Stochastic Methods for Material Science (Winter term 2023/24)

Programming Project – Part 1

(Deadline for submission: Februrary 4th 2024 at 11:59 pm)

Notice: You have to submit an R script including comments and the commands for solving the subsequent tasks. The evaluation and interpretation of the results and data analysis in R can be done via comments within the R script.

1. Task

In a determination of benzo[a]pyrene in a particular soil the following values were obtained (mg/kg dry matter):

i	1	2	3	4	5	6	7	8	9	10	11	12	13
x_i	9.89	10.06	9.88	10.09	10.15	10.34	9.96	10.1	9.97	10.27	9.78	10.19	10.02

a) Draw a histogram and describe your observations.

b) Draw a box plot and describe your observations.

c) Which distribution might the data follow? Check your assumption via a corresponding QQ plot.

d) The limit value for benzo[a]pyrene in Germany is 10 mg/kg.

   i. Test at the level $\alpha = 0.05$ whether the average content of benzo[a]pyrene in the soil at hand exceeds this limit value. To this end, also check the necessary assumptions for testing this hypothesis by applying appropriate other tests.

   ii. Additionally, we want to test at the level $\alpha = 0.05$ whether the official limit is exceeded in significantly more than 50% of all $n = 13$ cases! To this end, we transform the data to Bernoulli data by checking whether x_i is above or below the limit and then apply a *binomial test* for the success probability p. The null hypotehsis is, thus, $p = 1/2$. Read about the binomial test and apply it in R to the data. What is your answer to the above test question?

2. Task

We want to sample the Weibull distribution in R via inversion sampling.

a) Implement inversion sampling of the general Weibull distribution in R.

b) Now sample 25 points of the $\mathrm{Wei}(5, 2)$ distribution and estimate its mean.

c) Repeat this process for 1,000 times and store each time the estimates for the mean. Finally, plot the histogram of the estimates. Which distribution do they follow asymptotically?

3. Task

We want to simulate the uniform distribution $\mathrm{U}(\mathbb{D})$ on the unit disk in two dimensions

$$\mathbb{D} = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 \leq 1\}.$$

To this end, we apply rejection sampling.

a) Implement rejection sampling based on the uniform distribution $\mathrm{U}[-1, 1]^2$ as proposal distribution. The latter can be implemented easily by stochastically independent draws of $X \sim \mathrm{U}[-1, 1]$ and $Y \sim \mathrm{U}[-1, 1]$.

b) Generate $n = 500$ samples of the uniform distribution on $\mathbb{D}$ via your code and count each time the number of tries until one of the proposal draws is accepted. Then plot the generated 500 samples similar to the figure below

c) Finally, derive an estimate on the number π based on the stored number of tries of rejection sampling in b). Recall how the success probability of rejection sampling relates to the average number of tries. How large is the success/acceptance probability in this case?