3
4
A[, 1]


Histogram of rnorm(2000)
Frequency
0
50
100
150
−4
−2
0
2
rnorm(2000)