Data Analysis - 2021

Exercise sheet no 6: 7. December 2021
The maximum likelihood method

Exercise 1: Maximum Likelihood Estimator (4 Points)

Consider a random variable x, distributed according to the probability density function

$$P(x|\alpha) = \frac{1}{2}(1 + \alpha x) \quad \text{for } -1 \leq x \leq 1 \text{ and } 0 \leq \alpha \leq 1.$$

$n = 20$ measurements of this random variable gave the sample found in `MLE.txt` on the webpage.

(a) Calculate the negative-log-likelihood function $\text{NLL}(\alpha) = -\ln L(\alpha)$ for these data as a function of α. Plot it for $0 \leq \alpha \leq 1$.

(b) Determine the maximum likelihood estimator $\hat{\alpha}$ using your plot of the negative-log-likelihood.

Exercise 2: Maximum Likelihood and least squares (14 Points)

In the file `exponential_data.txt` there are 1000 decay time measurements between 0 and 5 µs of muons, which have a mean lifetime of $\tau = 2$ µs. The PDF used to generate this data is:

$$P(t) = \frac{1}{\tau \cdot (1 - e^{-5/\tau})} \cdot e^{-t/\tau}$$

where τ is the mean decay time in μs and t is the measured decay time.

(a) Calculate twice the negative-log-likelihood ($2{\cdot}\text{NLL}(\tau)$) as a function of the lifetime τ. Shift the values of the $2{\cdot}\text{NLL}(\tau)$ such that the minimum is zero and plot it for $1.8 < \tau < 2.2$ µs.

(b) Bin the time measurements into a histogram of 40 equally sized bins, from 0 to 5 µs. Use the binned maximum likelihood method to calculate twice the negative-log-likelihood ($2{\cdot}\text{NLL}(\tau)$) for these 40 bins as a function of τ. Shift the values of the $2{\cdot}$NLL such that the minimum is zero and overlay the plot with the unbinned calculation in part a). Does it matter if you approximate the integral of the function when computing the prediction for each bin (f_i)?

(c) Now calculate the χ^2 function as a function of τ and overlay it on the NLL plot with the other two calculations in parts a) and b). Shift the χ^2 also such that the minimum is at zero. Compare the results.

(d) Bin the dataset into a wider histogram of 2 bins and calculate the binned $2{\cdot}$NLL and the χ^2 function again. How is the agreement now between them? How do they agree agree with the unbinned calculation in part a)?

Exercise 3: Fitting with more than one parameter (polynomial) (12 Points)

In the file `polynomial_data.dat` you can find 30000 measurements of the variable x between the range -1 and 1.

(a) Bin these measurements as a histogram with 20 bins and calculate the uncertainty on number of measurements in each bin using the Poisson distribution. Plot this histogram with error bars.

(b) Use the least-squares method `scipy.optimize.curve_fit` to fit a polynomial of first, second, third and fourth degree to the histogram bins, taking the x-value to be centre of each histogram bin and the y-value to be the number of measurements. Plot the resulting polynomial fits overlaid on the histogram.

(c) For each fit, calculate the uncertainties on the parameters using the covariance matrix returned by the `scipy.optimize.curve_fit` method.

(d) Use the optimal fitting parameters to calculate the χ^2/ndf for the different fits and plot it as a function of the degrees of the polynomial.

(e) What do you think was the degree of the polynomial used to generate the data?

Deadline for submission: Thursday, 23. December 2021 17:00
Form: Please submit your solutions to da@physik.uzh.ch. The solutions should be submitted as a single python script with answers to specific questions in the comments.