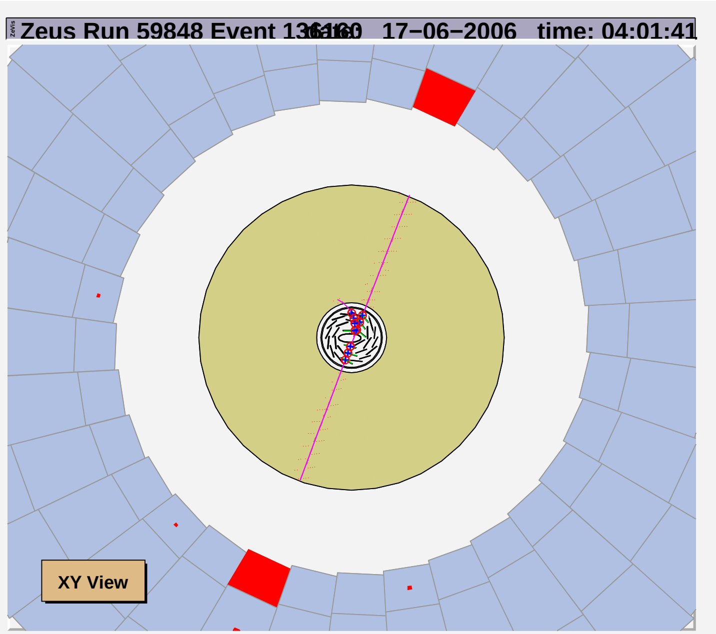
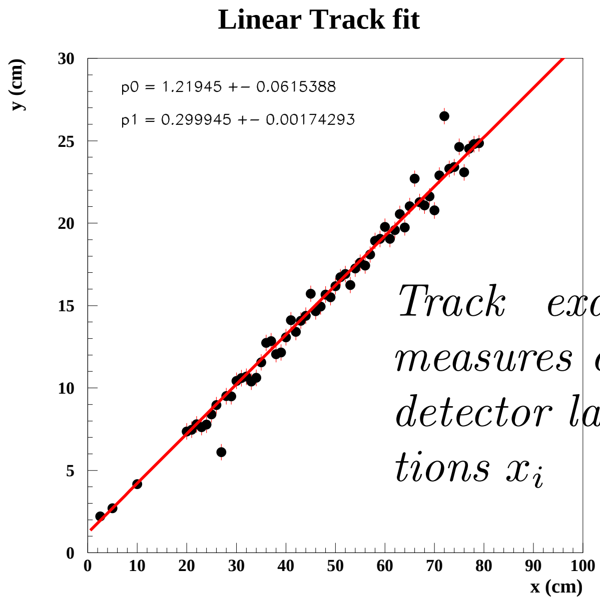
Linear Least square fit - introductory example

Example: Precise muon track fits for possible discovery $Z^* \to \mu^+\mu^-$





Track example: Detector measures coordinate y_i in n detector layers at fixed positions x_i

- Necessary conditions for linear least square fit:
  - Measurements with gaussian uncertainties
  - Linear model, here: $y = a_0 + a_1 x + a_2 x^2$
- Fit construction:
  - $\chi^2 = \sum_i \frac{\left[y_i - (a_0 + a_1 x + a_2 x^2)\right]^2}{\sigma_i^2}$
  - Determine a_0, a_1, a_2 by finding χ^2 total minimum (normal equations)
- Check consistency
  - χ^2 and fit probability
  - Outlier rejection
- Detailed error analysis
  - Parameter errors and correlations (error ellipses), track trajectory error band
  - Momentum calculation (error propagation)
- Possible Extensions:
  - Apply constraint fits to both tracks, e.g. $p_t(\mu^+) = p_t(\mu^-)$ $\to$ covered in session on extended fits
  - Analysis of obtained $\mu^+\mu^-$ mass spectrum containing background and possible signal events $\to$ covered in session on non-linear least square fits

Overview of Linear least square fit section

Part I	Part II	Part III
• Reminder of χ^2-fit method • Linear χ^2-fit examples (Constant, straight line, parabola, etc.) • **Fit of a constant (averaging measurements)** • One single measurement: χ^2_{min} and $\chi^2_{min}+1$, Hesse matrix • Exercise: Two measurements: perform fit by adding χ^2-parabolas • Averaging many measurements, results • Exercise: Compare weighted vs unweighted average	• **χ^2-fit-quality test:** Example: χ^2 of two measurements and known true value • χ^2-function for n degrees of freedom and χ^2-fit probability Exercise: plot and study features of the χ^2-function vs n using the parameterised function New: Generate 1000 random experiments with n degrees of freedom and obtain χ^2 and χ^2-fit probability distributions • χ^2 for two measurements with unknown true value • New exercise: Track position measurement in test beam using 10 detector layers, in each detector 99% chance for signal hit and 1% for random noise hit $\longrightarrow$ Generate 1000 tracks and corresponding hits and obtain χ^2, χ^2-fit probability and measured parameter distributions. Try to reject outliers: Method 1: reject track fits with small χ^2-fit probability, Method 2: iterative, repeat track-fit and downweight outliers • Exercise: **Outlier rejection**, case world average of m_W, study how the rejection of certain measurements change the average and the χ^2-fit probability • New exercise: Upscaling of errors a la PDG to obtain reasonable χ^2 • New exercise: Pulls of single measurements to the average	• General form of linear χ^2 • Solution by normal equations • Normal equation solution for **straight line fit** • Exercise: Learn qualitative features of straight line fits, e.g. importance of lever arm • Exercise: Straight line fit and detailed error analysis (error ellipse, trajectory error band) • New exercise: Coordinate transformation such that the coordinate center is in the middle of the points $\longrightarrow$ study the effect on the parameter errors and correlation • New exercise: Add a very precise point at the origin of the track such that the p_0 parameter is basically fixed. Repeat the track-fit and study the effect on the slope and error • Exercise: **Parabola track fit**, complete analysis: *fit, outlier-rejection, parameter errors/correlation, trajectory uncertainty,* ***momentum calculation***

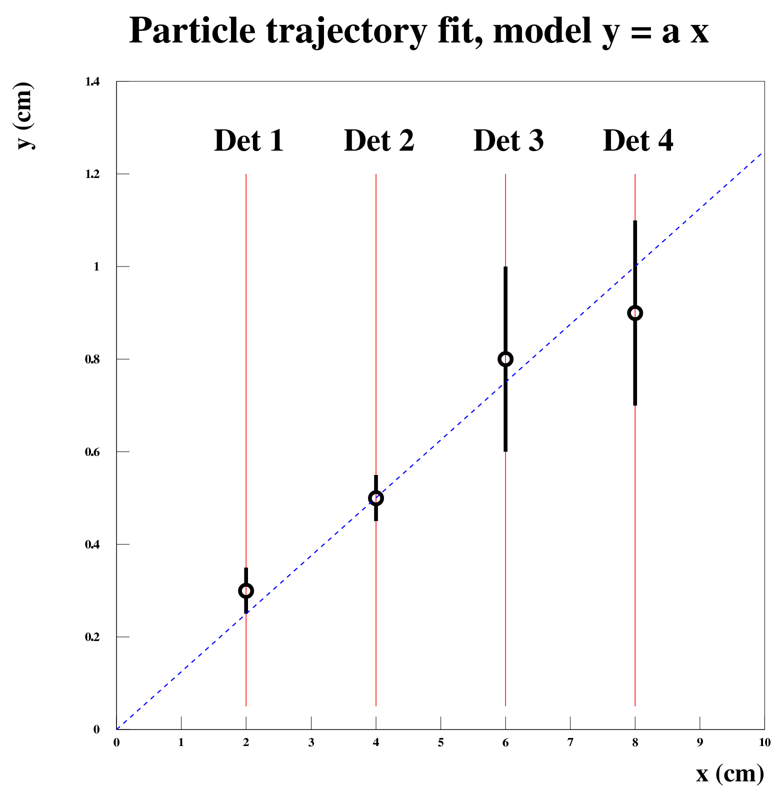
Lecture I on linear least square fits

Contents:

- Least square χ^2-fit method reminder
- Linear least square fits: definition and examples
- Fit of a constant: One, two and many measurements

Method of least squares fit - a reminder

Example problem: Particle trajectory measurement



general:

n-measurements y_i

with uncertainties σ_i

at fixed x_i

Model: $y = f(x, a)$

$\Rightarrow$ how to determine a?

Idea: for the correct a one expects: $|y_i - f(x_i, a)| \leq \sigma_i$,

i.e. curve describes data within measurement uncertainties

Method of least squares fit - a reminder

$\rightarrow \boxed{\chi^2 = \Sigma_{i=1}^{n} \frac{(y_i - f(x_i,a))^2}{\sigma_i^2}} \leftrightarrow$ Minimum with respect to a

$\Rightarrow$ determine a from $\frac{d\chi^2}{da} = 0$

$\rightarrow \boxed{\frac{d\chi^2}{da} = \Sigma_{i=1}^{n} \frac{(y_i - f(x_i,a))}{\sigma_i^2} \cdot \frac{df(x_i,a)}{da} = 0}$

In general not analytically solvable.
$\Rightarrow$ use iterative (numerical) methods (MINUIT, Mathematica)

Method of least squares fit - a reminder

Most general case

- y_i, y_j correlated measurements with covariance V_{ij}
- m fitparameters $\vec{a}$

$$\rightarrow \boxed{\begin{aligned} \chi^2 &= \Sigma^n_{i,j=1}(y_i - f(x_i,\vec{a}))V^{-1}_{ij}(y_j - f(x_j,\vec{a})) \\ &= (\vec{y} - \vec{f}(\vec{a}))^t V^{-1}(\vec{y} - \vec{f}(\vec{a})) \end{aligned}}$$

Linear least square fits

$\vec{y}$ vector of n measurements $\begin{pmatrix} y_1(x_1) \\ . \\ y_n(x_n) \end{pmatrix}$ with cov-matrix V

$\vec{a}$ vector of m fitparameters $\begin{pmatrix} a_1 \\ . \\ a_m \end{pmatrix}$

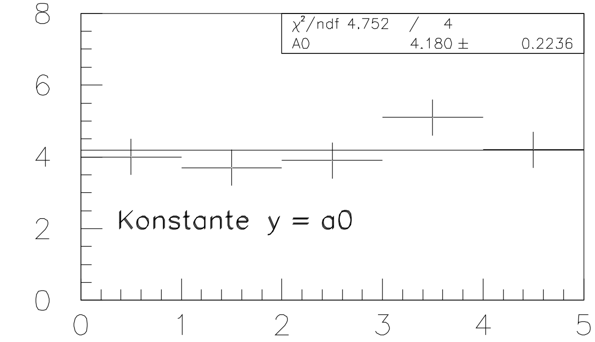
Model $\vec{y} : = A\,\vec{a}$ <u>Example:</u> $y = ax$; $\vec{a} = (a)$, $A = \begin{pmatrix} x_1 \\ .. \\ x_n \end{pmatrix}$

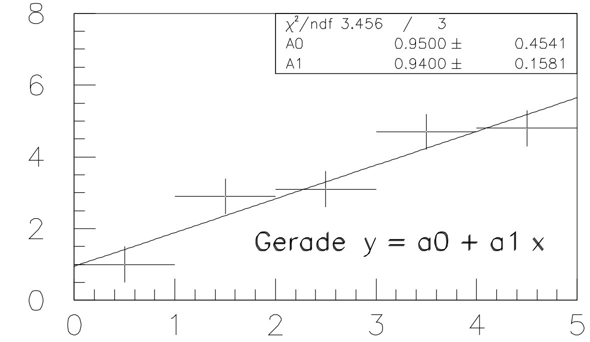
$$\boxed{\chi^2 = (\vec{y} - A\,\vec{a})^t\, V^{-1}\, (\vec{y} - A\,\vec{a})}$$

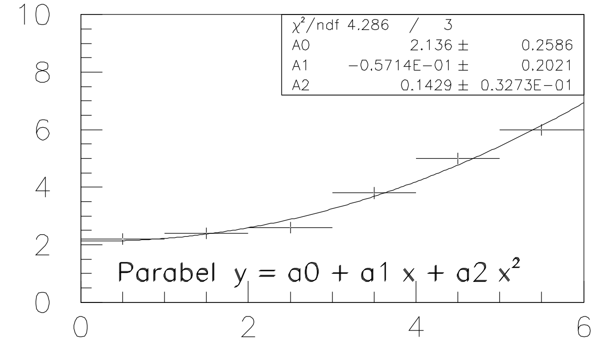
$\longrightarrow$ to be minimised w.r.t $\vec{a}$

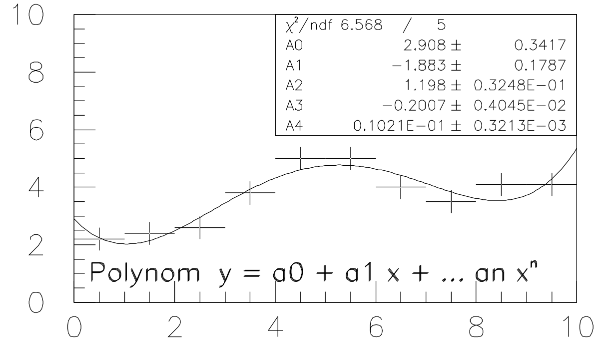
Examples for linear least square fits

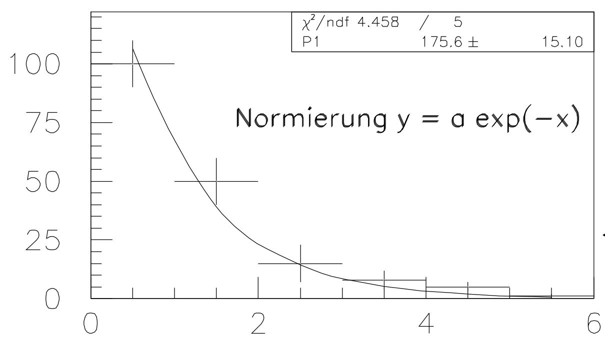
Linear means that y depends linearly on the fitparameters a_i.











← *Watch out: function can be highly non-linear in x*

Fit of a constant

Example: Averaging of n different measurements $a_i \pm \sigma_i$ of an observable a (e.g. $a = \alpha_s(m_Z)$)

$$\chi^2 = \Sigma_i^n \frac{(a_i - a)^2}{\sigma_i^2}$$

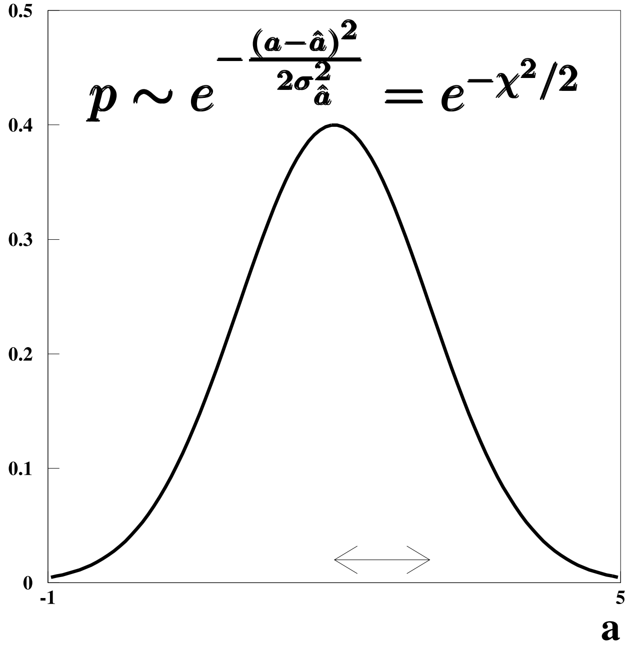
"Idiot example" of one single measurement $a_1 \pm \sigma_1$:

$$\chi^2 = \frac{(a_1 - a)^2}{\sigma_1^2}$$

$$Min.\chi^2 : \quad \frac{d\chi^2}{da} = 0 \ \rightarrow \ \text{Estimated value} \quad \hat{a} = a_1; \ \sigma_{\hat{a}} = \sigma_1$$

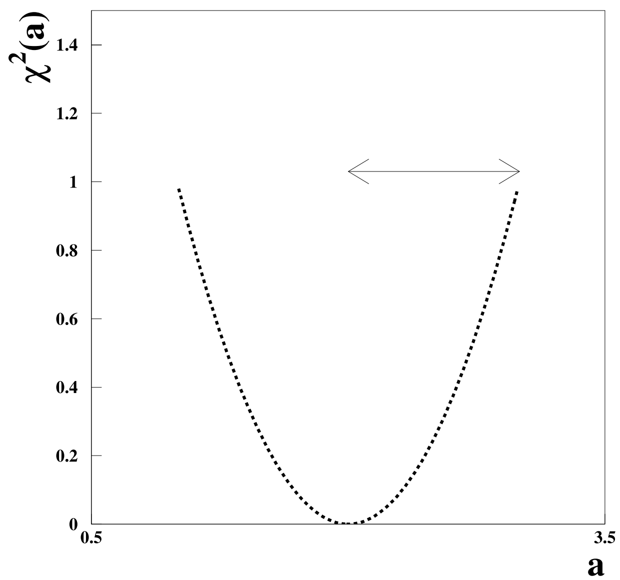
Fit of a constant *(single measurement)*

Probability density p for true value of a (*inverse probability*):



$\rightarrow$ Important general relation:

$$\sigma_{\hat{a}} = \left[-\frac{d^2\chi^2}{da^2}_{|a=\hat{a}} \right]^{-1/2}$$



$\rightarrow$ Important general relation:

$$\chi^2(\hat{a} + \sigma_{\hat{a}}) = 1$$

Generalisation to any one-parameter fit

Taylor expansion of χ^2 around estimated value $\hat{a}$:

$$\chi^2 = \chi^2(\hat{a}) + \frac{d\chi^2}{da}_{|a=\hat{a}} \cdot (a - \hat{a}) + \frac{1}{2}\frac{d^2\chi^2}{da^2}_{|a=\hat{a}} \cdot (a - \hat{a})^2 + \ldots$$

$$= \chi^2(\hat{a}) + H \cdot (a - \hat{a})^2 \quad \text{with} \quad H = \frac{1}{2}\frac{d^2\chi^2}{da^2}_{|a=\hat{a}} \quad \text{Hesse matrix}$$

Thus the *inverse probability density* is

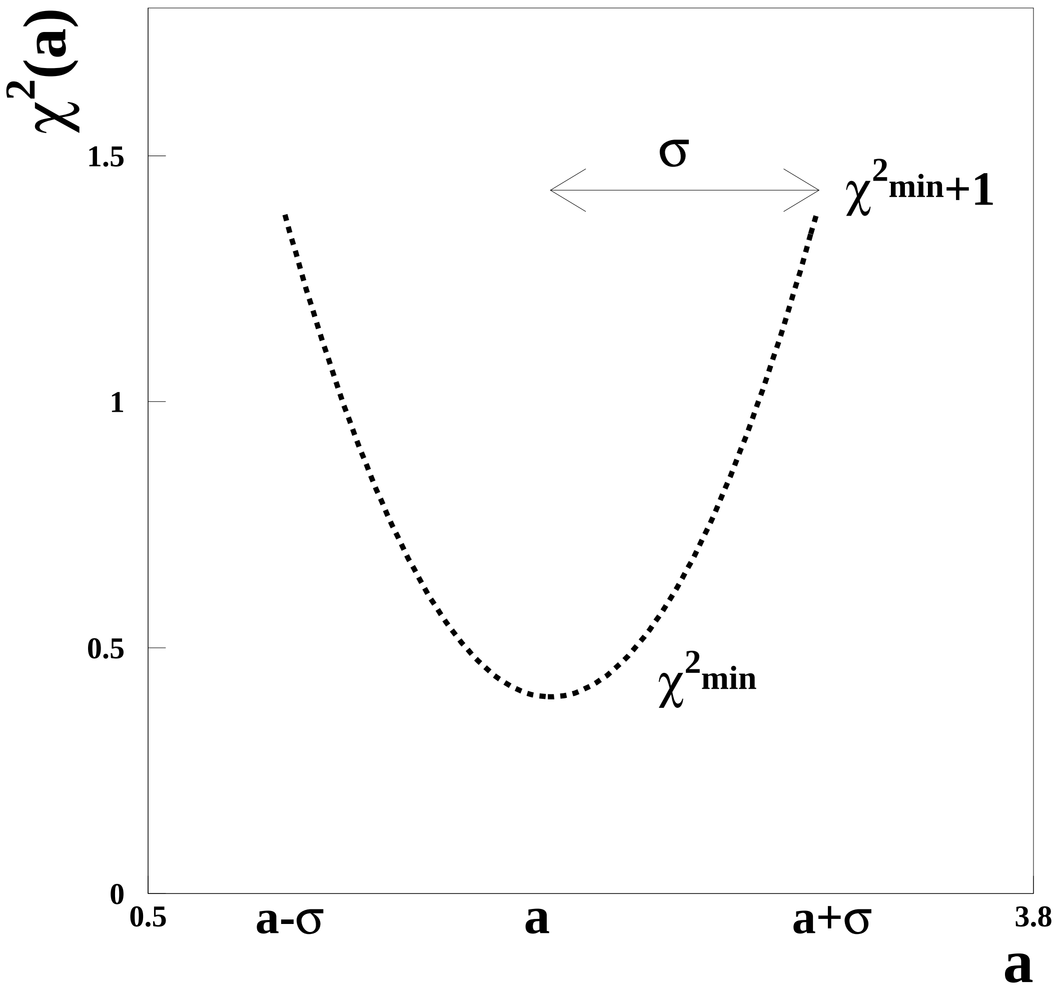
$$p(a|\hat{a}) \sim e^{-\frac{\chi^2(\hat{a})}{2}} \cdot e^{-\frac{1}{2}H\cdot(\hat{a}-a)^2)}$$

$\longrightarrow$ Gaussian distribution around $\hat{a}$ with width $\sigma = H^{-1/2}$

Generalisation to any one-parameter fit

$$\chi^2(a) = \chi^2(\hat{a}) + \frac{(a-\hat{a})^2}{\sigma^2}$$

$$\rightarrow \quad \chi^2(a \pm 1\sigma) = \chi^2(\hat{a}) + 1 = \chi^2_{min} + 1$$

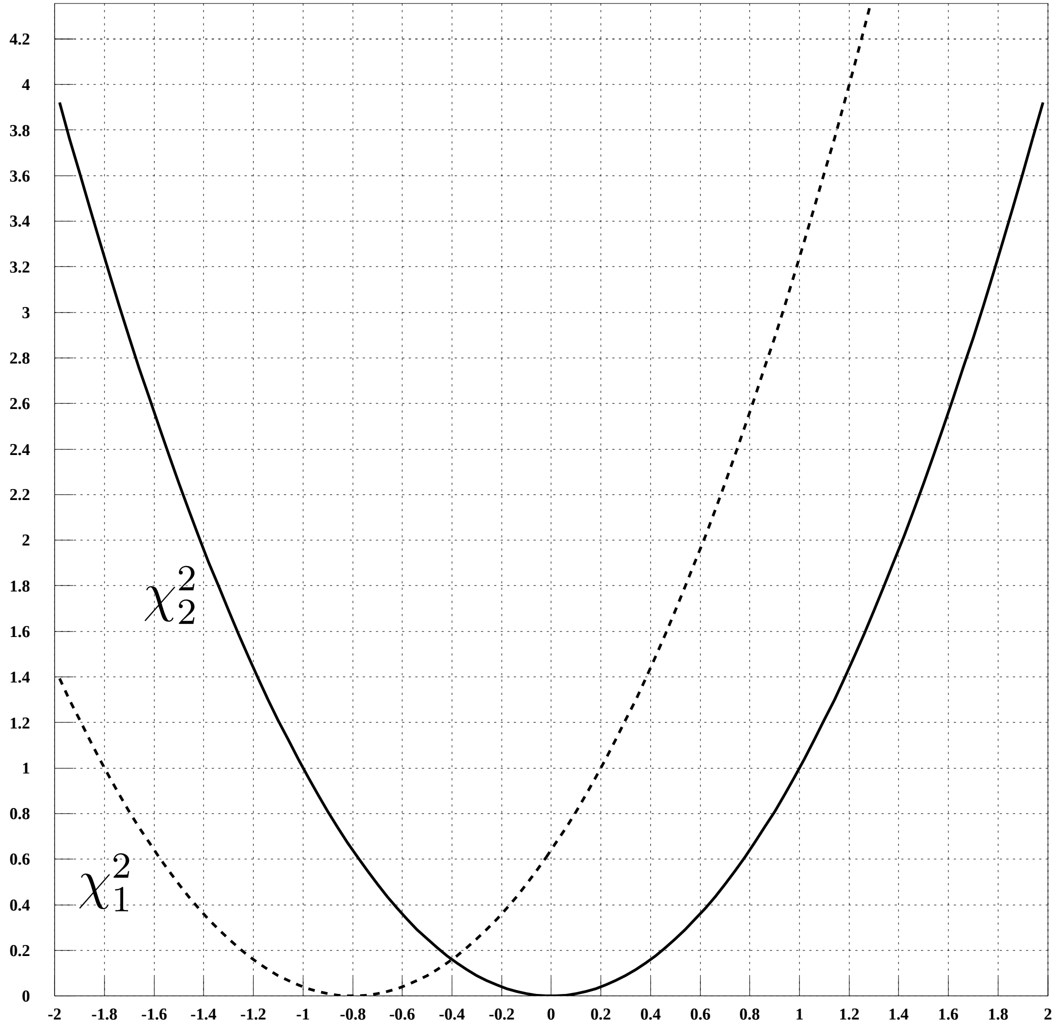


→ Read error directly from χ^2 curve

Mini-exercise Averaging of two meas. via χ^2 parabolas

Two measurements y_1, y_2 of the same quantity a can be represented by their χ^2 parabolas: $\chi_i^2 = (y_i - a)^2/\sigma_i^2; \quad i = 1, 2$

- Draw for the example on the right the total χ^2, i.e. the sum of the two parabolas (yes, do it simply by hand :-))
- Read off the value $\hat{a}$ (where the total χ^2 is minimal)
- Estimate the error $\sigma_{\hat{a}}$ from the points where $\chi^2 = \chi^2_{min} + 1$
- How much is the error $\sigma_{\hat{a}}$ reduced compared to the errors of the two original measurements?



Averaging several measurements

n measurements $y_i \pm \sigma_i$ of the same quantity a $\longrightarrow$ what is the best way to average?

(Why is $\frac{1}{n}\Sigma y_i$ not the best? $\longrightarrow$ measurements with large errors get too much weight and can spoil the average!)

$$\chi^2 = \Sigma_{i=1}^n \frac{(y_i - a)^2}{\sigma_i^2}$$

$$\frac{d\chi^2}{da} = 0 = \Sigma_{i=1}^n \frac{-2(y_i - a)}{\sigma_i^2} = -2\Sigma_{i=1}^n \frac{y_i}{\sigma_i^2} + 2a\Sigma_{i=1}^n \frac{1}{\sigma_i^2}$$

$$\longrightarrow \quad \boxed{\hat{a} \;=\; \Sigma_{i=1}^n \frac{y_i}{\sigma_i^2} / \Sigma_{i=1}^n \frac{1}{\sigma_i^2}} \tag{1}$$

Averaging several measurements

$\longrightarrow$ Single measurements contribute with weight $G_i = \dfrac{1}{\sigma_i^2}$;

Define $G_s := \Sigma_{i=1}^n G_i$; Hesse matrix $H = 1/2 \dfrac{d^2\chi^2}{da^2} = G_s$

$$\longrightarrow \quad \boxed{\hat{a} \;=\; \frac{1}{\Sigma_{i=1}^n G_i} \cdot \Sigma_{i=1}^n G_i y_i = \frac{1}{G_s} \cdot \Sigma_{i=1}^n G_i y_i} \tag{2}$$

Error on $\hat{a}$:

$$\boxed{\begin{array}{l} \sigma_{\hat{a}}^2 = \Sigma_{i=1}^n \left(\dfrac{d\hat{a}}{dy_i}\right)^2 \cdot \sigma_i^2 = \Sigma_{i=1}^n \left(\dfrac{G_i}{G_s}\right)^2 \cdot \sigma_i^2 \\ = \dfrac{1}{G_s^2} \cdot \Sigma_{i=1}^n G_i = \dfrac{1}{G_s} = \dfrac{1}{\Sigma_{i=1}^n 1/\sigma_i^2} \end{array}}$$

In short

$$\boxed{\hat{a} = \frac{\Sigma_{i=1}^n y_i/\sigma_i^2}{\Sigma_{i=1}^n 1/\sigma_i^2} \;\pm\; \frac{1}{\sqrt{\Sigma_{i=1}^n 1/\sigma_i^2}}}$$

Practical work: gain from weighted average

A radioactive source is completely surrounded by two hemispherical counters, one with 100% efficiency and the other with 10% efficiency only. An experimenter observes within one minutes 100 ± 10 decays in the first counter and 9 ± 3 in the second.

Estimate the total decay rate of the source per minute using the following two methods - compare the results and their precision:

A) unweighted average:

1. Correct for each counter the observed number for the counter inefficiency
2. Sum the corrected numbers in both counters to the total number of decays and determine an uncertainty using error propagation

B) weighted average:

1. Assume that without any inefficiency each of the counters has 50% geometrical acceptance, i.e. would see half of the decays. Determine two separate measurements of the total decay rate by correcting for each counter the efficiency corrected rate additionally by the geometric acceptance.
2. Calculate the weighted mean of the two separate measurements and an error

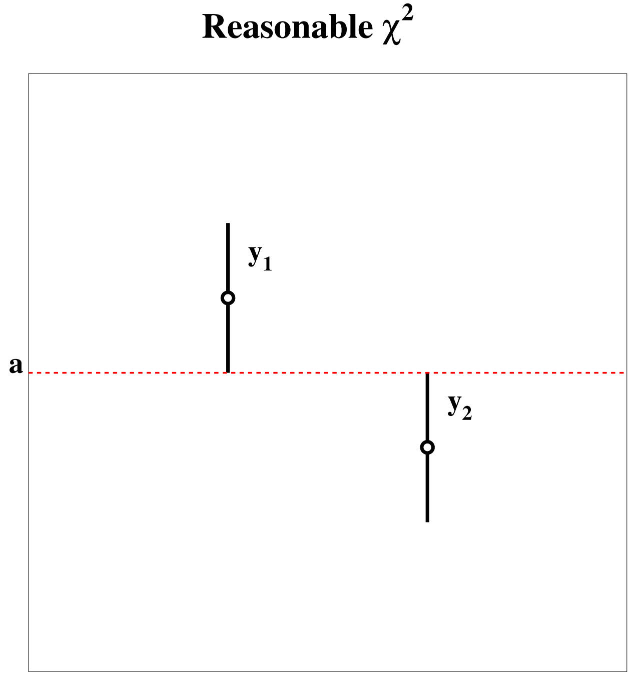
Lecture II on linear least square fits

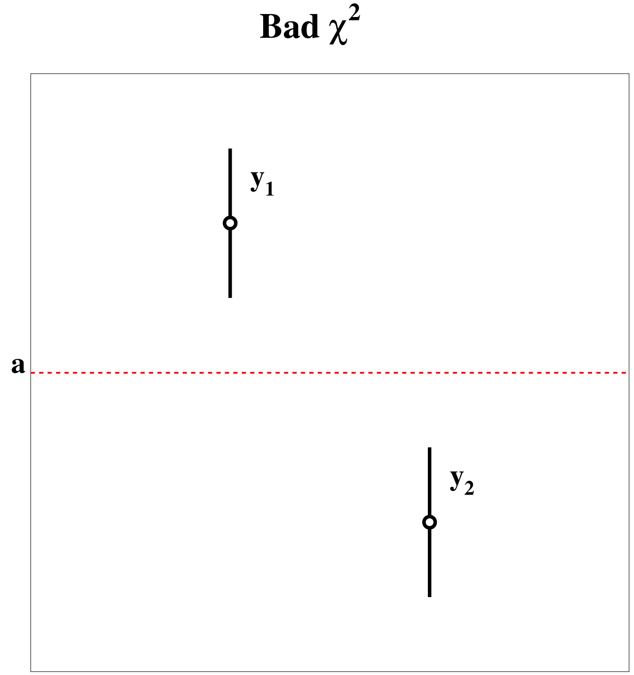
Contents:

- Consistency of measurements $\rightarrow \chi^2$ fit quality test
- Outlier rejection

Consistency of measurements

Example: Two measurements y_1 and y_2 with errors σ_1 and σ_2; the true value a is known, are the measurements consistent with a?:



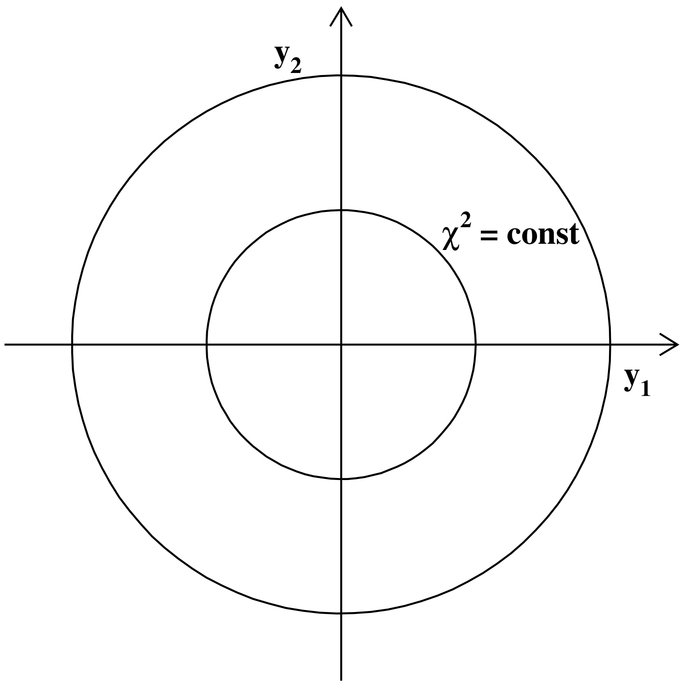


$$\chi^2 = \frac{(y_1 - a)^2}{\sigma_1^2} + \frac{(y_2 - a)^2}{\sigma_2^2} = 2 \qquad \chi^2 = 8$$

$\rightarrow \chi^2$ is a measure of the consistency

Consistency of measurements

Expected probability density for $\vec{y} = (y_1, y_2)$ (case $a = 0; \sigma_1 = \sigma_2 = \sigma$):



$$f(\vec{y})\, dy_1 dy_2 = \frac{1}{2\pi} e^{-y_1^2/2} e^{-y_2^2/2}\, dy_1 dy_2$$

$$= \frac{1}{2\pi} e^{-r^2/2}\, dy_1 dy_2 \quad \text{with } r = \sqrt{y_1^2 + y_2^2}$$

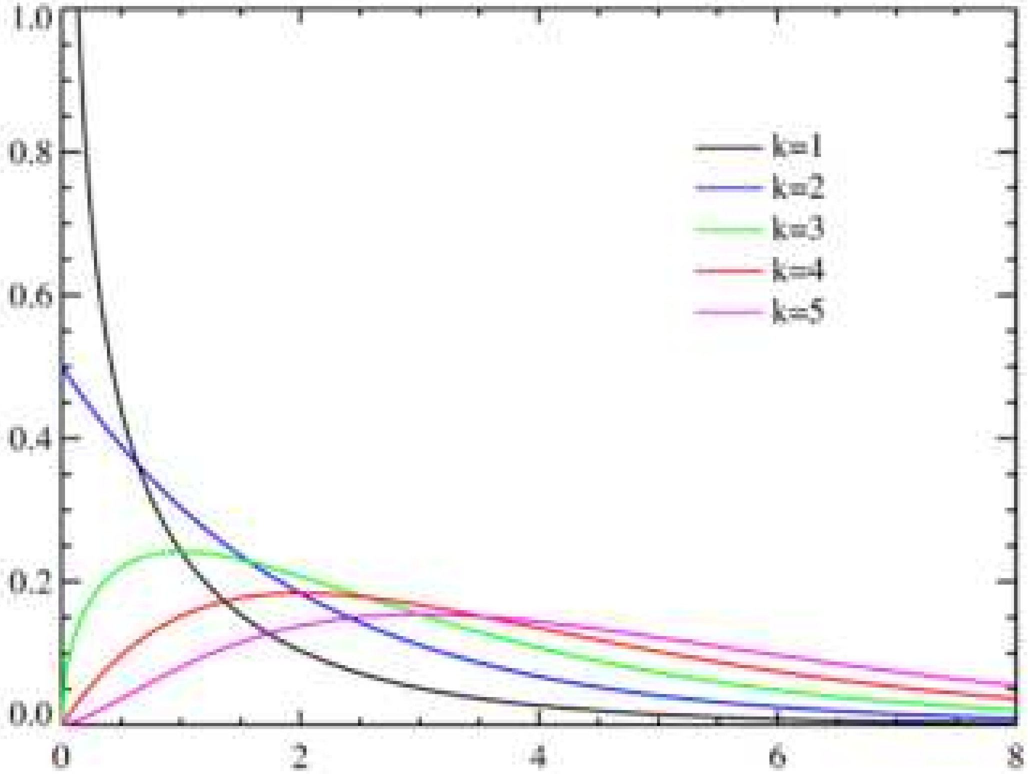
$$f(r)\, dr = \frac{2\pi r}{2\pi} r e^{-r^2/2}\, dr = r e^{-r^2/2}\, dr$$

$$\chi^2 = r^2 \longrightarrow$$

$$f(\chi^2)\, d\chi^2 = \frac{1}{2} e^{-\chi^2/2}\, d\chi^2$$

$\longrightarrow$ introduces χ^2-distr. for $z = \chi^2$ and two dimensions (ndf=2):
$f(z, 2) = \frac{1}{2} e^{-z/2}$

χ^2-function for n degrees of freedom



$$f(\chi^2, n) = \frac{1}{\Gamma(n/2)2^{n/2}} \cdot (\chi^2)^{n/2-1} \cdot e^{-\chi^2/2}$$

with $\Gamma(n/2) = \int_0^\infty dt e^{-t} t^{n/2-1}$

$$\int_0^\infty f(\chi^2, n) d\chi^2 = 1$$

$$\langle \chi^2 \rangle = n$$

$$V(\chi^2) = 2n;\ \sigma(\chi^2) = \sqrt{2n}$$

$$\rightarrow \langle \chi^2/n \rangle = 1$$

$$\sigma(\chi^2/n) = \sqrt{2/n}$$

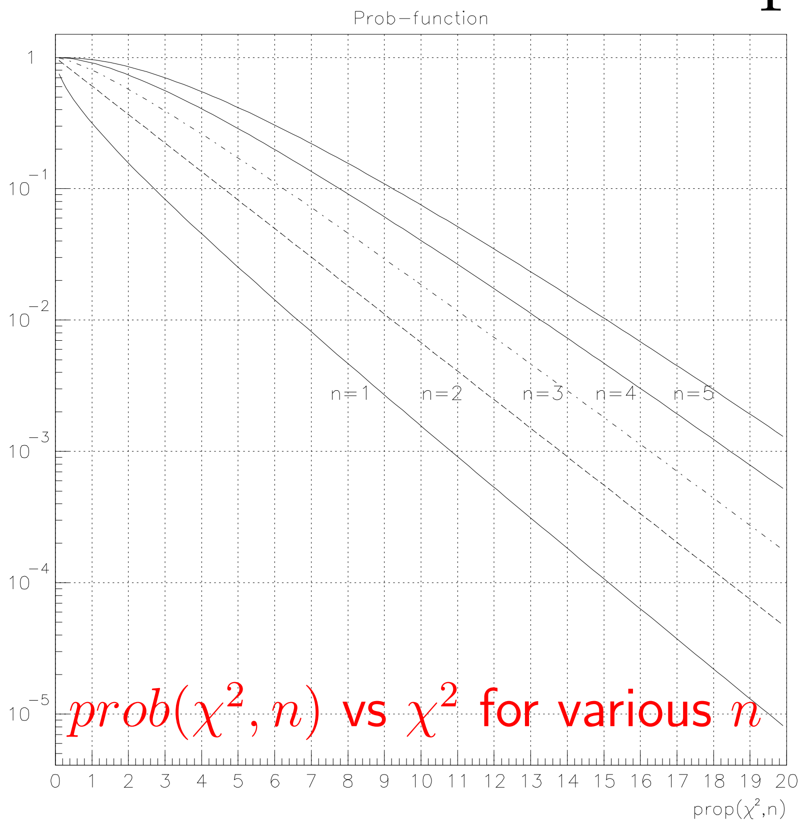
χ^2-fit probability

Common measure for consistency of measurements:

Probability that for repeated experiments a $\chi^2 \geq \chi^2_{actual}$ is observed

$$prob(\chi^2, n) = \int_{\chi^2}^{\infty} f(\chi^2, n) d\chi^2 \quad \text{subst. } t = \chi^2/2$$

$$= \frac{1}{\Gamma(n/2)} \cdot \int_{\chi^2/2}^{\infty} dt\, e^{-t}\, t^{n/2-1}$$



Add plot of expected flat distribution

χ^2 for two measurements with unknown true value

Until now the true value of a was assumed to be known, now replace by estimated $\hat{a}$;

Example of two measurements:

$$\chi^2_{min} = \frac{(y_1 - \hat{a})^2}{\sigma_1^2} + \frac{(y_2 - \hat{a})^2}{\sigma_2^2}$$

Using the weighted average $\hat{a} = \dfrac{G_1 y_1 + G_2 y_2}{G_1 + G_2}$ with $\quad G_i := 1/\sigma_i^2 \rightarrow$

χ^2 for two measurements with unknown true value

$$
\begin{aligned}
\chi^2_{min} &= G_1 \cdot \left(y_1 - \frac{(G_1 y_1 + G_2 y_2)}{G_1 + G_2} \right)^2 + G_2 \cdot \left(y_2 - \frac{(G_1 y_1 + G_2 y_2)}{G_1 + G_2} \right)^2 \\
&= G_1 \cdot \left(\frac{(G_2 y_1 - G_2 y_2)}{G_1 + G_2} \right)^2 + G_2 \cdot \left(\frac{(G_1 y_2 - G_1 y_1)}{G_1 + G_2} \right)^2 \\
&= \frac{G_1 G_2^2}{(G_1 + G_2)^2} (y_1 - y_2)^2 + \frac{G_2 G_1^2}{(G_1 + G_2)^2} (y_1 - y_2)^2 \\
&= \frac{G_1 G_2 (G_1 + G_2)}{(G_1 + G_2)^2} \cdot (y_1 - y_2)^2 = \frac{G_1 \cdot G_2}{G_1 + G_2} \cdot (y_1 - y_2)^2 \\
&= \frac{1}{1/G_1 + 1/G_2} \cdot (y_1 - y_2)^2 = \boldsymbol{\frac{1}{\sigma_1^2 + \sigma_2^2} \cdot (y_1 - y_2)^2}
\end{aligned}
$$

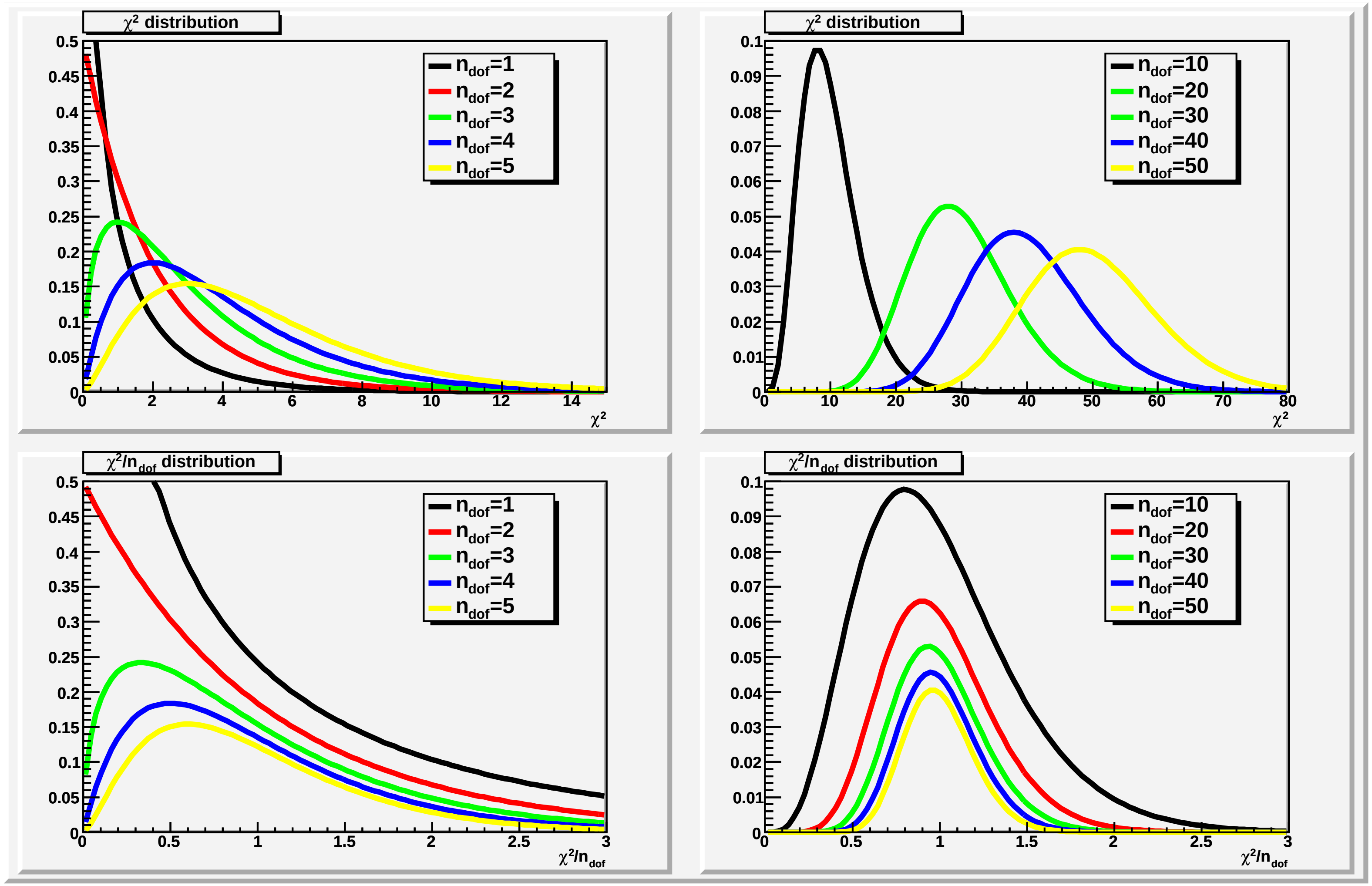
$\Delta = \frac{y_1 - y_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ should follow *(errorpropagation!)* gauss distribution $\sim e^{-\frac{\Delta^2}{2}}$

$\rightarrow \chi^2 = \Delta^2$ follows <u>1-dim</u> χ^2 distr.!

$\rightarrow$ One degree of freedom "sacrificed" for determination of $\hat{a}$.

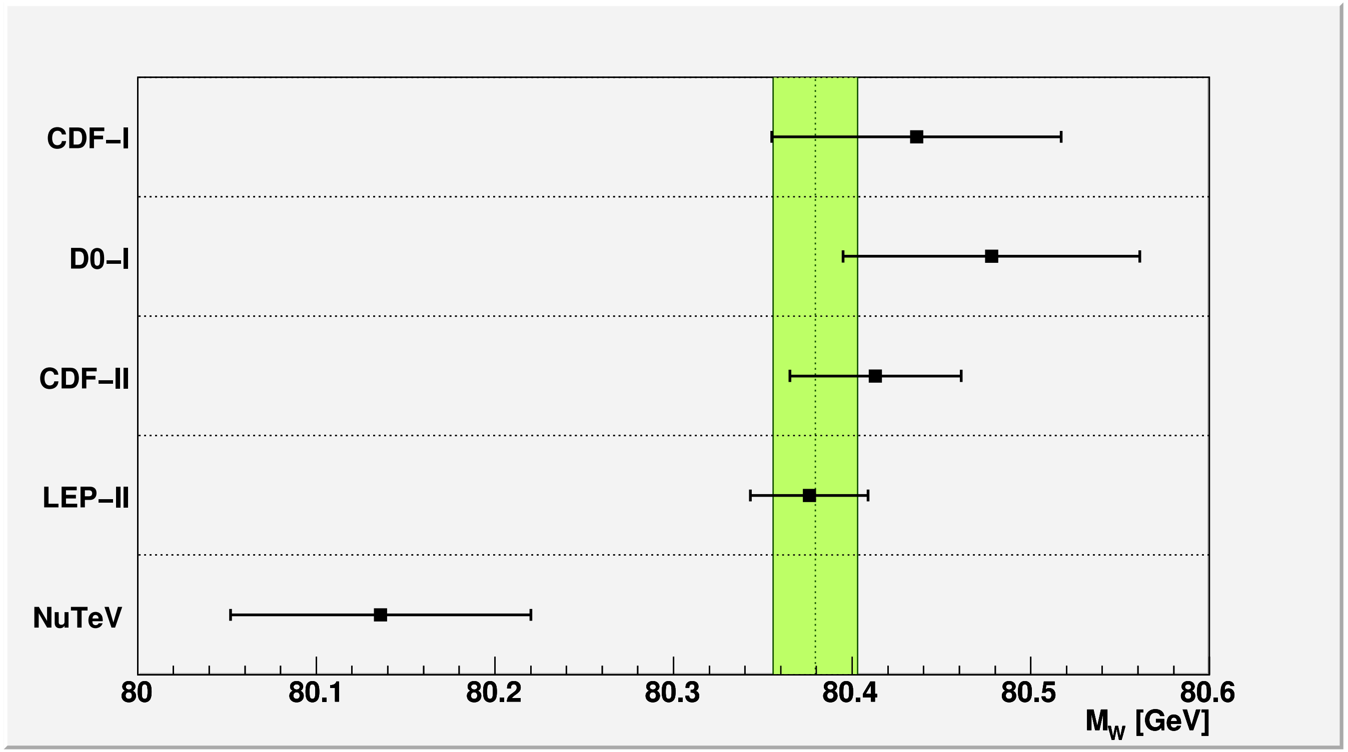
General: n-measurements with one unknown a $\rightarrow$ follows χ^2 distribution with $n-1$ degrees of freedom

Mini-exercise Plot χ^2 curves with computer



see $/afs/desy.de/user/m/mgoebel/public/StatisticsWS/ChiSquareDistribution.C$

Mini-exercise Outlier rejection



see $/afs/desy.de/user/m/mgoebel/public/StatisticsWS/AverageMW.C$

Tasks: Determine weighted average, the χ^2_{min} and fit probability, see what happens if one takes out the value with the largest deviation from the mean, etc.

Lecture III on linear least square fits

Contents:

- General solution (normal equations) of linear least square fits
- Straight line fits ... and other fit examples

Solution via normal equations

Linear fit $\longrightarrow$ determines estimator for $\vec{a}$

$$\chi^2 = (\vec{y} - A\vec{a})^t V^{-1} (\vec{y} - A\vec{a})$$

$$Min.\chi^2 \;\longrightarrow\; \frac{d\chi^2}{d\vec{a}^t} = 0$$

$$\longrightarrow\; -2A^t V^{-1}(\vec{y} - A\vec{a}) = 0$$

$$\longrightarrow\; A^t V^{-1} A\vec{a} = A^t V^{-1}\vec{y}$$

Side remark: why can't we just set: $\vec{y} = A\vec{a}$??? Because $\vec{y}$ has dimension n and $\vec{a}$ has dimension m !!! Solution:

$$\boxed{\begin{array}{rcl} \vec{a} & = & (A^t V^{-1} A)^{-1} A^t V^{-1} \vec{y} \\ & = & H^{-1} A^t V^{-1} \vec{y} \\ & = & U A^t V^{-1} \vec{y} \\ & & \text{with } U = H^{-1} = Cov(a) \end{array}}$$

Normal equations

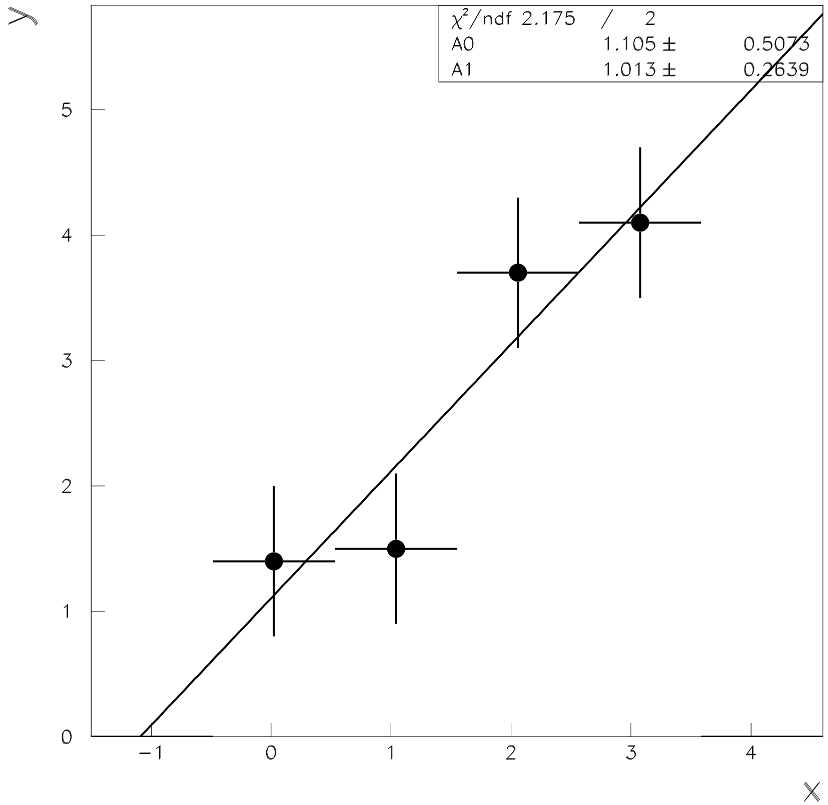
Straight line fit

Model: $\boldsymbol{y = a_0 + a_1 x}$

Simple example: All y_i have the same uncertainty and are uncorrelated

$$V = \begin{pmatrix} \sigma^2 & & 0 \\ & . & \\ 0 & & \sigma^2 \end{pmatrix}$$

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - a_0 - a_1\, x_i)^2}{\sigma^2}$$



Matrix notation:

$$\vec{y} = A\vec{a}; \qquad \vec{a}\begin{pmatrix} a_0 \\ a_1 \end{pmatrix} \quad A = \begin{pmatrix} 1 & x_1 \\ . & . \\ 1 & x_n \end{pmatrix}, \quad A^T = \begin{pmatrix} 1 & ... & 1 \\ x_1 & ... & x_n \end{pmatrix}$$

$$\hat{\vec{a}} \;=\; \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \end{pmatrix} = (A^T V^{-1} A)^{-1} A^T V^{-1} \vec{y}; \qquad V^{-1} = \begin{pmatrix} 1/\sigma^2 & & 0 \\ & . & \\ 0 & & 1/\sigma^2 \end{pmatrix}$$

Straight line fit

$$
\begin{aligned}
\hat{\vec{a}} &= \sigma^2 (A^T A)^{-1} \cdot \frac{1}{\sigma^2} A^T \cdot \vec{y} \;=\; (A^T A)^{-1} A^T \cdot \vec{y} \\
&= \begin{pmatrix} \sum_i 1 & \sum_i x_i \\ . & . \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix} = \begin{pmatrix} N & N\overline{x} \\ N\overline{x} & N\overline{x^2} \end{pmatrix}^{-1} \cdot \begin{pmatrix} N\overline{y} \\ N\overline{xy} \end{pmatrix} \\
&= \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \overline{y} \\ \overline{xy} \end{pmatrix} = \frac{1}{\overline{x^2} - \overline{x}^2} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix} \begin{pmatrix} \overline{y} \\ \overline{xy} \end{pmatrix} = \frac{1}{V[x]} \cdot \begin{pmatrix} \overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy} \\ -\overline{x}\,\overline{y} + \overline{xy} \end{pmatrix}
\end{aligned}
$$

Covariance matrix:

$$
\boldsymbol{U} = \begin{pmatrix} \sigma_{a_0}^2 & cov(a_0, a_1) \\ cov(a_0, a_1) & \sigma_{a_1}^2 \end{pmatrix} = (A^t V^{-1} A)^{-1} = \boldsymbol{\frac{\sigma^2}{N V[x]} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}}
$$

Mini exercise: straight line track-fit with N detectors

The covariance formula

$$\begin{pmatrix} \sigma^2_{a_0} & cov(a_0, a_1) \\ cov(a_0, a_1) & \sigma^2_{a_1} \end{pmatrix} = \frac{\sigma^2}{NV[x]} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$$

is valid for e.g. a straight line track fit in N detectors of resolution σ:

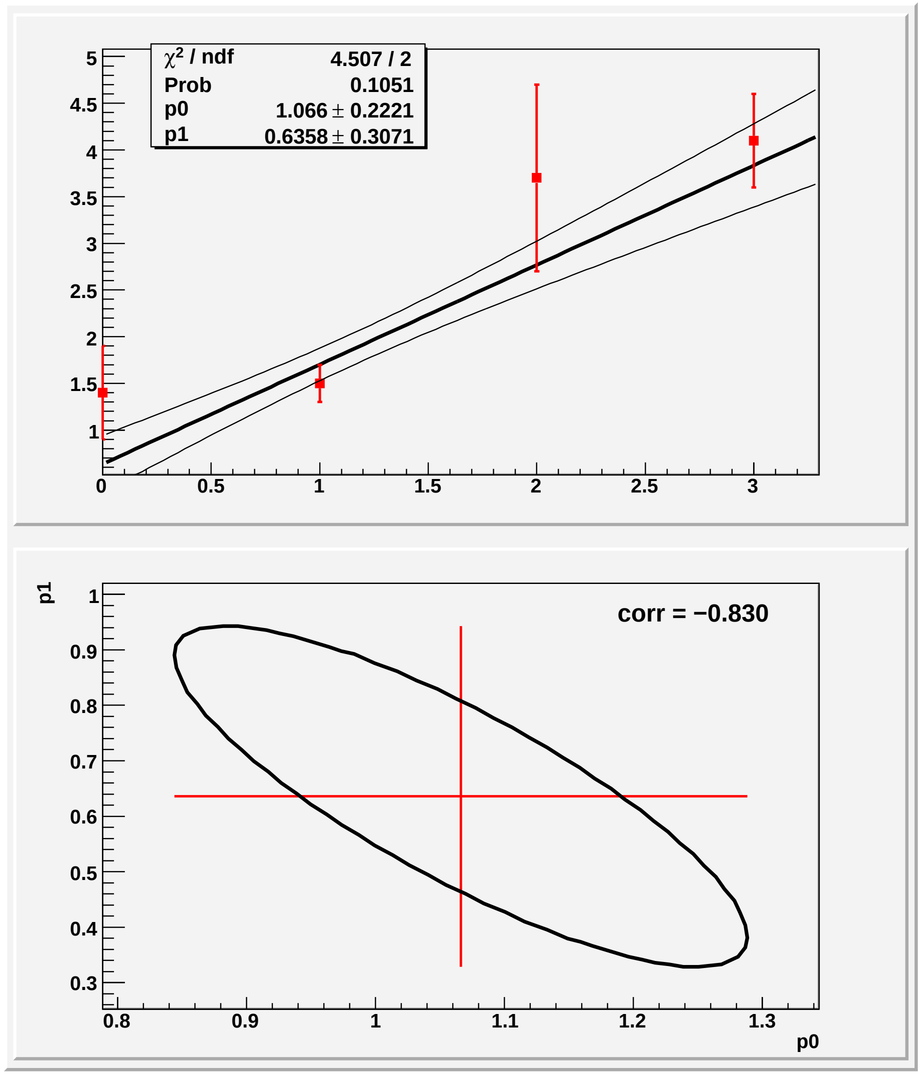
Determine the improvements on the slope error σ_{a_1} by:

a) **Doubling the number of detector layers within the same interval in x**

b) **Distributing the detector layers over an interval in x twice as large**

c) **Buying detectors with measurement uncertainties reduced by a factor two**

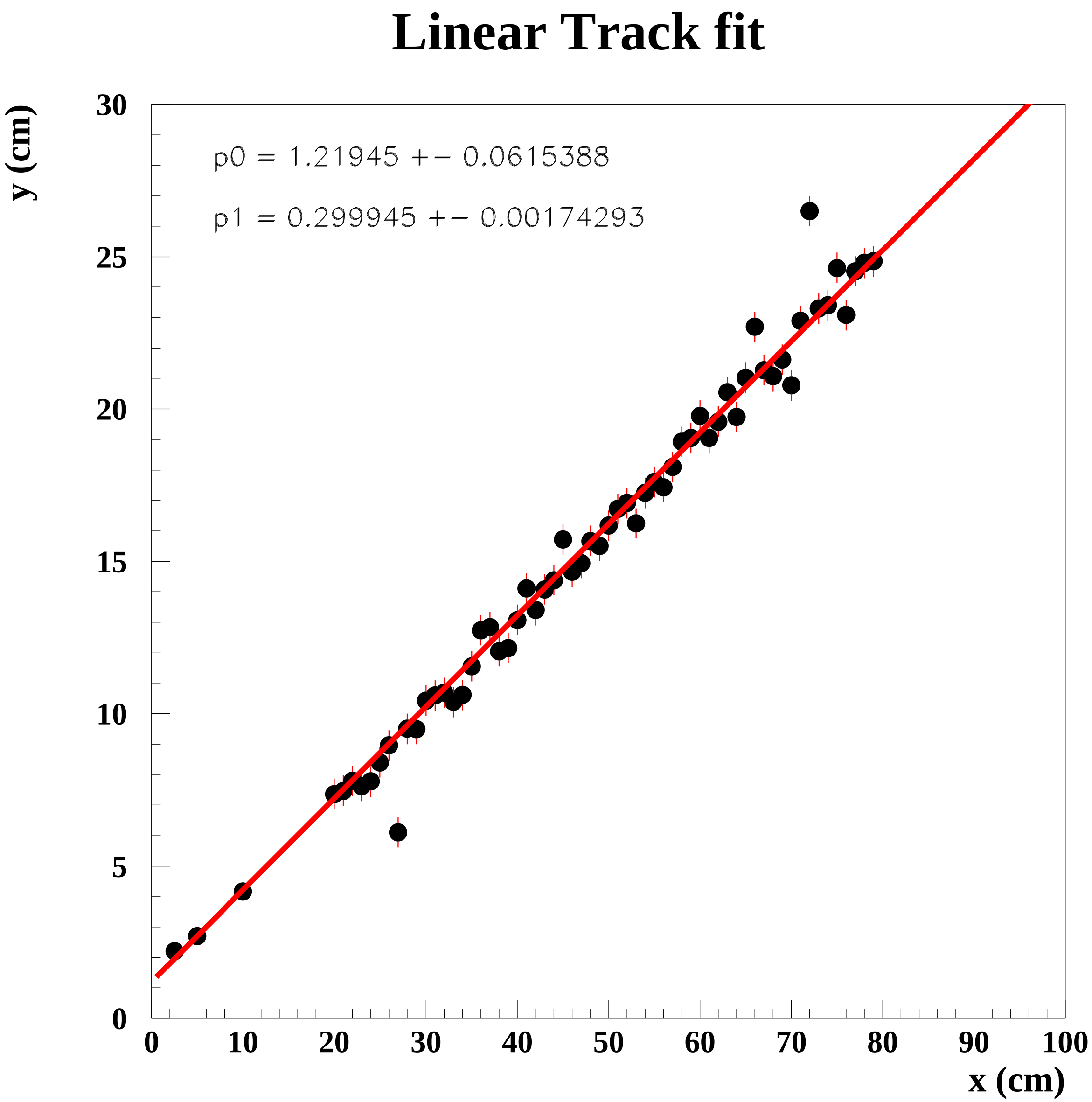
Mini-exercise Straight-line fit example



Tasks: Do the fit, plot the error ellipse, get the 1-sigma band, judge fit quality, etc.

see $/afs/desy.de/user/m/mgoebel/public/StatisticsWS/StraightLineFit.C$

Mini-exercise Real Trackfit in Si + driftchamber



Tasks: Do the straight line fit, plot the error ellipse, get the 1-sigma band, judge fit quality, reject outliers!, fit with parabola, determine momentum and error, extension: add vertex constraint