χ^2 and maximum likelihood fits

Alexandre Brea, Frédéric Blanc

04/04/2025

What is a fit?

In physics, very often we want to adjust (or fit) a model to our experimental data

E.g: perform a fit using the linear model

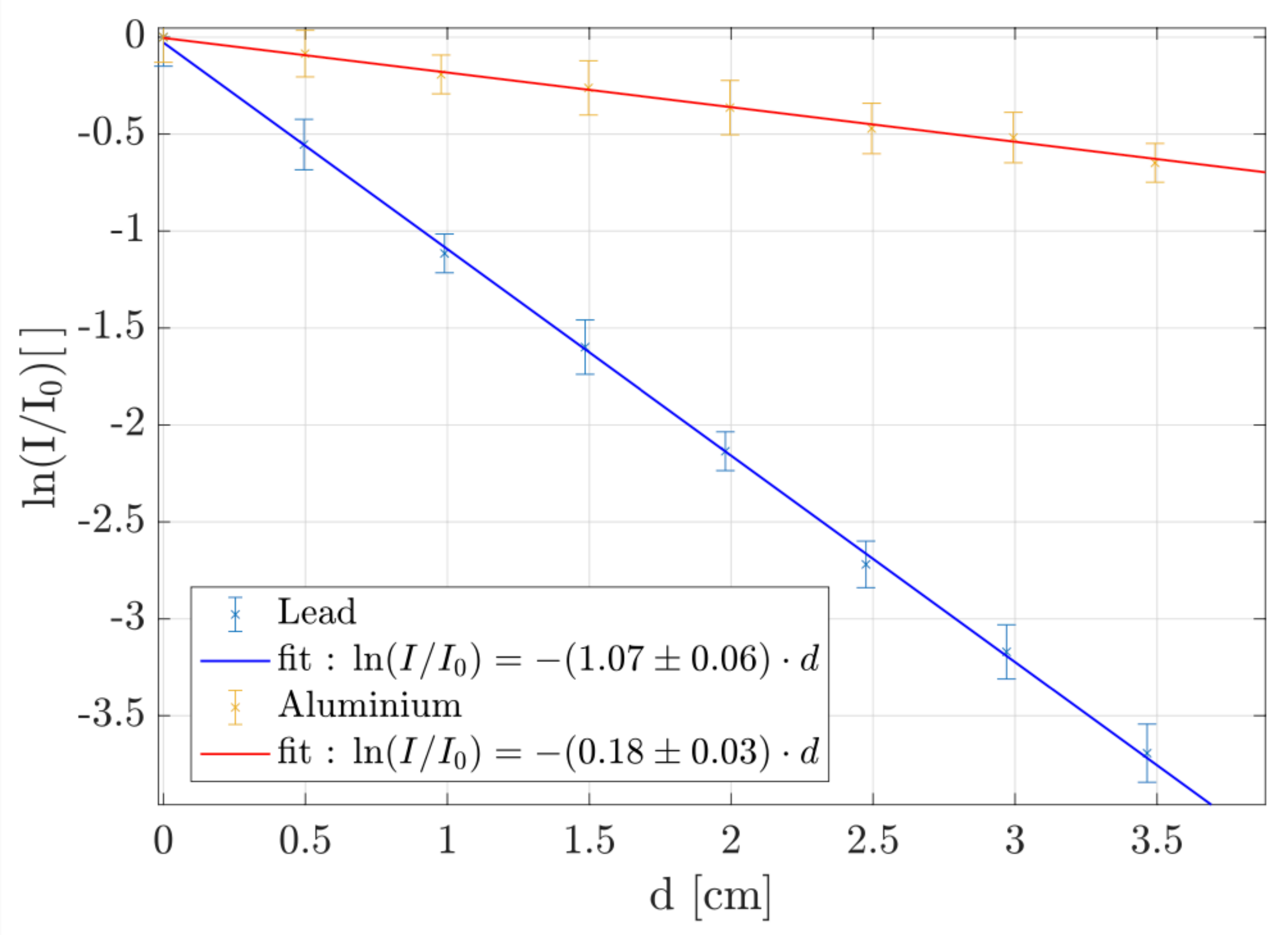
$$f(x;\vec{\theta}) = mx + h, \text{ where } \vec{\theta} = (m,h)$$

to a set of N measurements $\{y_i(x_i) \pm \sigma_i, i = 1,\ldots,N\}$

The fitting function depends on N_p parameters encoded in $\vec{\theta}$

We want to find the best fit function by adjusting these parameters

From TP3 lab:



Fitting procedure:

1) Determine the best estimators $\hat{m}$, $\hat{h}$ for the parameters m,h

2) Compute the uncertainties on these estimators

3) Provide a measure of how good the fit is

χ^2 fit

- Widely used statistical method for parameter determination
- We first write the χ^2 function, which is the sum of all "normalised distances" squared:

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(y_i - f(x_i; \vec{\theta})\right)^2}{\sigma_i^2}$$

- The best estimators $\hat{\theta}_j$ are obtained by <u>minimising</u> the χ^2:

$$\left.\frac{\partial \chi^2}{\partial \theta_j}\right|_{\theta_j = \hat{\theta}_j} = 0, \quad \text{for all parameters in } \vec{\theta} \; (\theta_1 = m, \; \theta_2 = h \text{ for the linear model})$$

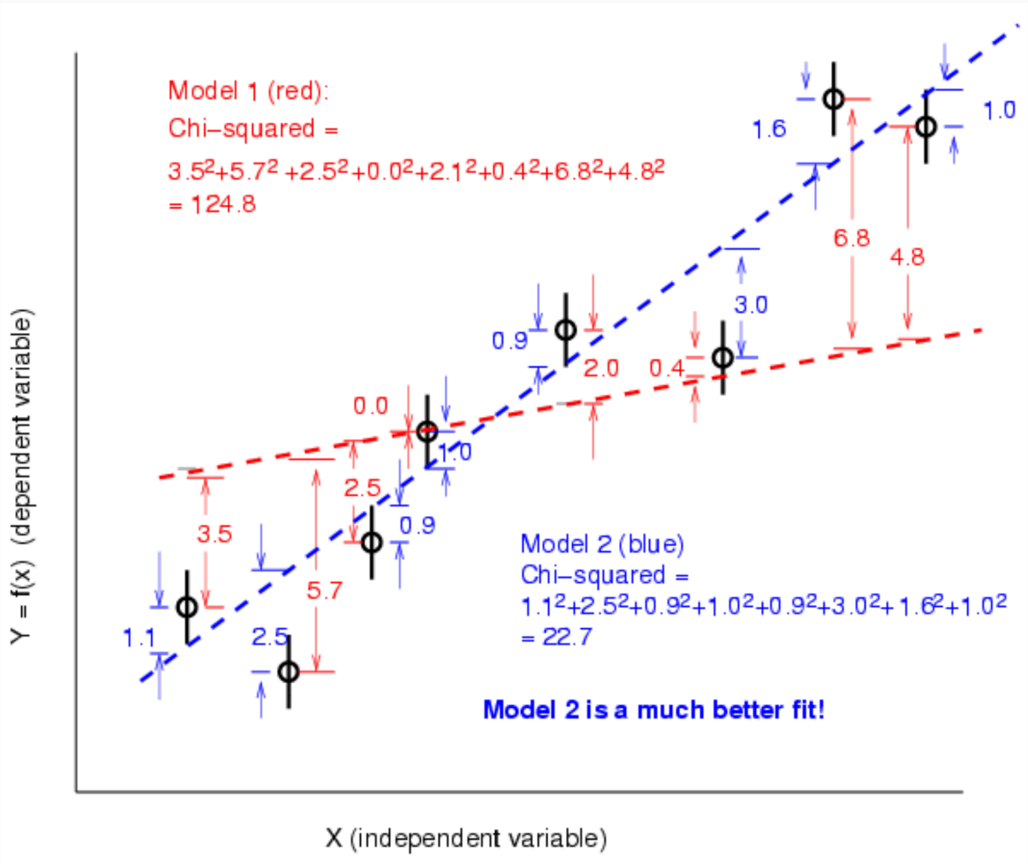


Figure taken from https://www.phys.hawaii.edu/~varner/PHYS305-Spr12/DataFitting.html

Maximum likelihood estimation

- The χ^2 fit is a special case of the Maximum Likelihood (ML) fit
- The likelihood function is defined as $\mathscr{L}(\vec{\theta}) = \mathscr{P}(\vec{y};\vec{\theta})$, where $\mathscr{P}(\vec{y},\vec{\theta})$ is the probability of observation $\vec{y}$ [depends only on the data measured]
- If the y_i measurements are independent, we can write:

$$\mathscr{L}(\vec{\theta}) = \prod_{i=1}^{N} \mathscr{P}(y_i(x_i), \sigma_i; \vec{\theta})$$

- The best estimators $\hat{\theta}_j$ are obtained by maximising $\mathscr{L}(\vec{\theta})$ [normally we minimise $-2\ln\mathscr{L}(\vec{\theta})$ because $\ln(\prod_i x_i) = \sum_i \ln x_i$ is easier to perform computation wise]

χ^2 fit: a particular case of a ML fit

- In the particular case when the y_i are Gaussian distributed random variables, the likelihood function becomes:

$$\mathscr{L}(\vec{\theta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(y_i - f(x_i, \vec{\theta}))^2}{2\sigma_i^2} \right)$$

- This implies that:

$$-2\ln\mathcal{L} = -2\sum_{i=1}^{N} \ln\mathcal{P}(y_i(x_i), \sigma_i; \vec{\theta}) = -2\sum_{i=1}^{N}\left[-\frac{(y_i - f(x_i, \vec{\theta}))^2}{2\sigma_i^2} \underbrace{- \ln\sigma_i - \ln\sqrt{2\pi}}_{=\text{constant}} \right]$$

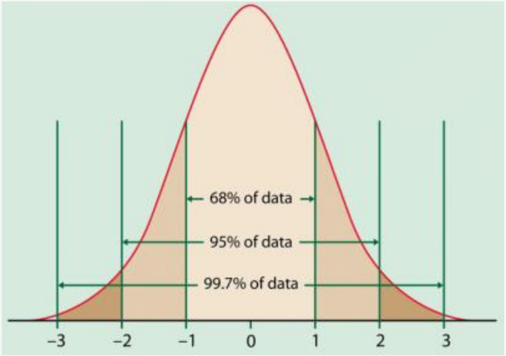
$$= \sum_{i=1}^{N} \frac{(y_i - f(x_i, \vec{\theta}))^2}{\sigma_i^2} \underbrace{+2\ln\sigma_i + \ln(2\pi)}_{=\text{constant}} = \chi^2 + \text{constant}\,.$$

Therefore, the χ^2 minimisation method is a particular case of the more general ML method

$$\mathscr{L} \propto e^{-\chi^2/2}$$

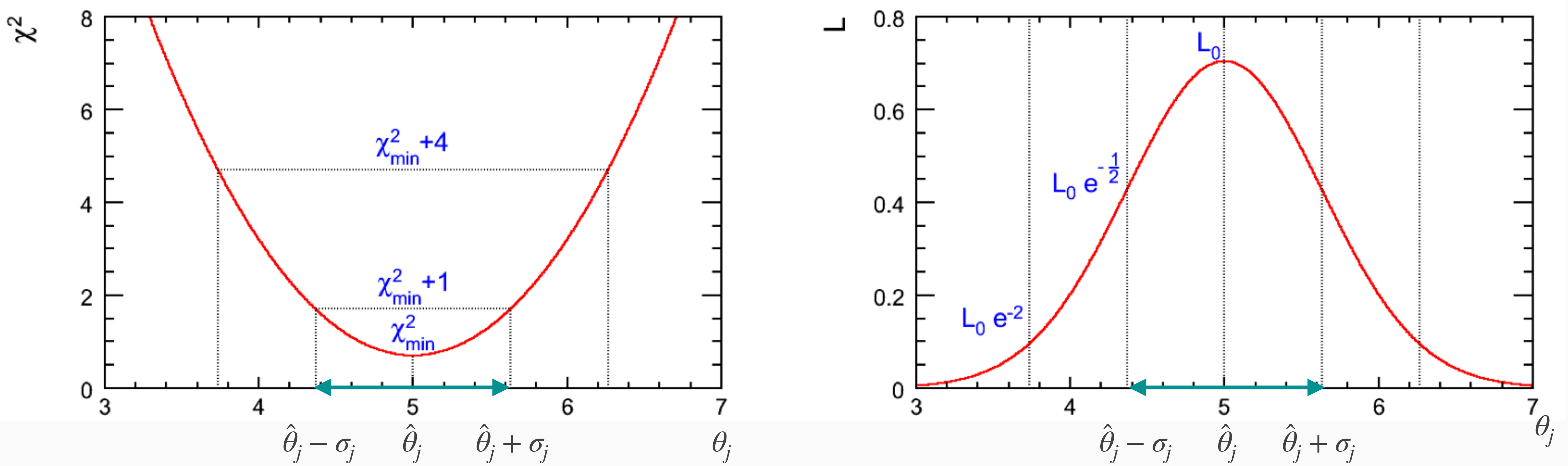
Uncertainties on the estimators $\hat{\theta}_j$

- The uncertainty on $\hat{\theta}_j$ ($\pm 1\sigma_j$) is given by the interval around $\hat{\theta}_j$ by which the χ^2 increases by 1 (or $\mathscr{L}$ decreases by $e^{-1/2}$)



1) For measurements w/ Gaussian errors, the likelihood is Gaussian

2) If $f(x;\vec{\theta})$ is linear on $\vec{\theta}$, the χ^2 is parabolic on $\vec{\theta}$



Figures taken from Mark Thomson's slides on statistics, 2015

- Under hypotheses 1) and 2), the uncertainties on the estimators are related to the second derivative of the χ^2 function:

$$\sigma_j^2 = \left(\frac{1}{2}\frac{\partial^2\chi^2}{\partial\theta_j^2}\bigg|_{\theta_j=\hat{\theta}_j}\right)^{-1} = \left(-\frac{\partial^2 \ln\mathscr{L}}{\partial\theta_j^2}\bigg|_{\theta_j=\hat{\theta}_j}\right)^{-1}$$

Extended maximum likelihood fit

- So far we have regarded the number of events, n, as being fixed, but in many cases we are interested in measuring absolute rates
- The number of events, n, follows a Poisson distribution with mean λ: $\mathscr{P}(n;\lambda) = \dfrac{\lambda^n e^{-\lambda}}{n!}$
- This distribution can be incorporated in the likelihood as a multiplicative factor:
  $\mathscr{L}(\vec{\theta}, n) = \dfrac{\lambda^n e^{-\lambda}}{n!} \prod_{i=1}^{N} \mathscr{P}(y_i(x_i); \vec{\theta})$
- In this case, minimising the χ^2 is no longer equivalent to minimising $-2\ln\mathscr{L}$

$$-2\ln\mathscr{L} = -2\left[\sum_{i=1}^{N} \ln \mathscr{P}(y_i(x_i), \vec{\theta}) + n\ln\lambda - \lambda + \text{const.}\right]$$

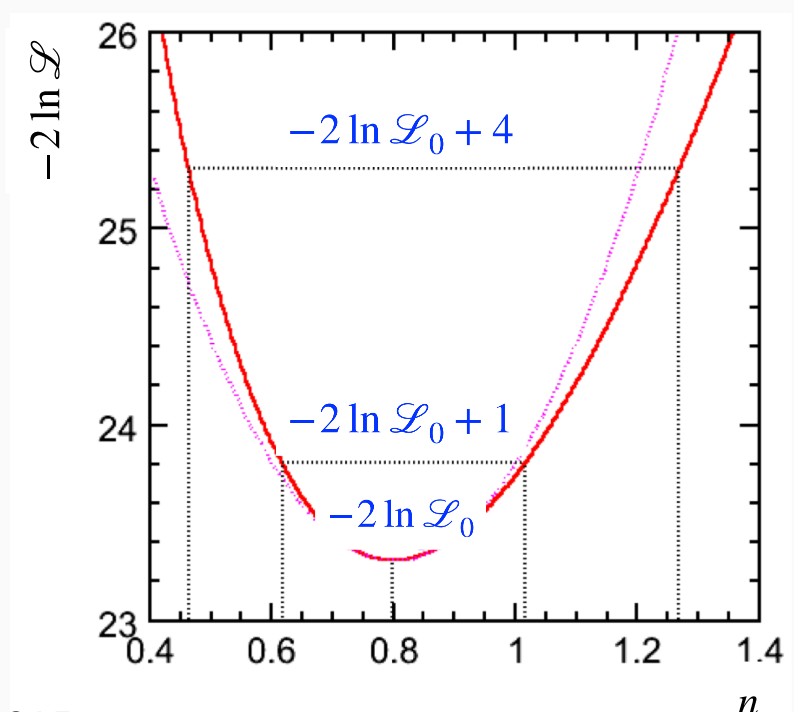


Figure taken from Mark Thomson's slides on statistics, 2015

χ^2 distribution

- If the model $f(x;\vec{\theta})$ is correct and the measurements y_i have Gaussian uncertainties, then the χ^2 is a random variable that follows the χ^2 p.d.f:

$$P_{\chi^2}(x;\nu) = \frac{x^{\frac{\nu}{2}-1}e^{-\frac{\nu}{2}}}{2^{\frac{\nu}{2}}\,\Gamma(\frac{\nu}{2})},\ \nu > 0$$

with mean $<\chi^2> = \nu$ and

variance var(χ^2)=2ν

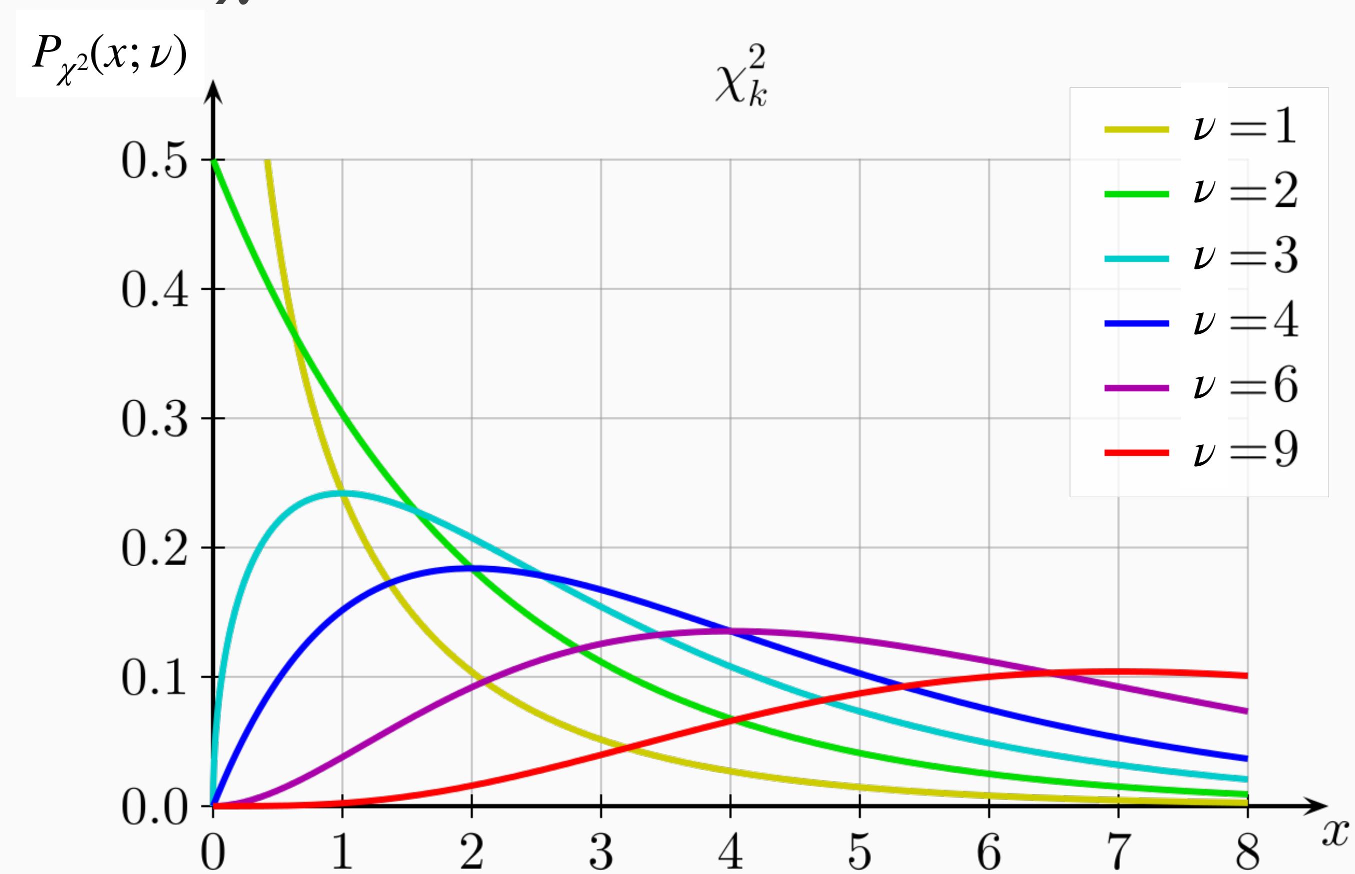


Figure taken from Wikipedia

- ν is the number of degrees of freedom

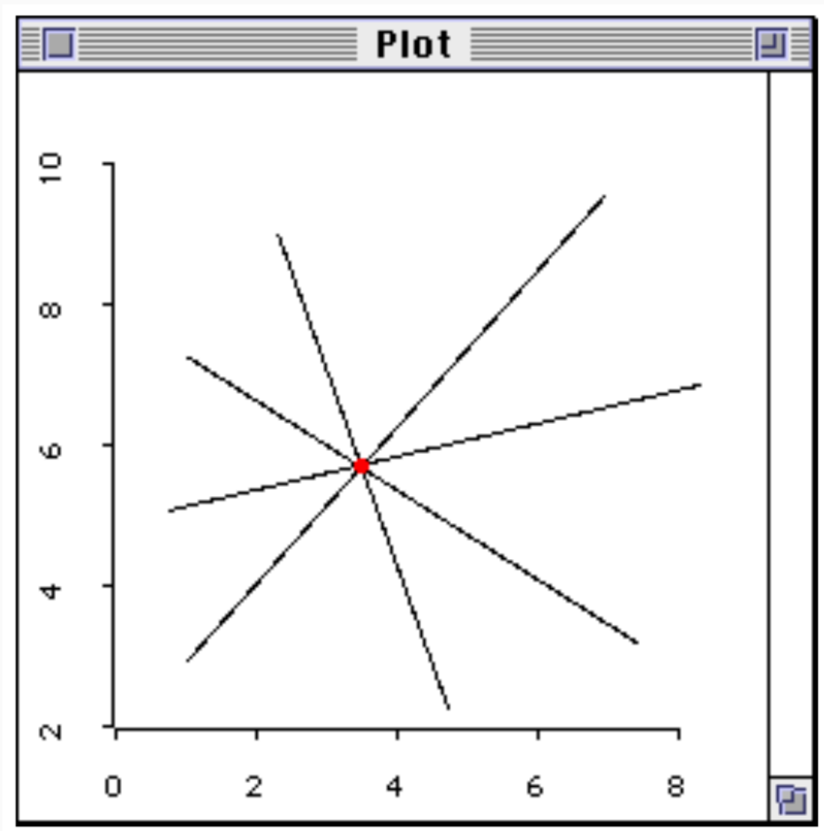
$$\nu = N - N_p$$

Number of measurements y_i

Number of fitted parameters θ_j

Number of degrees of freedom

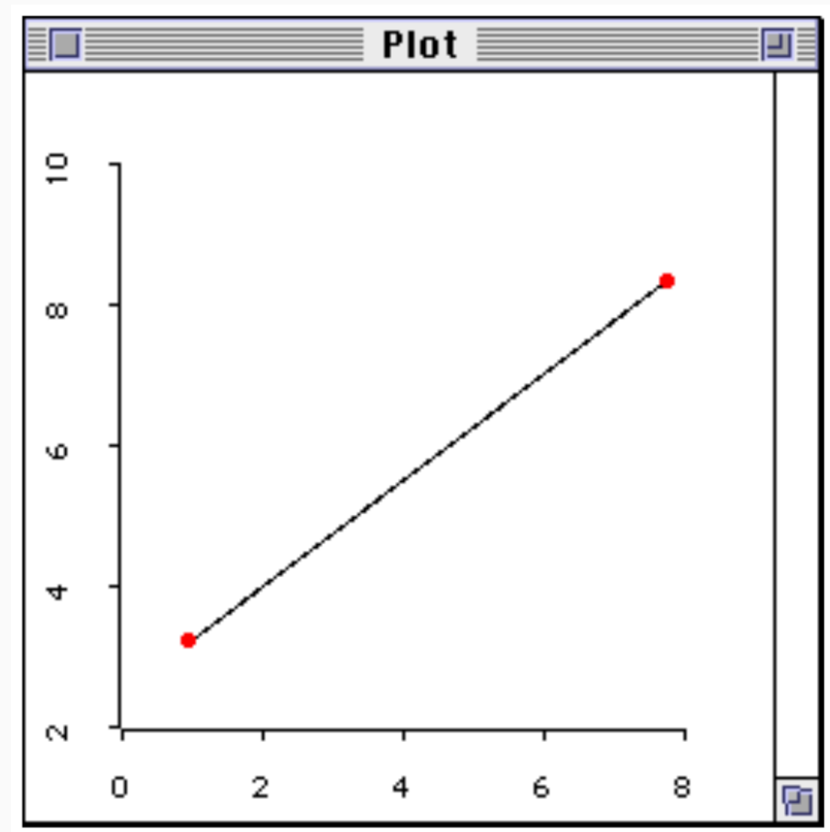
- Fitting a line to:



N = 1, m = 2

$\nu = -1$

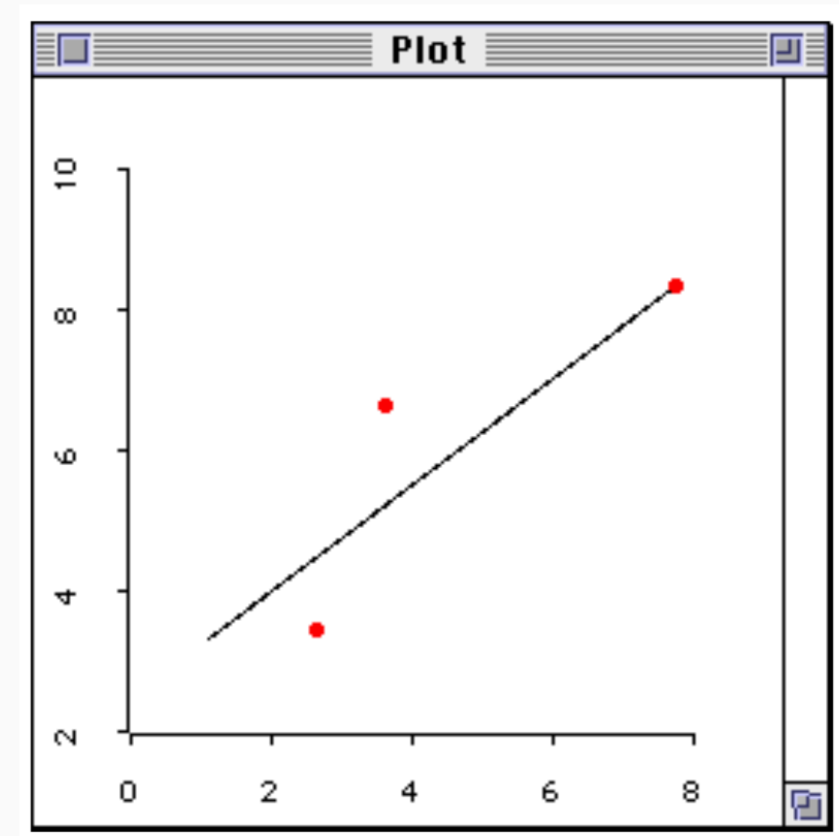
Any line can pass by 1 point. Every line is the “best” line. Impossible to solve.



N = 2, m = 2

$\nu = 0$

2 points define exactly a line. The method has no freedom to choose different values for the parameters.



N = 3, m = 2

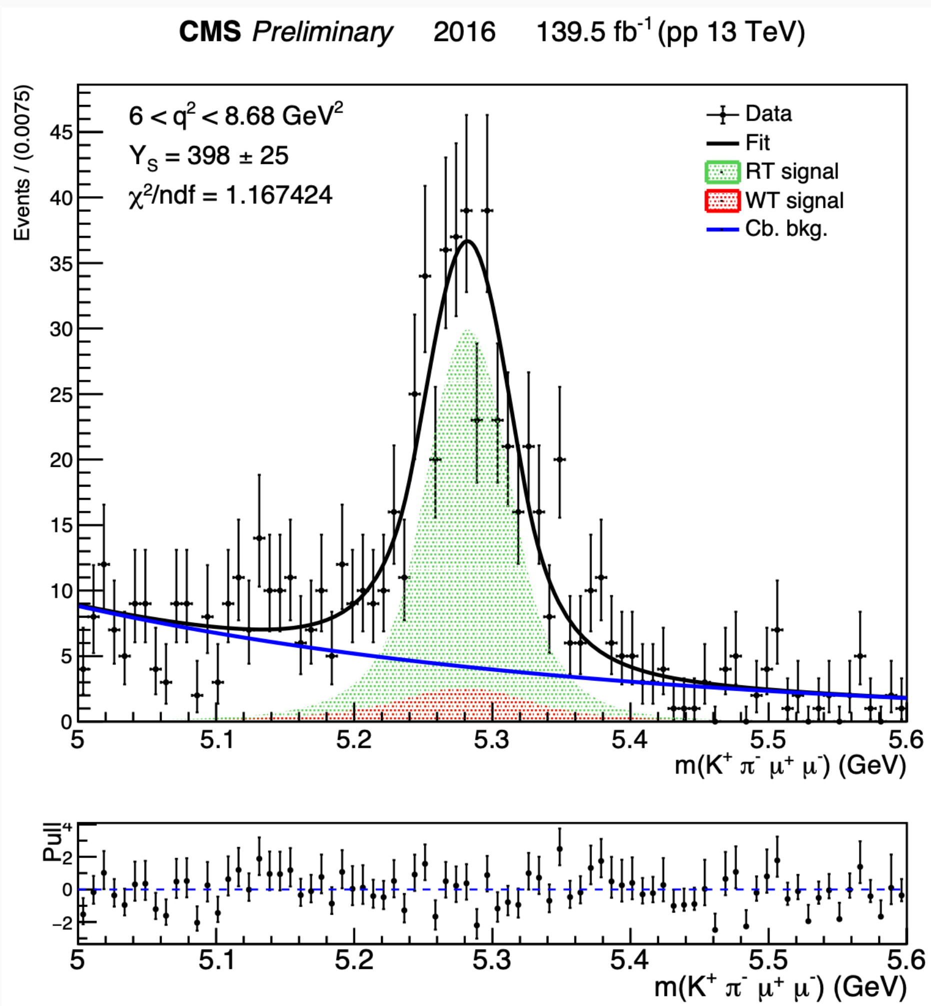
$\nu = 1$

Method will adjust the value of the parameters to find the best fit. It has some freedom.

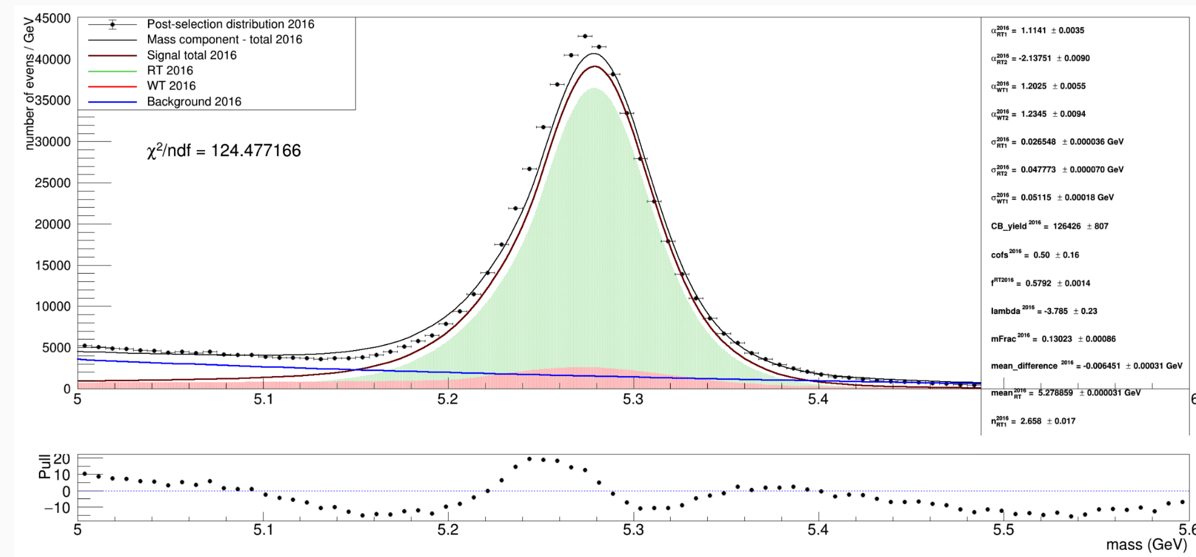
Figures taken from https://creative-wisdom.com/computer/sas/df.html

Goodness-of-fit test

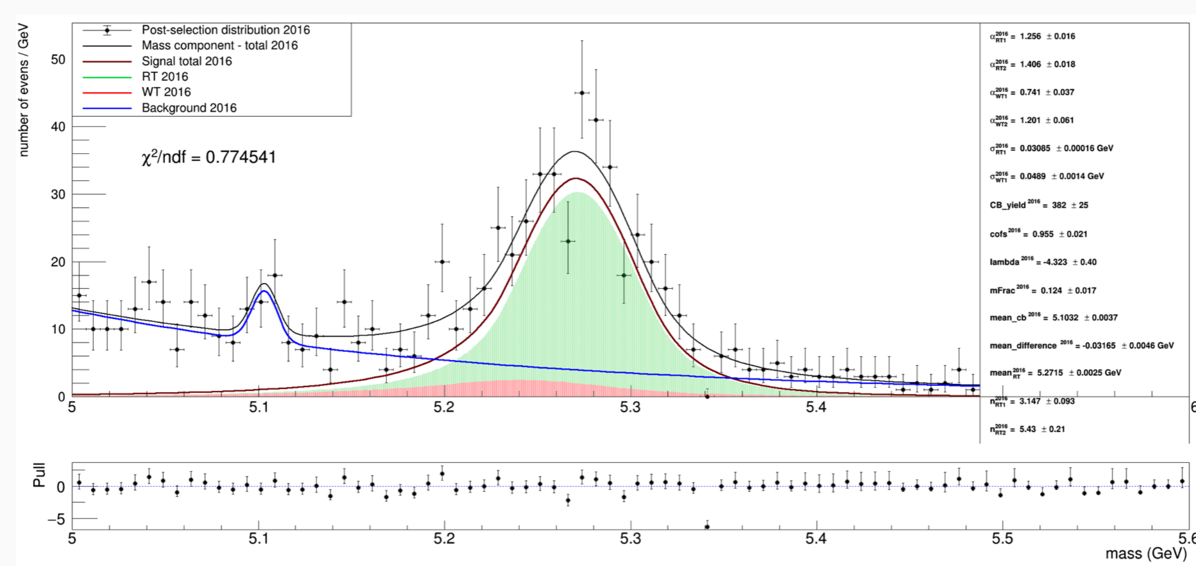
- We want to quantify the level of agreement between the data and the fit model
- This can be determined from the ratio χ^2/ν



$\chi^2/\nu \approx 1$, good fit



$\chi^2/\nu \gg 1$, poor fit: the model does not fully capture the data



$\chi^2/\nu < 1$, model is over-fitting (improperly fitting noise, following a statistical fluctuation or uncertainties were overestimated)

Hypothesis testing

- We want to know whether we can accept or reject the hypothesis, H_0, that "the fitted model correctly describes the measured data points" (null hypothesis) against the alternate hypothesis, H_1, that "the fitted model does not correctly describe the measured data points"

- We use as "test statistic" the χ^2, whose value reflects the level of agreement between the data and the hypothesised model

- We define the p-value as the probability of finding a value of χ^2 equal or greater than the observed one χ^2_{obs}:

$$p = \int_{\chi^2_{obs}}^{\infty} P_{\chi^2}(x;\nu)dx$$

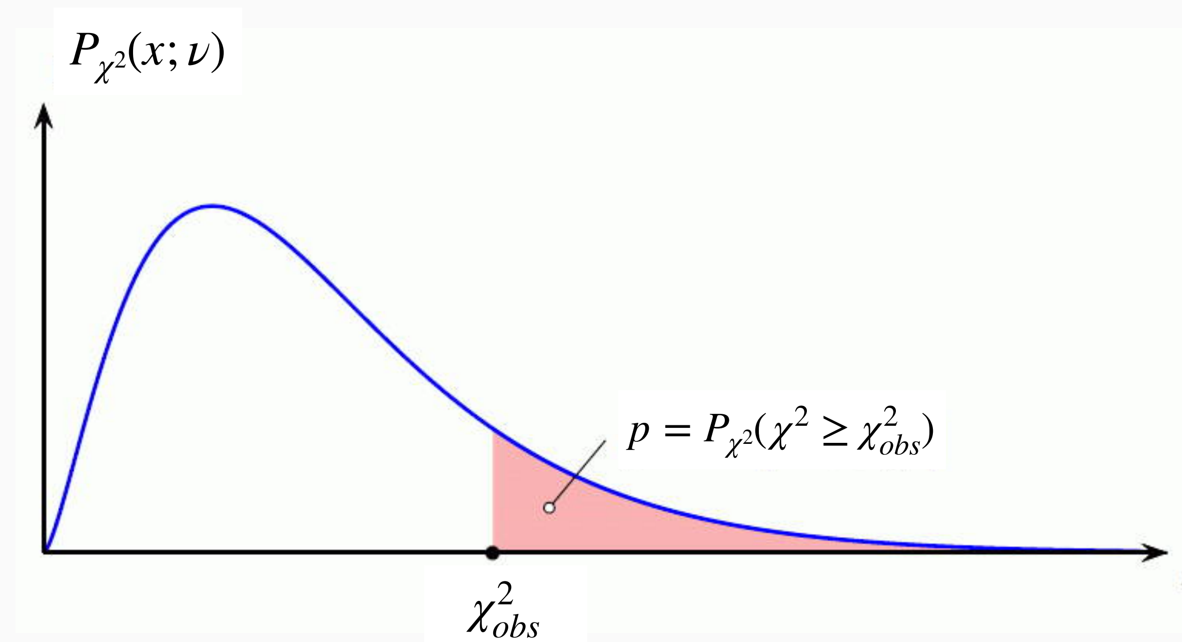


Figure taken from https://www.di-mgt.com.au/chisquare-calculator.html

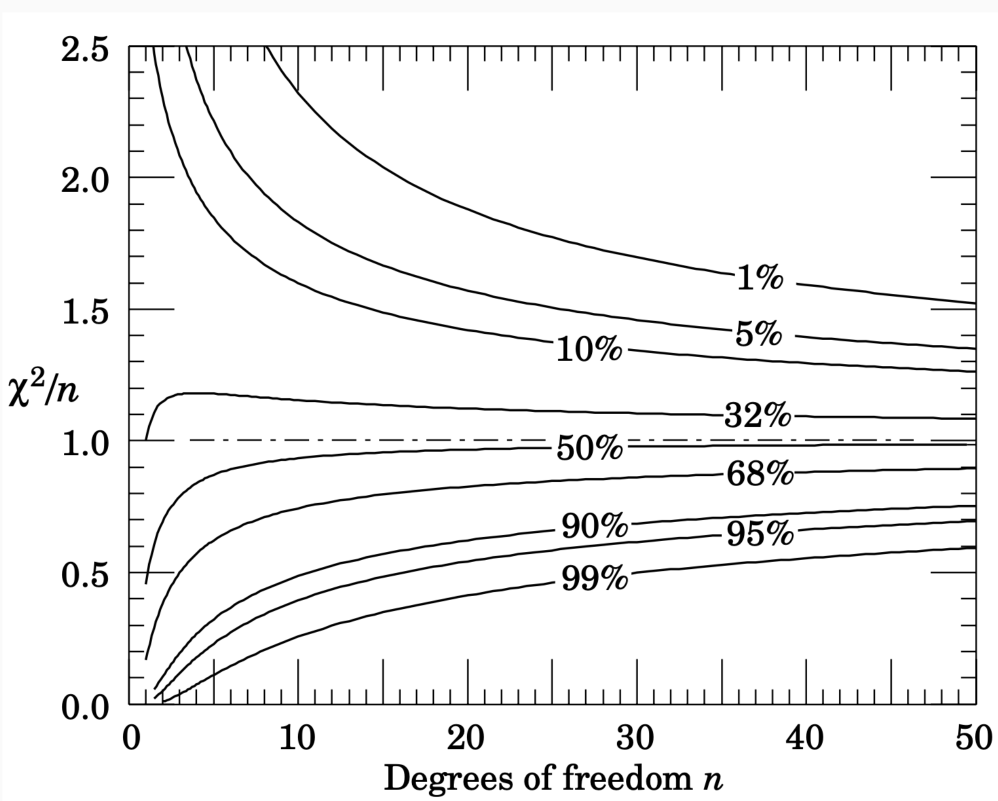


Figure taken from PDG Ch.40 Statistics

Standard guidelines:

$p > 0.10$ →	not significant	→ do not reject null hypothesis
$p \leq 0.10$ →	marginally significant	
$p \leq 0.05$ →	significant	
$p \leq 0.01$ →	highly significant	→ reject null hypothesis

Backup