*Institut für Mathematik*
*Abteilung Numerik*
*Jana de Wiljes*

# 2. Problem sheet for **Statistical Data Analysis**

**Exercise 1** (4 Points)
Consider a random variable $T$ with probability density

$$f(t) = \begin{cases} \frac{1}{2\theta\sqrt{t}} \exp(\frac{-\sqrt{t}}{\theta}), & \text{for} \quad t > 0 \\ 0, & \text{for} \quad t \leq 0 \end{cases} .$$

where $\theta$ is an unknown model parameter. Moreover, we have the following i.i.d. samples

| $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ |
|---|---|---|---|---|
| 11300 | 5000 | 4300 | 8500 | 7900 |

1. Derive the log likelihood function and the maximum likelihood estimate (in general and for the specific sample).
2. Use the method of moments in order to find an estimate for $\theta$ (Hint: $E(T) = 2\theta^2$).

**Exercise 2** (4 Points)
Consider a Poisson distributed random variable with probability density function

$$f_\theta(x) = \frac{\theta^x}{x!} e^{-\theta}, \quad x = 0, 1, 2, \ldots$$

and independent samples $(x_1, \ldots, x_n)$. Derive the maximum likelihood estimate $\hat{\theta}$ of $\theta$. Is the estimate unbiased and consistents

**Exercise 3** (8 Points)
The file `wine.txt` contains data on the wine production for a certain wine region in tons per $100m^2$ and the average number of berries in a bunch of grapes for the years 1971 to 1983 (no data were taken in 1972 due to a storm). Load the data in python and

1. Produce a scatterplot of the data.
2. Assume the a simple linear regression model and estimate the parameters $\beta_0$ and $\beta_1$ in the regression model (use your own derivation.
3. Plot the regression line.
4. Predict the yearly production of wine if the number of berries in a bunch of grapes is 100.