Statistical Data Analysis
Problem Sheet 2

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

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

   where θ is an unknown model parameter. Moreover, we have the following i.i.d. samples:

$$\frac{t_1}{11300}, \quad \frac{t_2}{5000}, \quad \frac{t_3}{4300}, \quad \frac{t_4}{8500}, \quad \frac{t_5}{7900} \tag{1}$$

   1. Derive the log likelihood function and the maximum likelihood estimate (in general and for the specific sample).

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

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

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

3. **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

   (a) Produce a scatterplot of the data.
   (b) Assume the a simple linear regression model and estimate the parameters β_0 and β_1 in the regression model (use your own derivation).
   (c) Plot the regression line.
   (d) Predict the yearly production of wine if the number of berries in a bunch of grapes is 100.

Note that you need to register via the PULS system until the 10th of november 2023. After that date, registration is no longer possible..