Optimization and Data Science
8. Homework exercises

Theoretical exercise 1:
Let X be a random variable uniformly distributed on $[a, b]$ with $a, b \in \mathbb{R}$ and $a < b$. Prove that:

a) $\mathbb{E}(X) = \frac{1}{2}(b + a)$

b) $\mathbb{V}(X) = \frac{1}{12}(b - a)^2$

Hint: $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ (slide 9 lecture 15) and $b^3 - a^3 = (b - a)(a^2 + ab + b^2)$.

Programming exercise 1:
Consider the values in the file

a) "data1.txt"

b) "data2.txt"

as realizations of a random sample, i. e., as realizations of random variables with the same probability distribution.

- *Estimate the corresponding expected value and the corresponding variance.*
- *Plot the probability density function of a normal distribution with your estimated expected value and variance.*
- *Plot a histogram for the random sample.*
- *Do you think the values are realizations of your plotted normal distribution?*

(You can use some external function for the calculation of the expected value and the variance.)

Hint: scipy.stats could be used for this purpose.

Programming exercise 2:
Let X be a random variable with

a) $X \sim \mathcal{U}(0, 4)$ *(uniform distribution with $a = 0$ and $b = 4$)*

b) $X \sim \mathcal{N}(0, 4)$ *(normal distribution with $\eta = 0$ and $\sigma^2 = 4$)*

c) $X \sim \mathcal{LN}(0, 4)$ *(log-normal distribution with $\eta = 0$ and $\sigma^2 = 4$).*

Create a plot (for each distribution), which shows:

- *the probability density function (in some appropriate range)*

- *the expected value* $\mathbb{E}(X)$
- *the interval* $[\mathbb{E}(X) - \sqrt{\mathbb{V}(X)}, \mathbb{E}(X) + \sqrt{\mathbb{V}(X)}]$
- *the 0.95-confidence interval (centered at the expectation)*

(You can use some external function for the calculation of the expected value, the variance and the confidence interval.)

Hint: scipy.stats could be used for this purpose.

Programming exercise 3:
Let X *be a random sample with random variables* X_i *uniformly distributed in* $[0,1]^2$ *for all* $i \in \{1, \ldots, n\}$. *The function*

$$f(X) = \frac{4}{n}\ |\{i \in \{1, \ldots, n\}\ :\ \|X_i\|_2 \leq 1\}$$

is an estimator for the mathematical constant Π. *(The estimate is four times the number of values with euclidean norm lower or equal to one divided by the total number of values.)*

Calculate the estimate for different values of n *and plot the difference between the estimate and* Π *depending on* n. *Do you think that this estimator is biased?*

The solutions of the theoretical exercises will be discussed on 08. June 2020.