Excursus: Likelihood function and maximum likelihood (ML) method

● Point estimation in inductive statistics by maximum likelihood

given: sample (x_1, x_2, …, x_n) of n identically and independently distributed (i.i.d) sample variables X_1, X_2, …, X_n from a population with an unknown parameter θ

Known type of distribution of the random variable X in the population (e.g. normal distribution)

Density function: $f(x|\theta)$ for all x

As the sample variables X_i are distributed like X and independently drawn from the population, the **joint density function** $f_{X1,\dots,Xn}$ of the n sample variables is given by the product of the **marginal density functions** f:

(A5.1) $$f_{X_1,\dots,X_n}(x_1,\dots,x_n|\theta) = f(x_1|\theta)\dots f(x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

● **Idea of the maximum likelihood method**:
Which value of the parameter θ in the population will maximise the probabilty of the given sample values to be realised?

Before drawing the sample:
Parameter θ is a fixed quantity, wheras the potential sample values $X_1.X_2,\ldots.X_n$ are variable

After the drawing of the sample:
Sample values $x_1,x_2,\ldots,x_n$ have been realised (fixed), so that $f_{X1,\ldots,Xn}$ can be viewed as a function of the unknown parameter θ

In order to stress the situation after drawing the sample, one replaces the **joint density function** $f_{X1,\ldots,Xn}$ in (A5.1) by the **likelihood function** L:

(A5.2) $$L(\theta|x_1,\ldots,x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

The likelihood function of an unknown parameter θ renders the joint densitiy for each value of θ for a given sample $(x_1,x_2,\ldots,x_n)$.

● Maximising of the likelihood function

Maximising the likelihood function L with respect to θ is based on differential calculus. Because sums can be easier differentiated than products, usually the **log-likelihood function** (logarithm of the likelihood function) is considered:

(A5.3) $$\ln L(\theta|x_1,\ldots,x_n) = \sum_{i=1}^{n} \ln f(x_i|\theta)$$

As the logarithm of a function is a monotonic transformation, the value of θ that maximises ln L as well maximises L. The necessary condition for a maximum of ln L reads

(A5.4) $$\frac{d \ln L(\theta|x_1,\ldots,x_n)}{d\theta} \overset{!}{=} 0$$

The **maximum likelihood estimator** (ML estimator) $\hat{\theta}_{ML}$ for the unknown parameter θ is obtained by solving the equation (A5.4).