MATH5835M Statistical Computing
Exercise Sheet 2

https://www1.maths.leeds.ac.uk/~voss/2023/MATH5835M/
Jochen Voss, J.Voss@leeds.ac.uk
2023/24, semester 1

This does not count towards your final mark, the questions are for self-study only.

Exercise 4. The t-test at significance level 5% accepts or rejects the hypothesis that data $x_1, \ldots x_n$ might be a sample from a probability distribution with mean μ, using the following procedure: We compute the test statistic

$$t = \sqrt{n}\,\frac{\bar{x} - \mu}{\mathrm{s}_x}, \tag{1}$$

where $\bar{x}$ is the sample mean and s_x is the sample standard deviation of the x_i, and we determine the 97.5%-quantile $t_{n-1}(0.975)$ of the t-distribution with $n-1$ degrees of freedom. If $|t| > t_{n-1}(0.975)$, we reject the hypothesis and if $|t| \leq t_{n-1}(0.975)$, we accept the hypothesis that the mean was μ.

A type I error occurs when the test wrongly rejects the hypothesis, while in fact the data was sampled using mean μ. If the data $x_1, \ldots, x_n$ are a random sample from a normal distribution, one can show that the probability of type I errors occurring equals 5%. In this exercise we will use Monte Carlo estimation to find the probability of type I errors, when the data is exponentially distributed.

Throughout this exercise we will assume $\mu = 2$ and that $X_1, \ldots, X_n \sim \mathrm{Exp}(1/\mu)$. Since it is known that the $\mathrm{Exp}(\lambda)$-distribution has expectation $1/\lambda$, the X_i then have expectation μ.

a) Assume $n = 10$, $\mu = 2$ and $X_1, \ldots, X_n \sim \mathrm{Exp}(1/\mu)$. Use the function `rexp()` in R to generate a random sample from this distribution, and then compute the value T of the test statistic (1) for this sample. Hint: Use a combination of the functions `sqrt()`, `mean()` and `sd()`.

b) The value T computed in the previous question can be used as a single sample in a Monte Carlo estimate. Use a loop in R to generate vector of N such samples. You can use $N = 100$ for now, but we will later change N. Hints: Be careful not to confuse N with n. You need to generate N groups of n exponentially distributed random numbers. Loops are described in appendix B.3.1.1 of the book.

c) Use the vector obtained in the previous part to get a Monte Carlo estimate for the probability of type I errors. Determine the (approximate) root mean-squared error of your estimate. Hints: You will likely need to use the function `abs()`. You can use the command `qt(0.975,` $n-1$`)` to compute the critical value $t_{n-1}(0.975)$.

d) Determine an N which is large enough to show that the probability of type I errors is different from 5%, but small enough that your computer can still compute the estimate in a reasonable amount of time.

e) Use a second loop to repeat your code from part (c) for different sample sizes n, and generate a plot which shows how the probability of type I errors depends on n.