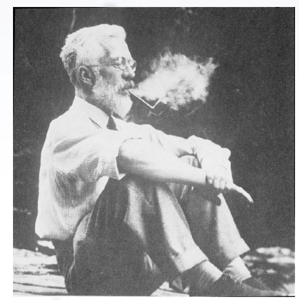
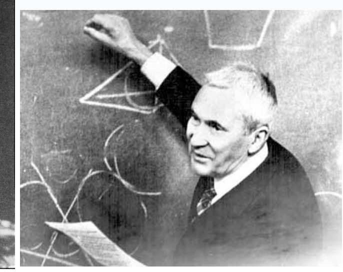
Applied Statistics

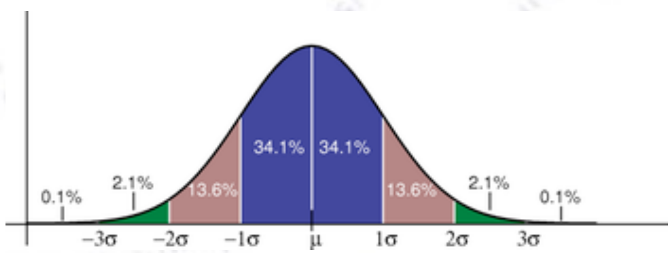
Two comments on ChiSquare





Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense"

Defining the Chi-Square - for a fit

<u>Problem Statement:</u> Given N data points (x,y), adjust the parameter(s) θ of a model, such that it fits data best.

The best way to do this, given uncertainties σ_i on y_i is by minimising:

$$\chi^2(\theta) = \sum_i^N \frac{(y_i - f(x_i, \theta))^2}{\sigma_i^2}$$

The power of this method is hard to overstate!

Not only does it provide a simple, elegant and unique way of fitting data, but more importantly it provides a **goodness-of-fit measure**.

This is the Chi-Square test!

Chi-Square for two histograms

The Chi-Square is generally defined as:

$$\chi^2 = \sum_{i \in bins} \frac{\text{Difference}^2}{\text{Error on Difference}^2}$$

Comparing two histograms,
the difference is simple to define:

$$\text{Difference} = O1_i - O2_i$$

The uncertainty on the count in a bin
is (assumed) Poisson distributed:

$$\text{Error on } O1_i = \sigma(O1_i) = \sqrt{O1_i}$$

Now we can calculate the denominator:

$$\text{Error on Difference} = \sqrt{\sigma(O1_i)^2 + \sigma(O2_i)^2} = \sqrt{O1_i + O2_i}$$

Inserting all of this yields the final result:

$$\chi^2 = \sum_{i \in bins} \frac{(O1_i - O2_i)^2}{O1_i + O2_i}$$