Freie Universität Berlin
Prof. Dr. Max von Kleist
Liuwei Wang
Antonia Liebenberg

4. Assignment
Foundations of Mathematics and Statistics
WiSe 2025/26

Deadline: Nov 26th, 23:59 (the midnight **before** the lecture)

The homework should be worked out individually, or in groups of 3-4 students. Pen & paper exercises need not be handed in. The solutions will be discussed in the tutorial sessions. Programming exercises must be submitted via Whiteboard. The file containing the submission must include the last names of all group members in alphabetic order, e.g. "AlbertRamakrishnan-Romano", for group members Mandy Albert, Mike Ramakrishnan, and Marcus Romano.

Pen & Paper Exercise 1 (Statistical Inference)
Complete the following questions from the textbook (if you missed the lecture on 6th Nov).

a) Chapter 5, Section 5.8, Question 6

b) Chapter 6, Section 6.6, Question 1 and 3

Pen & Paper Exercise 2 (Empirical CDF)
Prove the following statements from Theorem 7.3,

$$\mathbb{E}\left(\widehat{F}_n(x)\right) = F(x),$$

$$\mathbb{V}\left(\widehat{F}_n(x)\right) = \frac{F(x)(1 - F(x))}{n},$$

$$\text{MSE} = \frac{F(x)(1 - F(x))}{n} \to 0,$$

Pen & Paper Exercise 3 (Bootstrapping)
We collected the following dataset with 4 samples: 3 8 1 8, from which we obtained 5 Bootstrap samples (each column is one bootstrap trial):

8	8	1	8	8
1	1	3	1	8
3	1	3	8	1
8	1	3	1	8

Compute a bootstrap 60% percentile confidence interval and a normal 95% confidence interval for the median.

Programming Exercise 1 (Bootstrapping, 2+2+1 points)
(to be uploaded via Whiteboard)

a.) The input file "DeathsByState_2019_2020.xlsx" on Whiteboard contains the number of deaths on each day of 2019 and 2020 in each of the federal states of Germany. Read in the data using the pandas library.

b.) Reproduce Figure 7.1 from the textbook (without the confidence interval) using the above dataset by plotting the empirical CDF of daily deaths in Germany (i.e. summing up all states) for 2019 and 2020, respectively (meaning two empirical CDF curves on the same plot).

c.) Perform bootstrapping to compute the 95% percentile confidence interval for the median daily deaths of Germany in 2019 and 2020, respectively. You should

- Set your random seed using `np.random.default_rng(seed = 123)` and then use `rng.choice` to perform resampling on the data.

- Use the following bootstrap sample sizes B: 100, 500, 1000.

- Save the bootstrap sample sizes and the respective confidence intervals in "output.txt". Each row of the file have the first number being the bootstrap sample size and the second and third being the bounds of the confidence intervals for each year (CI for 2019 is (a1, b1), 2020 (a2, b2):

$$100, a1, b1, a2, b2$$
$$500, a1, b1, a2, b2$$
$$1000, a1, b1, a2, b2$$

d) Interpret your results above in 30 words.

Your code should print the requested plot (e.g. using plt.show()), save the file "output.txt", and print the interpretation using python's print function. Call this program "Ex4Bootstrap.py" and submit it via the Whiteboard system.

Good luck!