Laboratory Session 02 : April 8, 2020

Exercises due on : April 22, 2020

Exercise 0

- practice with the discrete probability distributions in R: make plots of the standard pdfs and cdfs, playing with the function parameters

Exercise 1

- a set of measurements have been performed on the concentration of a contaminant in tap water. The following tables reports a set of values (x), with the corresponding probabilities given by the two methods (p_1 and p_2)

x	15.58	15.9	16	16.1	16.2
p_1	0.15	0.21	0.35	0.15	0.14
p_2	0.14	0.05	0.64	0.08	0.09

- Evaluate the expected values, $E[X]$, and the variance, $Var(X)$, for both methods

Exercise 2

- the waiting time, in minutes, at the doctor's is about 30 minutes, and the distribution follows an exponential pdf with rate 1/30

A) simulate the waiting time for 50 people at the doctor's office and plot the relative histogram

B) what is the probability that a person will wait for less than 10 minutes ?

C) evaluate the average waiting time from the simulated data and compare it with the expected value (calculated from theory and by manipulating the probability distributions using R)

B) what is the probability for waiting more than one hour before being received ?

Exercise 3

- let's suppose that on a book, on average, there is one typo error every three pages. If the number of errors follows a Poisson distribution, plot the pdf and cdf, and calculate the probability that there is at least one error on a specific page of the book

Exercise 4

- we randomly draw cards from a deck of 52 cards, with replacement, until one ace is drawn. Calculate the probability that at least 10 draws are needed.

Exercise 5

- The file available at the URL[1] `https://userswww.pd.infn.it/~agarfa/didattica/sindaciincarica.csv`) contains the list of all mayors currently in charge in the Italian mayors working in local towns in Italy. (Updated to April 6, 2020).
- open R and import the file in a `tibble` or `data.frame`

- (For the `tibble` option, import the `tidyverse` library and load the file with: `read_csv2("sindaciincarica.csv", skip=2)`, since the first two lines of the file have to be skipped, being only a comment to the file content)

- (For the `data.frame` option, use the `read.csv2()` function: `read.csv2("sindaciincarica.csv", skip=2)`)

- plot the gender distribution among the mayors (column name `sesso`)
- plot the number of towns grouped per province (`codice_provincia`) and per region (`codice_regione`)
- plot a distributions of the age (years only) of the mayors. In the `data_nascita` column the birthday is available
- plot a distribution of the time the mayor is in charge. The starting date is in column `data_elezione`. Since elections happen every 5 years, how many of them are going to complete their mandate this year ? And how many in 2021 ?

[1] This file belongs to the open-data file sets, for the Italian public administration, available at the URL `https://www.dati.gov.it/`