Mandatory assignment 1

Deadline: Sunday September 29th at 23:59 (Norwegian time).

Read carefully through the information about the mandatory assignments on Canvas. Notice in particular that the assignments should be solved individually.

Hand in on Canvas. Submissions should be of **either** of the following types

- Submit two files: One pdf-file with a report containing the answers to the theory questions, and one R-file including the R-code.
- Submit two files: One R markdown (Rmd) file containing both theory answers and R-code, and a pdf-file with the output you obtain when running (knitting) you R markdown file. See tutorial to get started.

The first line of R-code should be: `rm(list=ls())`. Check that the Rmd/R-code file runs before you submit it. Use comments in the R-code to clearly identify which question each part of the R-code belong to. Also try to add some comments to explain important parts of the code. The file ending of the R-code file should be .Rmd, .R or .r. The report can be handwritten and scanned to pdf-file, or written in your choice of text editor and converted to pdf. Cite the sources you use.

Problems marked with an $^{\mathrm{R}}$ should be solved in R, the others are theory questions.

Problem 1:

We are interested to find out which of the events have the greatest probability.

a) Assume we throw three fair dices with numbers 1 to 6. Which event has the larger probability, the sum of the three dices is 9 or the sum is 10?

b) In this experiment, we want to find out if it is more likely to get at least one six out of 4 dice throws or to get at least one time two sixes when throwing two dices 24 times.

c) Finally, we want to find out which of these events are most likely,

   - to get at least one time a 6 when throwing a dice 6 times
   - to get at least two times a 6 when throwing a dice 12 times
   - to get at least three times a 6 when throwing a dice 18 times

d) What is the probability that we have to throw a dice 8 times in order to get three sixes?

e)$^{\mathrm{R}}$ Calculate all the probabilities above with 10000 simulations in R, i.e. the probabilities in a.), b.), c.) and d.).

Problem 2:

In this problem, we consider a random variable T with cumulative distribution function (CDF) given by

$$F_a(t) = e^{-e^{-(t-a)}}, \quad t \in \mathbb{R} \tag{1}$$

a) Verify that this is a proper cumulative density function.

b) Show that the mean is $E(T) = a + \gamma$ where $\gamma = -\int_0^\infty \log(x)e^{-x}dx \approx 0.5772$ is the Euler's constant (hint: consider integration by substitution, $s = (t-a)$ and $x = e^{-s}$).

c) Given $a = 5$, calculate $P(T \leq 4)$, $P(T > 9)$ and $P(5 < X \leq 6)$.

d) Show that the inverse cumulative distribution function is given by

$$F_a^{-1}(U) = a - \log(\log(U^{-1})).$$

e)$^{\mathrm{R}}$ Write a function (with two arguments n and a) which produces n random numbers with density (1) using the inverse transform method. Ideally, the function should not contain any for-loops. Check that your function produces correct results by comparing a histogram and kernel density estimate (KDE) of $n = 1000$ generated random numbers to the probability density of (1) with $a = 5$.

f)$^{\mathrm{R}}$ Calculate the sample average and compare the result with $E(T)$.

g)$^{\mathrm{R}}$ Calculate the sample variance. How many samples do we need to achieve a precision of 0.01 for the sample average?

h)$^{\mathrm{R}}$ Plot the CDF $F_a(x)$ and the ECDF based on $n = 1000$ simulated random observations. Compare both plots and comment on what you find.

Problem 3: The rate of return of shares can be modeled by a distribution with parameter λ that has the following density function

$$f_\lambda(x) = \frac{1}{2\lambda}\exp(-\frac{|x|}{\lambda}), \quad x \in \mathbb{R}, \lambda > 0, \tag{2}$$

where $|x|$ is the absolute value of x with $E(|X|) = \lambda$.

a) Supposed we have n independent observations $x_i, i = 1, \ldots, n$. Show that the maximum likelihood estimator (MLE) of λ is given by

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}|x_i|.$$

Is $\hat{\lambda}$ an unbiased estimator of λ?

b)$^{\mathrm{R}}$ Assume we observe day's rate of return of $1.3, -0.6, 0.2, 0.4, -0.8$ during one week. Calculate the MLE $\hat{\lambda}$ given this sample.

c)$^{\mathrm{R}}$ Plot the density function $f_\lambda(x)$ (2) for $x \in [-5, 5]$ and $\lambda = 0.6$.

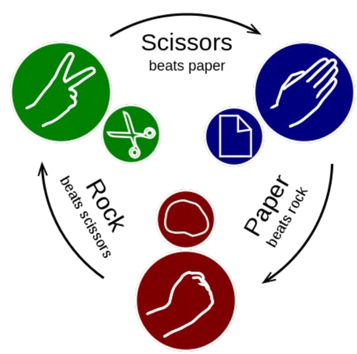
d)$^{\mathrm{R}}$ Write a function (with an argument n) which produces n random numbers with density $f_\lambda(x)$ (2) using the accept-reject method using a uniform(-5,5) proposal distribution $g(x)$. Check that your function produces correct results by comparing a histogram and KDE of $n = 500$ random numbers to the density $f_\lambda(x)$ (2) with $\lambda = 0.6$.

e)$^{\mathrm{R}}$ Investigate different bin-width rules for the histogram (Sturges' formula, Scott's formula, Freedman-Diaconis formula) and bandwidth rules for the KDE (Silverman's rule of thumb, Unbiased cross-validation, Biased cross-validation, Sheather & Jones). What are the selected bin-width and bandwidth? Comment on your findings!

f)$^{\mathrm{R}}$ It can be shown that the CDF of (2) is $F_\lambda(x) = 0.5 + 0.5 sign(x)(1 - \exp(-|x|/\lambda))$. Plot the CDF $F_a(x)$ and the ECDF based on $n = 500$ simulated random numbers. Comment!

Problem 4: Rock, Paper, Scissors (RPS)

According to Wikipedia *(https://en.wikipedia.org/wiki/Rock-paper-scissors)* **Rock-paper-scissors** is a zero-sum hand game usually played between two people, in which each player simultaneously forms one of three shapes with an outstretched hand. These shapes are "rock" (a simple fist), "paper" (a flat hand), and "scissors" (a fist with the index and middle fingers together forming a V). The game has only three possible outcomes other than a tie: a player who decides to play rock will beat another player who has chosen scissors ("rock crushes scissors") but will lose to one who has played paper ("paper covers rock"); a play of paper will lose to a play of scissors ("scissors cut paper"). If both players choose the same shape, the game is tied and is usually immediately replayed to break the tie. Other names for the game in the English-speaking world include roshambo and other orderings of the three items, sometimes with "rock" being called "stone".



Assume that the two players have equal possibilities of winning and that they choose R, P or S randomly. For one trial (in which each player simultaneously forms one of three shapes) the probability of a tie is then 1/3 and the probability that one of them wins is 2/3.

A variation of the game is that the players have a number of "lives". e.g. if the players have two lives, the one who is the first to lose two times (the two "lives"), loses the game and the other one wins.

If the players have only one live, the geometric distribution describes the numbers of trials until the first success, i.e. the number of trials until one of the players wins. Let $X =$ the number of trials until one of the players wins then

$$P(X = x) = (1 - p)^{x-1}p, \quad x = 1, 2, \ldots.$$

Let the random variable Y_k be the total number of trials until one wins when the players have k lives. Then at least k trials have to be made to decide a winner, and the possible outcomes of Y_k are $k, k+1, k+2, \ldots$. The task is now to use simulation to estimate the distribution of Y_k.

a)$^{\mathrm{R}}$ Write an R function which returns the number of trials needed to decide a winner when the players have k lives.

b)$^{\mathrm{R}}$ Check the routine by comparing the results of 10^5 simulations for $k = 1$ with the geometric distribution with probability of success $p = 2/3$.

c)$^{\mathrm{R}}$ Using 10^5 simulations, estimate the expected number of trials until there is a winner when $k = 3$ and also when $k = 5$.
In both situations estimate the probability of eight or more trials and make a plot of the estimated distribution of Y_3 and of Y_5.