SDA 2019 — Assignment 9

For Exercise 9.1 you can use the *R*-function `runif`, `rank`, `abs`, `sign`, `sum`, and for Exercise 9.2 the *R*-functions `wilcox.test` for the Mann–Whitney test and `ks.test` for the two sample Kolmogorov–Smirnov test.

Make a concise report of *all* your answers in *one single PDF file*, with only *relevant R* code *in an appendix*. It is important to make clear in your answers how you have solved the questions. Graphs should look neat (label the axes, give titles, use correct dimensions etc.). Multiple graphs can be put into one figure using the command `par(mfrow=c(k,l))`, see `help(par)`. Sometimes there might be additional information on what exactly has to be handed in. **Read the file AssignmentFormat.pdf on Canvas carefully**.

Exercise 9.1 — *Before making this exercise, you should read Section 6.2 in the syllabus.* In this exercise we investigate the behavior of two tests (sign and signed rank) for the location of a sample $X_1, \dots, X_n$ from the uniform distribution on $(-1+\theta, 1+\theta)$ with location parameter θ. The sign test for $H_0 : \theta = 0$ is based on the test statistic $S_n = \#(X_i > 0) = \sum_{i=1}^n 1_{\{X_i>0\}}$ and the (Wilcoxon) signed rank test is based on the test statistic $V_n = \sum_{i=1}^n R_i \,\mathrm{sgn}(X_i - 0)$. Here, R_i is the rank of $|X_i - 0|$ in the transformed sample $|X_1 - 0|, \dots, |X_n - 0|$.

We would like to find out whether or not the underlying distribution of the data is shifted to the right with respect to the uniform distribution on $(-1, 1)$, and we investigate this by testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$. Because under H_0 for large n, $S_n = \#(X_i \geq 0)$ is approximately normally distributed with expectation $n/2$ and variance $n/4$,[1] the first test follows the recipe

$$\text{“Reject } H_0 \text{ if } S_n > 1.644854\sqrt{n}/2 + n/2.\text{”}$$

Because under H_0 for large n, V_n is approximately normally distributed with expectation 0 and variance $\sqrt{n(n+1)(2n+1)/6}$,[2] the second test follows the recipe

$$\text{“Reject } H_0 \text{ if } V_n > 1.644854\sqrt{n(n+1)(2n+1)/6}.\text{”}$$

Here 1.644854 is the 95%-quantile of the standard normal distribution, so that both tests have, asymptotically for large n, a significance level of 0.05. The asymptotic efficiency of the sign test relative to the Wilcoxon test is given in Table 6.1 in the syllabus.

a. Implement the Wilcoxon test statistic V_n in `R` without using existing `R`-functions such as `wilcox.test(..)$statistic`. You may use the functions `rank`, `abs`, `sign`, `sum`.

b. Simulate 1000 times a sample of size $n = 200$ from the uniform distribution on $(-1 + \theta, 1 + \theta)$ with location parameter θ equal to 0, 0.05, 0.10 and 0.15, respectively. Perform for each sample the two tests based on S_n and V_n given above and report the results. (It is advised to present the resulting empirical rejection frequencies in a table.)

c. Which test performs best in your case? Is that in accordance with the theory? Motivate your answer.

d. Design and perform a simulation strategy to empirically verify the value of the asymptotic relative efficiency in Table 6.1.

[1] To see this, note that under H_0 the quantity $\#(X_i \geq 0)$ is binomially distributed with parameters n and $p = 0.5$. Hence for large n, this distribution can be approximated by a normal distribution with expectation $n/2$ and variance $n/4$.

[2] See page 73 in the syllabus.

e. Do your results of part d comply with the theory? Motivate your answer clearly and refer to the relevant part of the syllabus.

Hand in: your answers to all four parts b-e and the code for the implementation of the Wilcoxon test statistic in a.

Exercise 9.2 In 1882 S. Newcomb performed a series of experiments to determine the speed of light with the method of Foucault. Light bounced from a fast rotating mirror in Fort Meyer on the west bank of the Potomac in Washington to a fixed mirror at the foot of Washington monument, and back to the rotating mirror. The speed of the light was calculated from the measured distance between the mirrors (3721 meter) and the deflection angle of the emitted and received light on the rotating mirror. The file `newcomb.txt` contains Newcomb's data; the given values times 10^{-3} plus 24.8 are the original observations of the time (in $\sec^{-6}$) the light needed for traveling twice the 3721 meters.

a. Investigate whether there is a difference between the first 20 and the last 46 observations. Do this by making suitable graphical summaries, by determining an estimate of the difference, and by performing one or more suitable tests.
Explanation: 1. The graphical summaries should visualize the datasets and make them comparable. Possibly choose the same scales for the axes.
2. Choose a suitable estimator and a suitable test. Motivate your choice in one or two sentences.

b. Determine a 95% confidence interval for the traveling time of the light in Newcomb's experiment. According to present day physics the true value of the traveling time is equal to 24.8332 (in $\sec^{-6}$). Is this value in the computed confidence interval? What is your conclusion? *Explanation: choose a suitable location estimator. Motivate your choice in one or two sentences.*

Hand in: relevant plots and numbers and your answers to the questions.