Homework #3

Due Feb 14, 2022, 11:59pm

Problem 3.1.

Consider the data in the files `a100.csv, b100.csv, c100.csv` and `d100.csv`.

a. Determine the underlying probability distributions (and its parameters) of each data set, by creating a histogram and over-plotting with the most similar probability distribution, until the agreement is acceptable. Create a label with the name of the distribution, and its parameter values on the plot. Do not use a fitting function but determine the parameters by changing them manually until there is a good visual match. The goal of this exercise is to develop an intuition on how the shapes of the different distributions change as a function of the parameters.

b. Create a new series from each data set through the formula

$$y_p = \sum_{i=0}^{K-1} x_{p+i}$$

i.e. each new number is the sum of K adjacent elements of the original series (so called moving average). Determine the probability distribution and its parameter for each sequence for K=5, 20 and 80. Calculate the mean and variance of the original distributions and compare to the derived (summed) series.

Problem 3.2.

The files `noise01.csv` to `noise10.csv` contain a random noise from a real instrument, measuring the intensity of light as a function of the voltage on a light source. The voltage goes from 0.1V to 1.0V, encoded in the filename. (0.1V, 0.2V, 0.3V, 0.4V, 0.5V, 1.0V). Prove that the noise is due to the Poisson distribution of the discrete photons using iPython. Hint: Use the fact that a Poisson distribution has a single parameter, which determines both its mean and variance. Show that these quantities satisfy the appropriate scaling law for each data set.