142.351, 260014: Statistical Methods of Data Analysis

W. Waltenberger[1] (Lecture)
L. Einfalt[1,2] (Tutorials TU Wien and Universität Wien)
F. Reindl[1,2] (Tutorials TU Wien and Universität Wien)

[1] Institut für Hochenergiephysik der Österreichischen Akademie der Wissenschaften, A-1050 Wien, Nikolsdorfer Gasse 18
[2] Technische Universität Wien, Atominstitut, A-1020 Wien, Stadionallee 2

Winter term 2023/2024

Exercise sheet 4

Hand-in by: Friday, 17. November 2023 - 11:00
Class: Friday, 17. Nobember 2023 - 14:15

Example 4.1

For a series of measurements $x_1, \ldots, x_n$ of size $n = 250$ drawn from an exponential distribution $\mathrm{Ex}(\tau)$, the following holds true: $s = \sum x_i = 395.24$

a) Calculate the ML estimator $\hat{\tau}$ of the mean τ of the exponential distribution.

b) Show that the ML estimator $\hat{\tau}$ is efficient.

Example 4.2

The Laplace distribution $\mathrm{La}(m, s)$ with position parameter m and scale parameter s has the density function

$$f(x; m, s) = \frac{1}{2s} \exp\left(-\frac{|x - m|}{s}\right).$$

a) Calculate expectation value and variance of this distribution.

b) Determine the ML estimators of m and s.

Example 4.3

A survey of 1000 people yields the following preferences for parties A, B, C, D:

$$n_A = 253, n_B = 290, n_C = 349, n_D = 108$$

Estimate the voter shares p_A, p_B, p_C, p_D of the four parties in the population, assuming that the sample is representative. Find approximate expressions for the variance, joint covariance matrix and correlation matrix of the estimates.

Hint: Use the multinomial distribution as a generalization of the binomial distribution. You may use literature to find expressions for the variance and covariance of the binomial distribution, if you find these helpful in your calculations.

Example 4.4

Compulsory for TU (142.351), optional for University (260014)!
An unknown quantity μ is measured n times independently. For each measurement the accuracy is different but known, and without systematic error. Each measurement x_i comes from a distribution with mean μ and variance $\sigma_i^2, i = 1, \ldots, n$. The variable μ is estimated by a *weighted mean* of the form

$$\hat{\mu} = \sum_{i=1}^{n} w_i x_i$$

.

a) Determine the weights w_i such that the estimator $\hat{\mu}$ is unbiased and has the smallest possible variance among all estimators of this form.

b) Show that this estimator is identical to the ML estimator if the observations are normally distributed.

Example 4.5 (Prog)

Simulate N=150 samples of size $n = 250$ from the combination of two normal distributions (Hint: Exercise 3.2) with density

$$f(x) = p \cdot \varphi(x|\mu, \sigma_1^2) + (1 - p) \cdot \varphi(x|\mu, \sigma_2^2),\ \ \mu = 0, p = 0.7, \sigma_1^2 = 1, \sigma_2^2 = 10.$$

For each sample, calculate the ML estimators of μ and p by numerically maximizing the log-likelihood function and analyze the distributions of the estimates. Plot the log-likelihood function of one sample.

Example 4.6 (Prog)

Consider again the situation of exercise 4.5., but this time solve the problem with the EM-algorithm (see lecture slides from 326 and the notebook skeleton). Simulate N=150 samples of size $n = 250$ from the combination of two normal distributions described above.

For each sample, calculate the ML estimators of μ and p by numerically maximizing the log-likelihood function using the EM algorithm. To do this, follow the steps described on slide 327 and *do not use* a ready-made library implementation. Compare your results with the results of exercise 4.5, in terms of running time and precision of results.