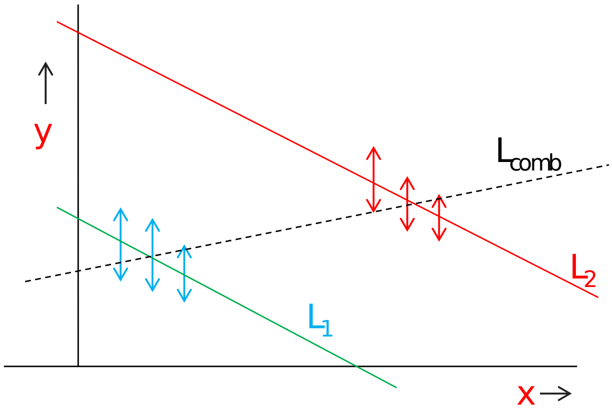
# Goodness-of-fit tests: exercises

## 1 Gaussian data

We come back to the straight line fitting example used in the context of data combination.




We consider a tracking detector without magnetic field (particle trajectories are straight lines $y = a + bx$). The detector is made of 2 tracking stations, each with 3 layers, situated at $x_{11} = 10$, $x_{12} = 11$, $x_{13} = 12$; $x_{21} = 20$, $x_{22} = 21$, $x_{23} = 22$ (distances in mm). We want to measure $a$, the impact parameter of the particle with respect to the origin, while we are not interested in $b$ (it is a "nuisance parameter"). Each layer has a resolution of 1 mm: in other words, each measurement $y_i$ is a random variable with a Gaussian distribution of mean $a + bx_i$ and variance 1.

In the following we assume the true values of $a$ and $b$ to be $a_0 = 3\,\text{mm}$ and $b_0 = 0.1$. The measured values are $y_{11} = 4.0$, $y_{12} = 3.8$, $y_{13} = 3.6$, $y_{21} = 5.2$, $y_{22} = 4.9$, $y_{23} = 4.8\,\text{mm}$.

1. What is the Gaussian $\chi^2$ corresponding to the measured values? What is the corresponding p-value? Same question replacing $a, b$ by $\hat{a}, \hat{b}$ in the model used in the $\chi^2$ computation. What number of degrees of freedom needs to be used in each case?

2. **Toy study**. In this question we want to generate pseudo-data from the true values of $a$ and $b$, and check the distribution (histogram) of the $\chi^2$ and the corresponding p-value, before and after fitting. We will generate $N$ pseudo-experiments ($N = 1000$) in the following way:

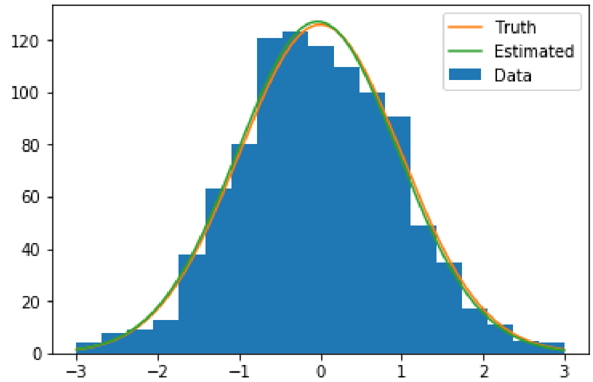
   (a) Generate values for the $y_i$ with a Gaussian distribution of mean $a_0 + b_0 x_i$ and variance 1.

   (b) Store the value of the "pre-fit" $\chi^2$ (with $a_0$ and $b_0$), as well as the corresponding p-value.

   (c) Compute the values of $\hat{a}$ and $\hat{b}$ (this can be done analytically).

   (d) Store the value of the "post-fit" $\chi^2$ (with $\hat{a}$ and $\hat{b}$), as well as the corresponding p-value.

## 2 Poisson data




We want to compare the performance of several goodness-of-fit tests: unbinned $\chi^2$, and several binned $\chi^2$ (Neyman, Pearson, and Baker–Cousins). For this, we will check the distribution of the corresponding test statistics and p-value, using toy experiments. We use a simple model consisting of a Gaussian with unknown mean and sigma, which we will simultaneously estimate using data. We assume the true values of the parameters to be $\mu = 0$ and $\sigma = 1$.

1. Write python functions for each of the g.o.f. tests.

2. Check and compare the properties of all four g.o.f. tests, and the corresponding p-values, in the following cases:

   (a) 20 bins between -3 and 3, 1000 events / toy experiment

   (b) 20 bins between -5 and 5, 100 events / toy experiment

   (c) 10 bins betwene -2 and 2, 20 events / toy experiment

Careful: make sure to be generating Poisson data, not multinomial data...