Statistical Methods Fall 2016

Assignment 2: Probability, Normality, CLT and Law of Large Numbers

Deadline: November 17, 23.59h

Topics of this assignment

The exercises below concern topics that were covered in Lectures 2, 3 and 4: probability, including the Law of Total Probability and Bayes' Theorem, random variables, the Central Limit Theorem (CLT), and the Law of Large Numbers (see the respective sections in Chapters 3, 4 and 5 of the book and the handouts of Lectures 2, 3, 4). Before making the assignment, study these topics.

How to make the exercises? See Assignment 1.

Theoretical exercises

Exercise 2.1 Exercise 1.8 from the set of additional exercises.
Hand in: Answers with your calculations and explanations.

Exercise 2.2 Consider the experiment of tossing a biased coin for which the probability of heads coming up equals 0.6.

a) Describe the probability distribution, i.e. the possible values and their probabilities, of the random variable 'number of heads in one coin toss'.

b) Describe the probability distribution, i.e. the possible values and their probabilities, of the random variable 'number of heads in two coin tosses'.

c) What is the expected number of heads in one coin toss?

d) Show analytically that the standard deviation of the variable 'number of heads in one coin toss' is approximately equal to 0.49.

e) Which distribution does the random variable 'the mean number of heads per coin toss after n tosses' have for large values of n (also give the expectation and standard deviation of the distribution)?

Hand in: Answers with your calculations and explanations.

R-exercises

Hints concerning R:

- Recall that a simple random sample of size n from a set of values `x` can be drawn in *R* using the function `sample(x,n)`. By default, the sample is drawn without replacement; by setting the additional parameter `replace` to `TRUE`, the sample is drawn with replacement. This function can be used to simulate a die.

- A sample from a certain distribution can be obtained in R with the function `rdist(n,par)` where `dist` stands for the name of the distribution, `n` for the sample size, and `par` for the relevant parameters: `x=rnorm(50,5,1)`, `x=rexp(25,1)`, `x=runif(30,-1,1)`, `x=rt(10,df=5)`, `x=rchisq(25,df=8)`. For example, the function `rnorm(n,mean,sd)` generates a sample of size `n` from the normal distribution with expectation `mean` and standard deviation `sd`. The parameters of the other distributions are documented in the help-function.

- A normal QQ plot can be obtained with `qqnorm(x)`.

- The command `dnorm(u)` computes the value of the probability density function of the standard normal distribution in `u`. For non-standard normal distributions adjust the arguments of the function.

- The command `lines(x,y)` joins the corresponding points in the vectors `x` and `y` with line segments. This is useful to draw a curve on top of an existing plot. Similarly, `abline(a,b)` draws the line $ax + b$ on top of an existing plot. Otherwise specify `type="l"` in the parameters of the function `plot()`.

- If you want to concatenate text and numbers (which could be useful for instance for titles of plots) you could use the R-function `paste()`.

Exercise 2.3

a) Generate the following samples and make for each of the four samples a normal QQ plot:

   (i) one sample of size 50 from the t-distribution with 2 degrees of freedom;

   (ii) one sample of size 30 from the $N(4, 1)$ distribution;

   (iii) one sample of size 25 from the exponential distribution with rate parameter 3;

   (iv) one sample of size 50 from the uniform distribution on the interval $[0, 1]$.

   What can you say about the shapes of these model distributions based on the QQ plots? Comment briefly on each plot.

   Hand in: Present the 4 plots concisely using the command `par(mfrow=c(2,2))`.

b) Answer for each of the data sets below the following question: "Is it reasonable to assume that the data come from a normal distribution?" In each case choose from the two answers: "Obviously not from a normal distribution" or "Normality cannot be excluded". Base your answer on histograms, boxplots and normal QQ-plots.

   (i) `klm.txt` (delivery time in days of products by Boeing to KLM)

   (ii) `iqdata2.txt` (IQ data)

   (iii) `dell.txt` (trading volumes Dell shares)

   (iv) `logdell.txt` (log trading volume Dell shares)

   (v) `simul102.txt` (simulated data)

   Hand in: Present for each data set: a suitable histogram, boxplot and QQ-plot, your answer to the question, and a short motivation of this answer. Use the function `par(mfrow=c(1,3))` to print the three plots next to each other. Adjust the size of the figure so that the ratio becomes approximately 1:3, and each plot is more or less square.

Exercise 2.4 Study the R-function `maxdice` from the file `function2.txt`. Load it by using the command `source("function2.txt")`.

a) Consider two dice and the random variable 'the maximum on two dice' (see Lecture 3). Illustrate the Law of Large Numbers for this random variable by considering 'the mean of the maximum on two dice' in n rolls for different values of n and making a plot similar to the one on slide 7 (page 16) of Lecture 4.

b) Use the function `maxdice` to find an approximate value of expectation of the random variable 'the maximum on m dice' for $m = 5, 10, 20$.

c) Investigate what the command `mean(maxdice(1000,2)==3)` does. Repeat this at most 6 times and briefly explain how the outcomes are related to the probability distribution of the random variable from part a).

Hand in: Properly described plot (part a), answer with motivation (parts b and c).

Exercise 2.5 Consider tossing the biased coin n times as taking a sample of size n from a population where all members have the probability distribution from Exercise 2.2, part a.

a) Write a function `cointoss(n,p)` that returns the outcome of n coin tosses, where each coin toss has probability of p being heads, independently of other coin tosses.

b) Use your function to illustrate the Central Limit Theorem for the random variable 'the mean number of heads per coin toss after n tosses' for the present context of the biased coin graphically, by making 4 plots similar to the 4 plots on slide 21 (page 42) of Lecture 4.

c) Explain briefly why the 4 plots of part b) illustrate the Central Limit Theorem in the present context.

Hand in: Plots (part b) and answers (part c). The code of the function should be placed in the appendix, and not in the main report.