Technische Universität München   SS 2022
Zentrum Mathematik
Lehrstuhl für Mathematische Statistik
Prof. Dr. Claudia Czado (cczado@ma.tum.de)
Ariane Hanebeck (ariane.hanebeck@tum.de)
Henrik Sloot (henrik.sloot@tum.de)

Exercise Sheet 1
Computational Statistics [MA4402]

If you are new to **R** and its functions, note that you can always type `?function` to get help for a specific `function` you are not familiar with (e.g. try `?runif`).

The video on sampling in **R** is also useful for this sheet (https://www.youtube.com/watch?v=SsPBX65otzk).

During the exercise sessions, you work on the exercises marked with *. You can take a look at the additional exercises at home. Their solutions will be uploaded and they are also relevant to the exam.

This sheet is discussed in the weeks of May 2nd and May 9th.

Problem 1 (*)   Let $\hat{\theta}$ be an estimator of θ. Prove that

$$bias(\hat{\theta})^2 + Var(\hat{\theta}) = MSE(\hat{\theta})$$

Problem 2 (*)   Consider $X_1, \ldots, X_n \sim N(0,1)$ i.i.d. and the estimator $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$ for the mean. We set $n = 100$.

   i) Calculate bias and mean squared error of $\hat{\mu}$.

   ii) Determine bias and mean squared error of $\hat{\mu}$ through simulation. You may use the **R**-functions `rnorm`, `var`, and `mean`.

   iii) How will bias and mean squared error change if n is increased?

Problem 3 (*)   Write an **R**-function, which generates random values from a $Bin(n,p)$-distribution. You may use the **R**-function `runif`, which generates random values from a $Uniform(0,1)$-distribution. Verify your function by drawing 100,000 observations from $Bin(10, 0.3)$ and plotting the observed relative frequencies (using the functions `table` and `barplot`). Compare to the true probabilities (`dbinom`) by plotting them in the same figure.

Problem 4 (Additional) Let $X \sim Bin(100, 0.7)$ and $Y \sim N(0, 1)$ be independent.

i) Determine $P((X + 1) \cdot Y^3 > 0)$ analytically and through simulation. For the simulation, make use of the **R**-function you implemented for Problem 3 and of the **R**-function `rnorm`. Set the size n of your data set to $n = 10,000$.

ii) Determine $P((X+1) \cdot Y^3 > 0.3)$ through simulation and write down your estimator. Here again, set the size n of your data set to $n = 10,000$.

iii) Is the estimator that you use in ii) for $P((X+1) \cdot Y^3 > 0.3)$ biased? Provide an analytical argument.