Statistics and Numerics

Lecture Prof. Dr. Jens Timmer

Exercises Helge Hass, Mirjam Fehling-Kaschek

Exercise Sheet Nr. 4

Exercise 1: Parameter estimation

Consider the linear regression

$$y_i = ax_i + \varepsilon_i, \quad \varepsilon_i \sim N(0,1) \; . \tag{1}$$

The maximum likelihood estimator $\hat{a}$ for the parameter a given N data tuples (x_i, y_i) is

$$\hat{a} = \frac{\sum_{i=1}^{N} y_i x_i}{\sum_{i=1}^{N} x_i^2} \; . \tag{2}$$

a) Simulate data points for $a = 1$ in the interval $x \in [0,20]$: generate $N = 10$ equidistant x values in the given range and simulate the respective y values by adding ε. Compute $\hat{a}$. Plot the data and add a line to the plot using the estimated $\hat{a}$.

b) Estimate parameter a for $N = 2,5,10,50,100,500,1000$ data points: calculate the mean and variance of $\hat{a}$, using $M = 1000$ realizations. Plot the variance and mean of $\hat{a}$ as a function of N.

c) For given N, does the estimator resemble a normal distribution?
Hint: Compute the histogram, cumulative density function and Q-Q-Plot.

Exercise 2: Uniform noise distribution

Replace the noise term ε_i in Eq. (1) by

$$\varepsilon_i \sim U(-b,b) \tag{3}$$

with probability density $f(x) = \begin{cases} 1/(2b) & -b \leq x \leq b \\ 0 & x \notin [-b,b] \end{cases}$ and repeat excercise 1.

- Choose the boundary b of the uniform distribution such that the variance equals one: $var(\varepsilon) = 1$.

- Use the same estimator Eq. (2). Does it represent the maximum likelihood estimator?

Exercise 3: Log-normally distributed noise

Repeat the exercise with multiplicative log-normally distributed noise by replacing Eq. (1) by

$$y_i = \exp(\ln(ax_i) + \sigma\varepsilon_i) \tag{4}$$

with $\varepsilon \sim N(0,1)$.

- Why does Eq. (4) represent a model with a multiplicative error? When does the log-normally distributed noise approximate normally distributed noise? *Hint:* Make use of the Taylor series.

- Repeat the exercise for $\sigma = 0.1, 0.5, 1$. How are the y values affected by the choice of σ.

- Use the same estimator Eq. (2). Does it represent the maximum likelihood estimator?