Parameter Estimation Sessions I+II : Linear Least Square Fits

Terascale Statistics Tools School

Mar 23-26 2010, DESY Hamburg

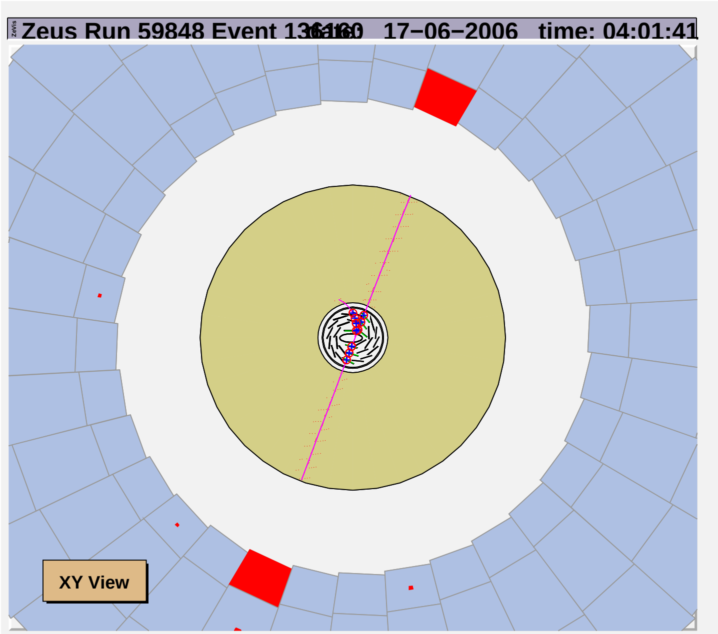
Olaf Behnke, DESY

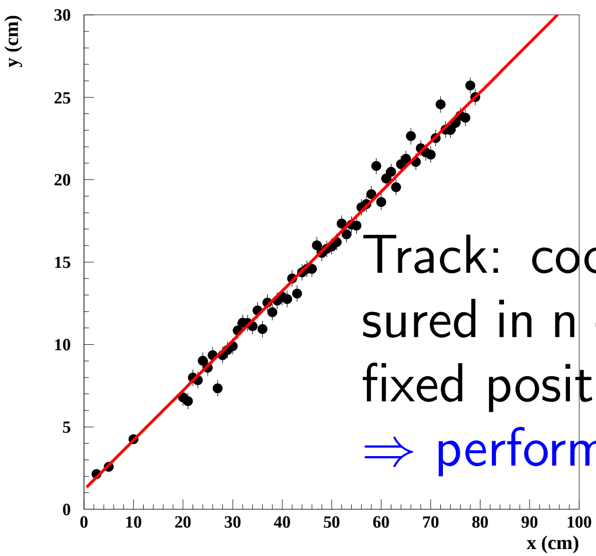
Literature:

- Roger Barlow: "Statistics, A Guide To The Use Of Statistical Methods In The Physical Sciences" Wiley & Sons, 1994
- Jay Orear: "Notes on Statistics for Physicists, Revised", 1958, *http://www.astro.washington.edu/users/ivezic/Astr507/orear.pdf*

Introductory Track fit example

Example: for possible discovery
$Z' \to \mu^+\mu^-$ need precise muon track fits

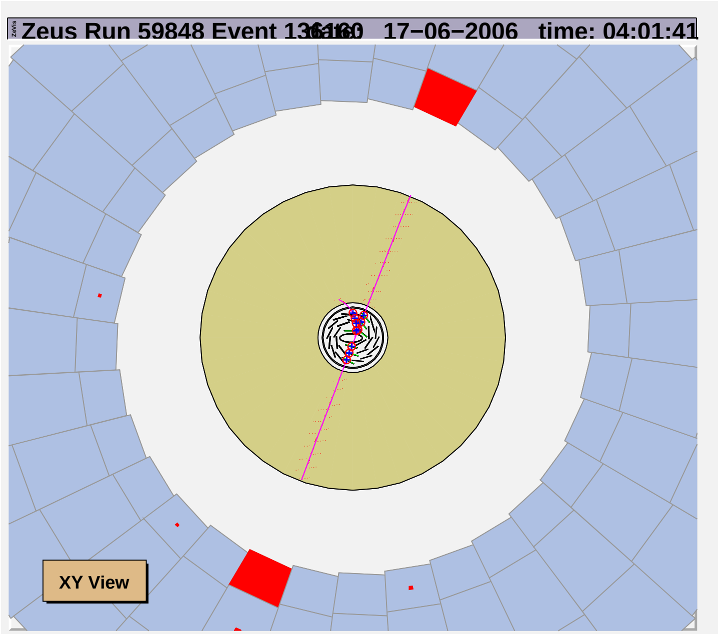


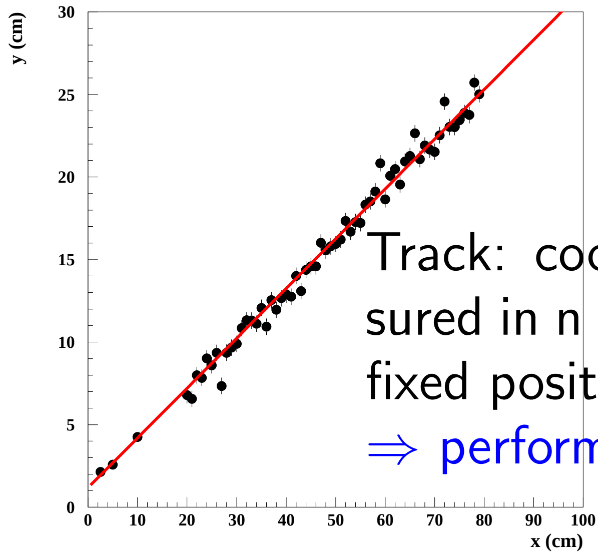


Track: coordinates y_i measured in n detector layers at fixed positions x_i
$\Rightarrow$ perform track fit

Introductory Track fit example

Example: for possible discovery $Z' \to \mu^+\mu^-$ need precise muon track fits





Track: coordinates y_i measured in n detector layers at fixed positions x_i

$\Rightarrow$ perform track fit

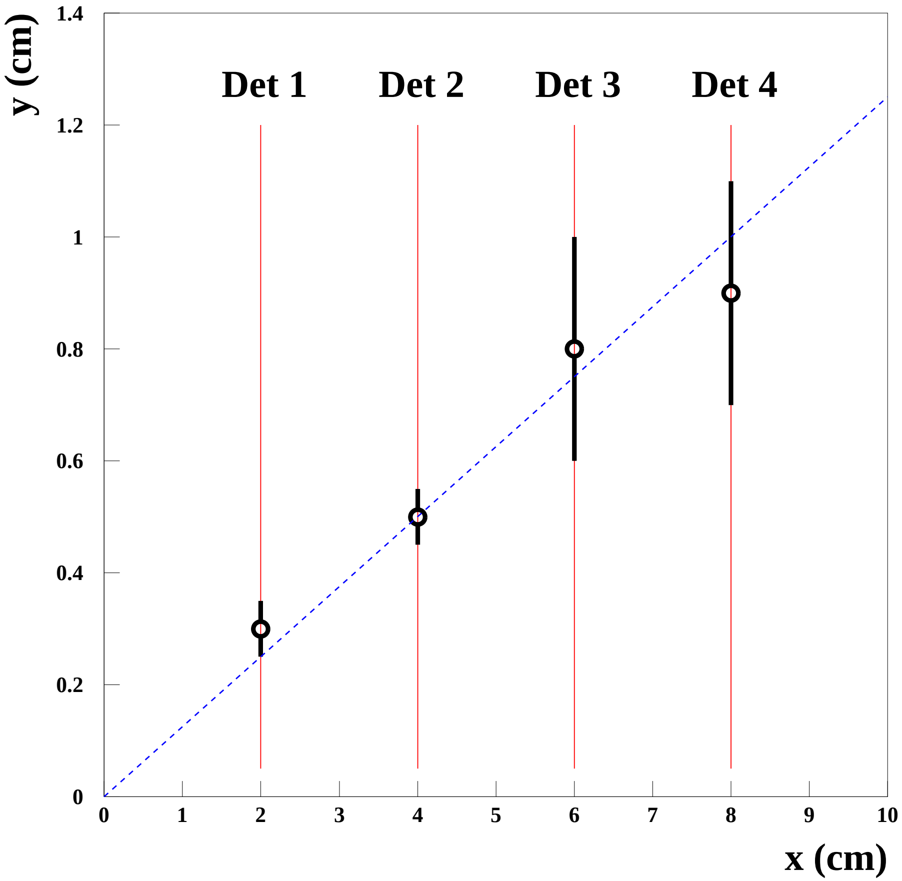
- Typical Assumptions:
  - Measurements with gaussian uncertainties
  - Linear(ized) model, here: $y = a_0 + a_1 x + a_2 x^2$ *(but could also use exact track helix model)*
- Construct χ^2:
  - $\chi^2 = \sum_i \frac{\left[y_i - (a_0 + a_1 x + a_2 x^2)\right]^2}{\sigma_i^2}$
  - Determine a_0, a_1, a_2 by finding χ^2 minimum (normal equations)
- Check consistency:
  - use χ^2 and χ^2-fit probability
  - reject outliers
- Analyse results:
  - parameters, errors and correlations (error ellipses), track trajectory error band
  - calculate momentum (error propagation)

Lecture Part 1

- Least square χ^2-fit method introduction
- Fit of a constant
- χ^2_{min} as consistency check

Method of least squares fit - Intro

Example: Particle trajectory measurement



n-measurements $y_i \pm \sigma_i$ at fixed x_i

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

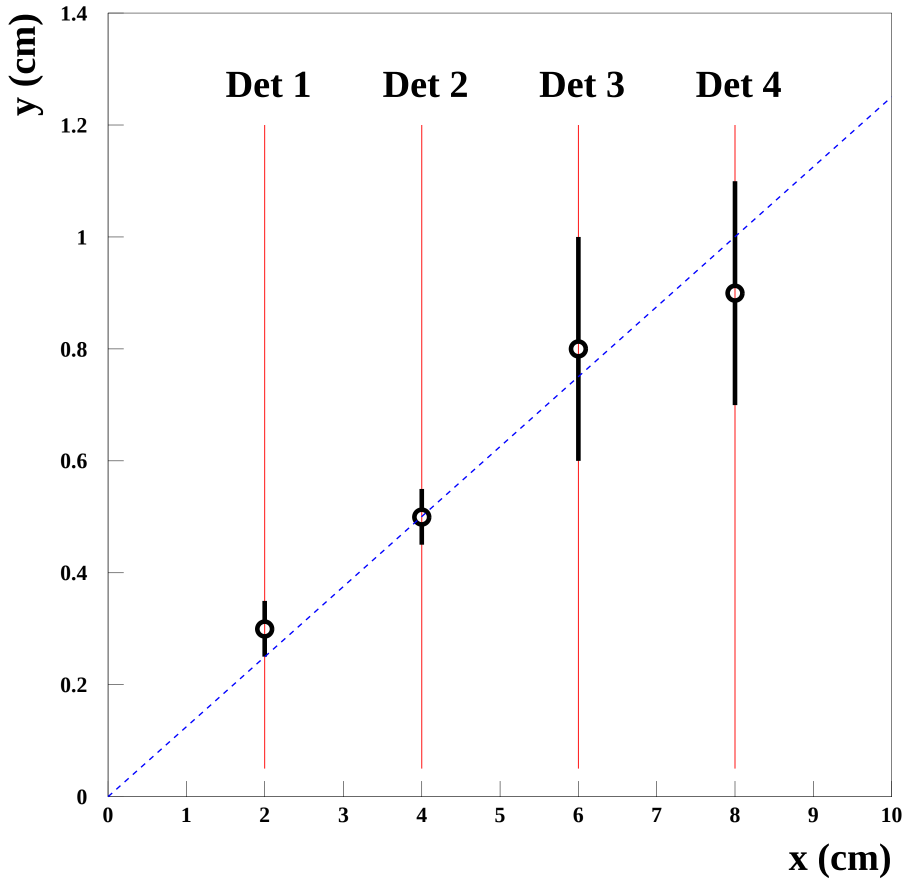
here: $y = ax$

$\Rightarrow$ how to determine a?

$\Rightarrow$

Method of least squares fit - Intro

Example: Particle trajectory measurement



n-measurements $y_i \pm \sigma_i$ at fixed x_i

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

here: $y = ax$

$\Rightarrow$ how to determine a?

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

Method of least squares fit - Intro

$\rightarrow$ $\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f(x_i, a))^2}{\sigma_i^2}}$ $\leftrightarrow$ Minimum w.r.t a

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

$\Rightarrow$

Method of least squares fit - Intro

$$\rightarrow \boxed{\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f(x_i, a))^2}{\sigma_i^2}} \leftrightarrow \text{Minimum w.r.t a}$$

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

$$\Rightarrow \boxed{\frac{d\chi^2}{da}_{|a=\hat{a}} = 2 \cdot \sum_{i=1}^{n} \frac{(y_i - f(x_i, a))}{\sigma_i^2} \cdot \frac{df(x_i, a)}{da} = 0}$$

In general not analytically solvable.

$\Rightarrow$ use iterative (numerical) methods (MINUIT, Mathematica)

Method of least squares fit

Most general case

- y_i, y_j correlated measurem. with cov. V_{ij}
- m fitparameters $\vec{a}$

$$\rightarrow \boxed{\begin{aligned} \chi^2 &= \sum_{i,j=1}^{n} (y_i - f(x_i, \vec{a})) V_{ij}^{-1} (y_j - f(x_j, \vec{a})) \\ &= \end{aligned}}$$

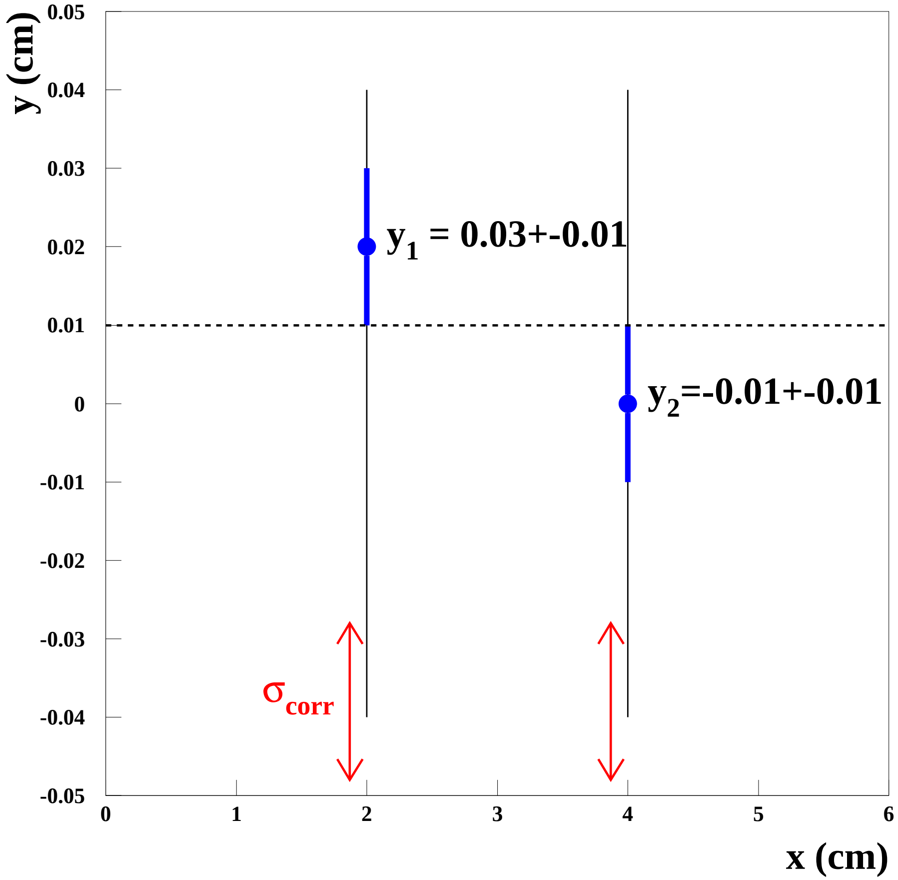
Method of least squares fit

Most general case

- y_i, y_j correlated measurem. with cov. V_{ij}
- m fitparameters $\vec{a}$

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

Example for two correlated measurements

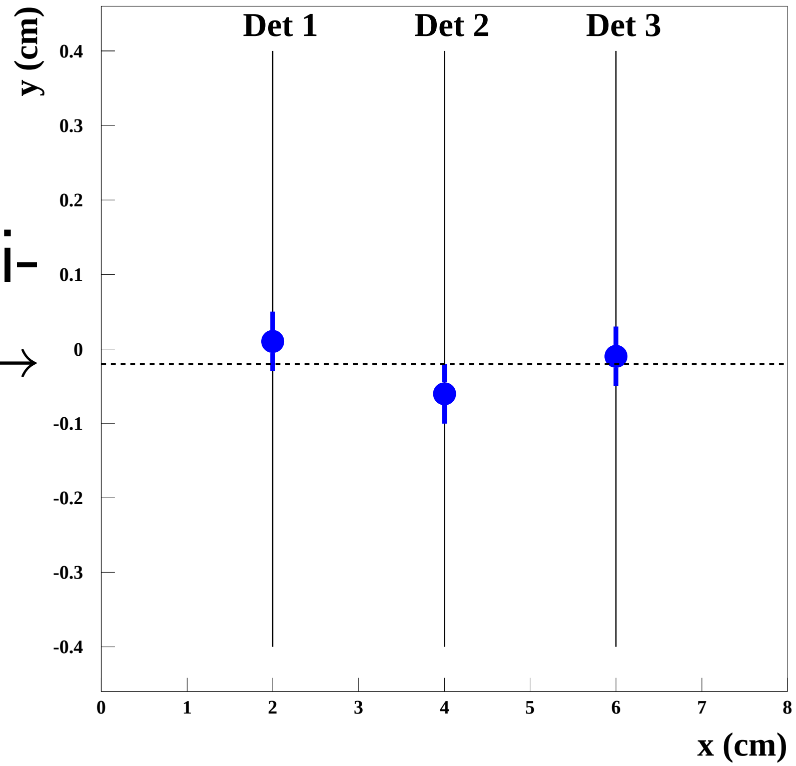


Measure track in two detector layers with global position uncertainty

$$V = \begin{pmatrix} 0.01^2 + \sigma^2_{corr} & \sigma^2_{corr} \\ \sigma^2_{corr} & 0.01^2 + \sigma^2_{corr} \end{pmatrix}$$

Fit of a constant

Measure position of horizontally flying particle $\longrightarrow$

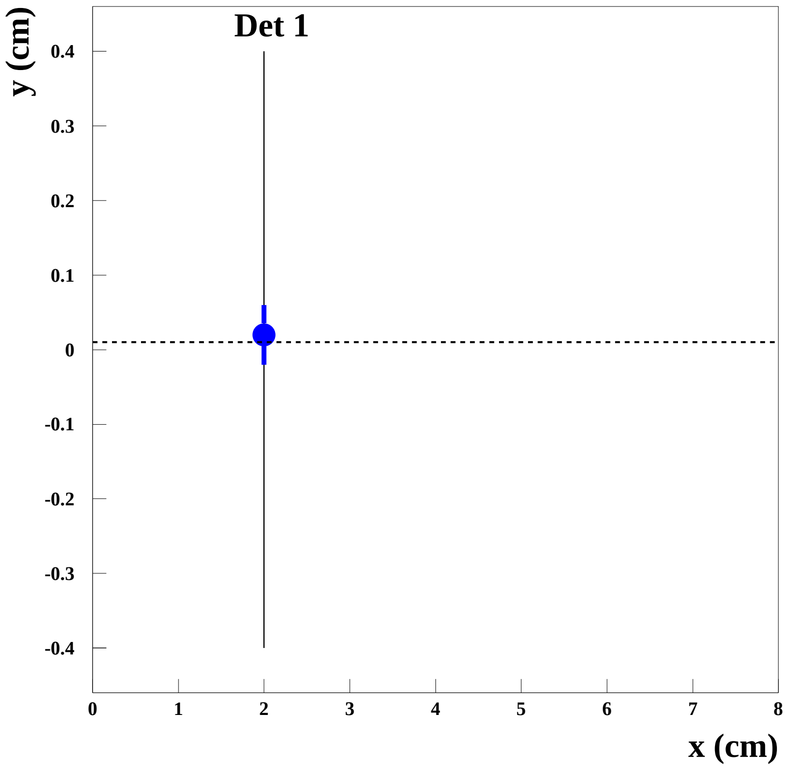


$\rightarrow$ Averaging of n measurements $y_i \pm \sigma_i$

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

Fit of a constant (one measurement)

"Idiot example" of one measurement $y_1 \pm \sigma_1$:



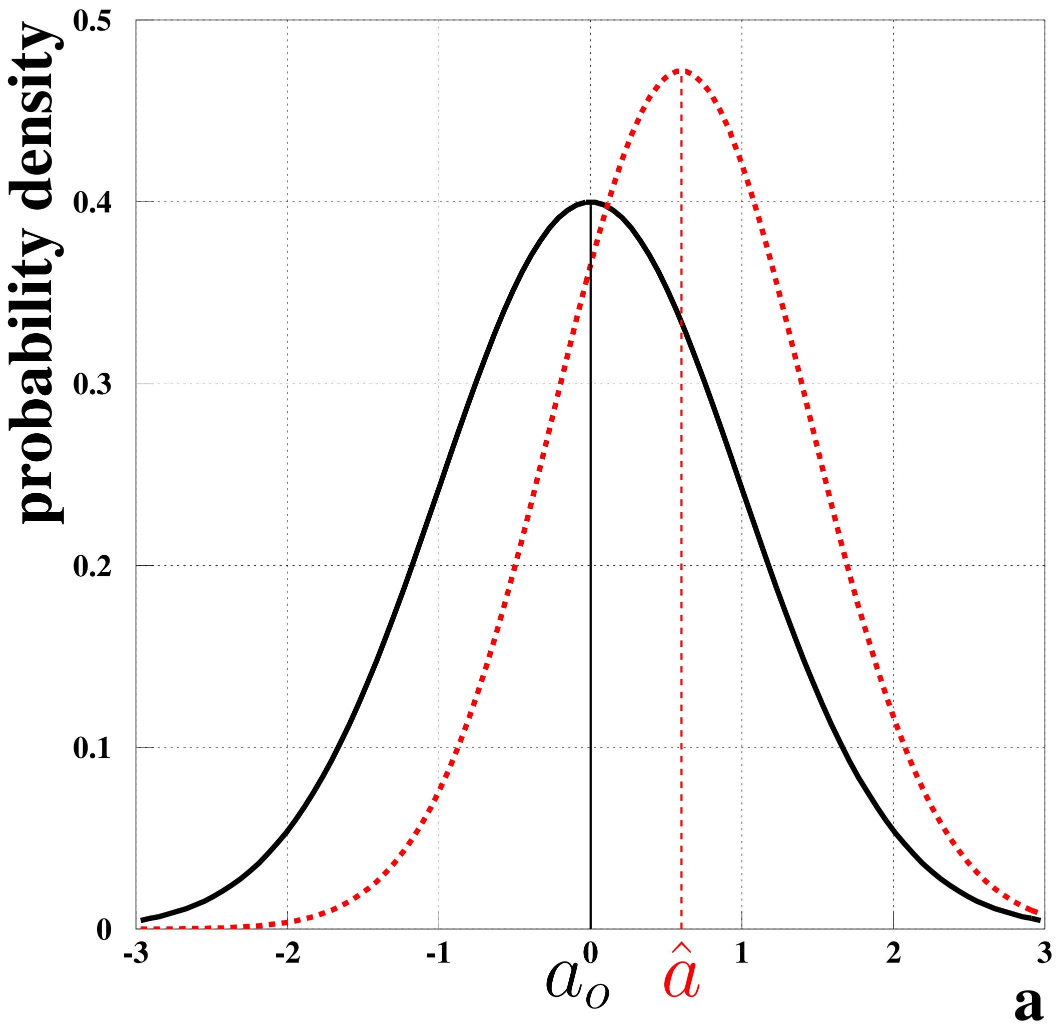
$$\chi^2 = \frac{(y_1 - a)^2}{\sigma_1^2}$$

$$Min.\chi^2 : \quad \frac{d\chi^2}{da} = 0$$

$\longrightarrow$ Estimated value: $\hat{a} = y_1$

$\longrightarrow$ Error propagation: $\sigma_{\hat{a}} = \sigma_1$

True and inverse probability densities for one measurement with gaussian uncertainty: $\hat{a} = y_1$, $\sigma_{\hat{a}} = \sigma_1$



True probability density to observe $\hat{a}$ for given true value a_o:

$$p = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(\hat{a}-a_o)^2}{2\sigma^2}}$$

But what if we don't know a_0?

Estimate "inverse probability density" for a_o from the measurement $\hat{a} \pm \sigma_{\hat{a}}$:

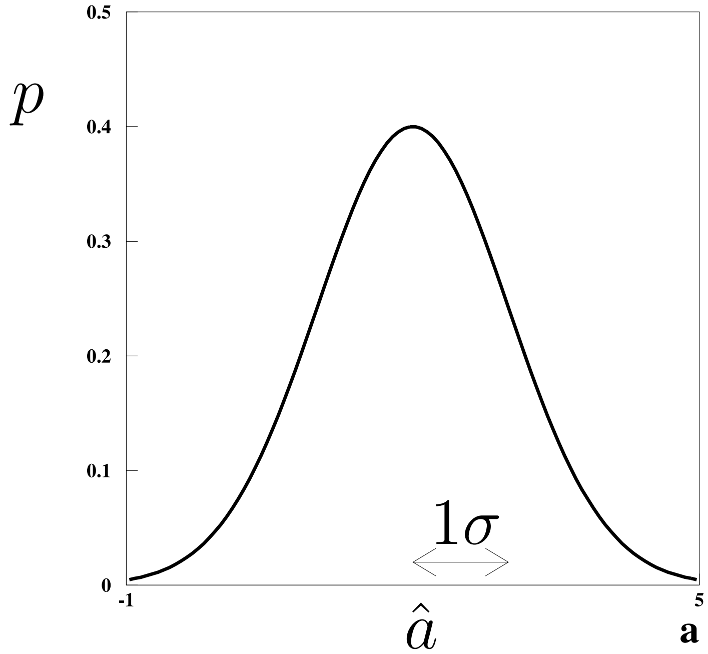
$$p = \frac{1}{\sqrt{2\pi}\sigma_{\hat{a}}} \cdot e^{-\frac{(\hat{a}-a_o)^2}{2\sigma_{\hat{a}}^2}}$$

Note: this is not a real prob. density!

from now on we will use a as synonym for a_0!

Fit of a constant (one measurement)

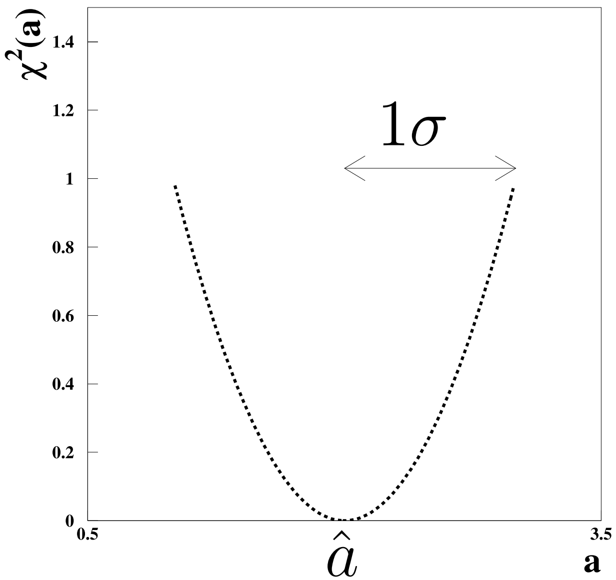
Inverse probability density for true a:



$$p \sim e^{-\frac{(a-\hat{a})^2}{2\sigma_{\hat{a}}^2}}$$

with $\chi^2 = \frac{(a-\hat{a})^2}{\sigma_{\hat{a}}^2} \Rightarrow p \sim e^{-\chi^2/2}$

$$\Rightarrow \frac{1}{\sigma_{\hat{a}}^2} = \frac{1}{2}\frac{d^2\chi^2}{da^2}_{|a=\hat{a}}$$



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

Note: These two relations hold for a large class of one parameter χ^2-fit-problems!

Fit of a constant - many measurements

Probability for true value a to observe measurements y_i, with $i = 1, n$:

$$p(y_1, y_2, ..., y_n|a) \propto \prod_{i=1}^{n} e^{-\frac{(y_i-a)^2}{2\sigma_i^2}}$$

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

but we don't know true a,

so *let's turn the whole thing around* to estimate probability density for true a from the measurements

Fit of a constant - many measurements

Recalling $p(y_1, y_2, ..., y_n|a) = e^{-\chi^2/2}$

Expand χ^2 around its minimum at $\hat{a}$:

$$\chi^2 = \chi^2(\hat{a}) + \underbrace{\frac{d\chi^2}{da}_{|a=\hat{a}}}_{=0} \cdot (a - \hat{a}) + \frac{1}{2}\frac{d^2\chi^2}{da^2}_{|a=\hat{a}} \cdot (a - \hat{a})^2$$

Fit of a constant - many measurements

Recalling $p(y_1, y_2, ..., y_n|a) = e^{-\chi^2/2}$

Expand χ^2 around its minimum at $\hat{a}$:

$$\chi^2 = \chi^2(\hat{a}) + \underbrace{\frac{d\chi^2}{da}_{|a=\hat{a}}}_{=0} \cdot (a - \hat{a}) + \frac{1}{2}\frac{d^2\chi^2}{da^2}_{|a=\hat{a}} \cdot (a - \hat{a})^2$$

$= \chi^2(\hat{a}) + H \cdot (a - \hat{a})^2$ with $H = \frac{1}{2}\frac{d^2\chi^2}{da^2}_{|a=\hat{a}}$ 'Hesse matrix' (for one par. a number)

$$\Rightarrow \quad p(y_1, y_2, ..., y_n|a) \propto \underbrace{e^{-\frac{\chi^2(\hat{a})}{2}}}_{\text{Fit consistency}} \cdot \underbrace{e^{-\frac{1}{2}H \cdot (\hat{a}-a)^2}}_{\text{gaussian density}}$$

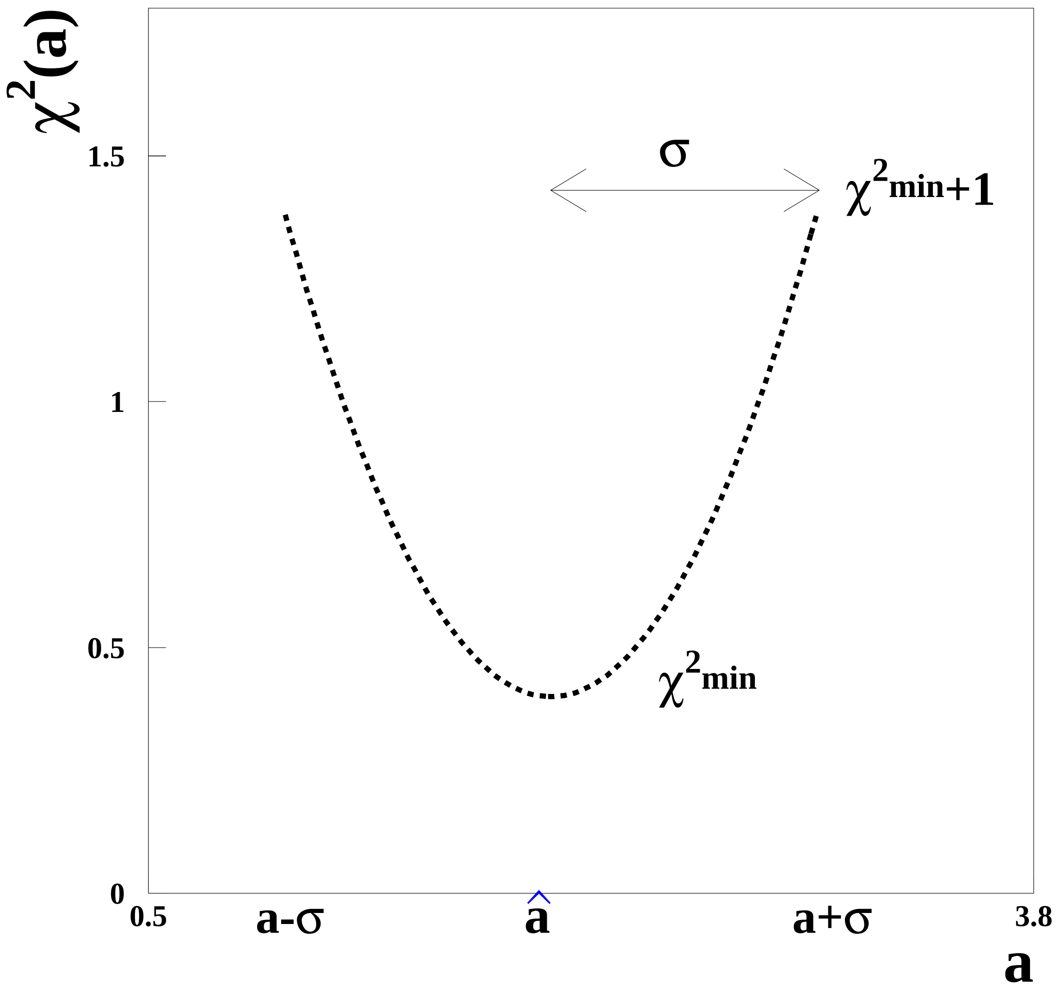
$\Rightarrow$ *interpreted as inverse probability density for true a:*

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

Generalisation to any one-parameter (linear) fit

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

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



→ Read error directly from χ^2 curve

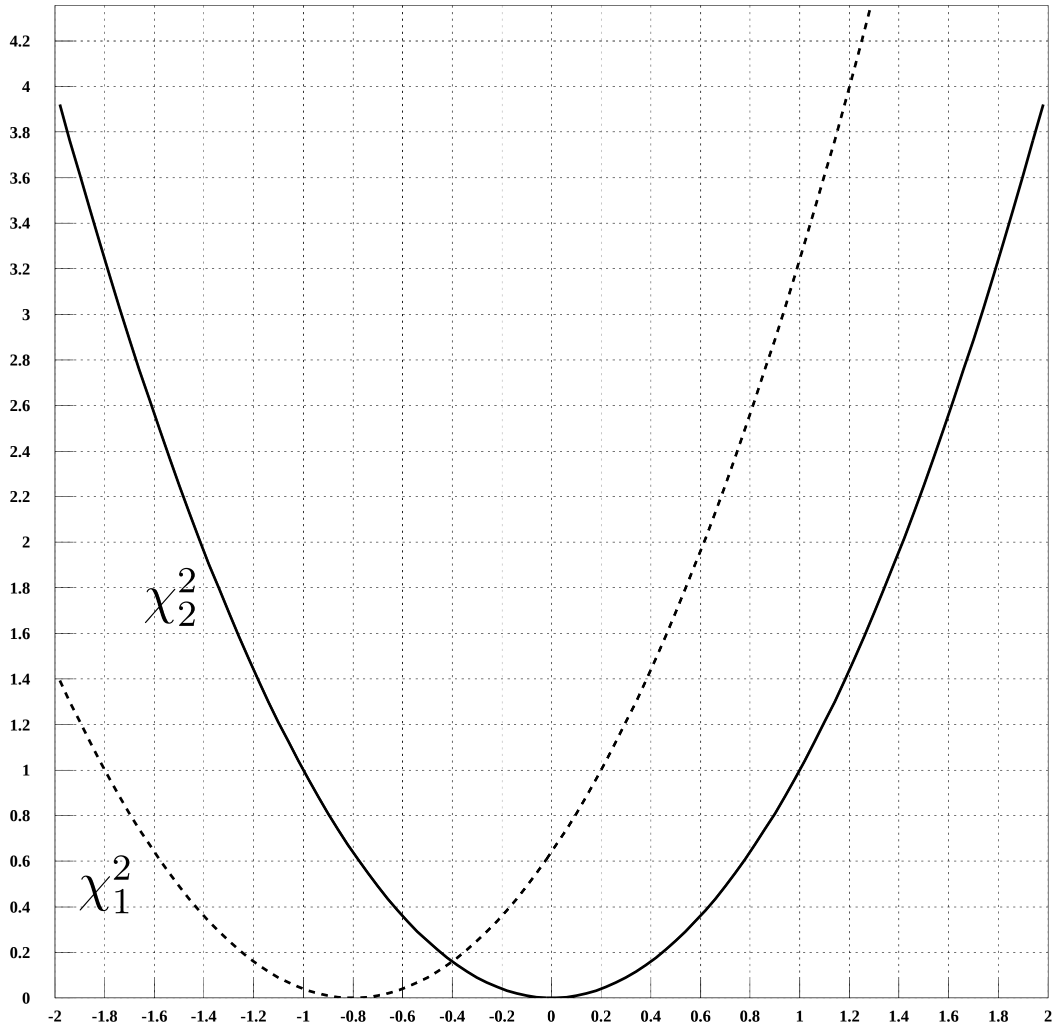
Mini-exercise Averaging of two meas. via χ^2 parabolas

Two measurements y_1 and y_2 of the observable a are represented in the figure by χ^2 parabolas:

$\chi_i^2 = (y_i - a)^2/\sigma_i^2; \quad i = 1, 2$

- Determine (yes by eye!) from the two χ^2 curves the values y_1, σ_1 and y_2, σ_2
- Draw the total χ^2, i.e. the sum of the two parabolas (yes, do it by hand :-)) and determine $\hat{a}$ and $\sigma_{\hat{a}}$

  (use χ^2_{min} and $\chi^2 = \chi^2_{min} + 1$)
- How much is the error $\sigma_{\hat{a}}$ reduced compared to σ_1 and σ_2?
- Relax your eyes and hands ;-)

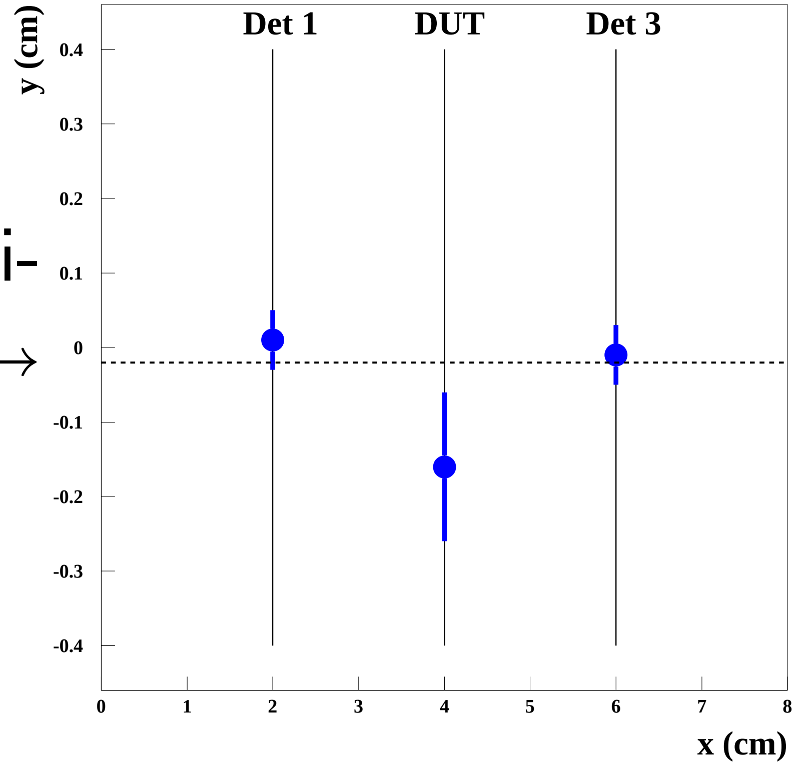


Averaging several measurements

n measurements $y_i \pm \sigma_i$ (Note: $\sigma_1 \neq \sigma_2$, etc.)

(Quiz question: Why is $\frac{1}{n}\Sigma y_i$ not the best average?

Measure position of horizontally flying particle $\longrightarrow$



Averaging several measurements

n measurements $y_i \pm \sigma_i$:

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

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

Averaging several measurements

n measurements $y_i \pm \sigma_i$:

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

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

$$\rightarrow \quad \boxed{\hat{a} \; = \; \sum_{i=1}^{n} \left[\frac{y_i}{\sigma_i^2}\right] / \sum_{i=1}^{n} \left[\frac{1}{\sigma_i^2}\right]}$$

$$\frac{1}{\sigma_{\hat{a}}^2} = \frac{1}{2} \frac{d^2\chi^2}{da^2} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}$$

Averaging - just reformulated

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

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

$$\rightarrow \quad \boxed{\hat{a} = \frac{1}{\sum_{i=1}^{n} G_i} \cdot \sum_{i=1}^{n} G_i y_i = \frac{1}{G_s} \cdot \sum_{i=1}^{n} G_i y_i}$$

$\sigma_{\hat{a}}$ from simple error propagation:

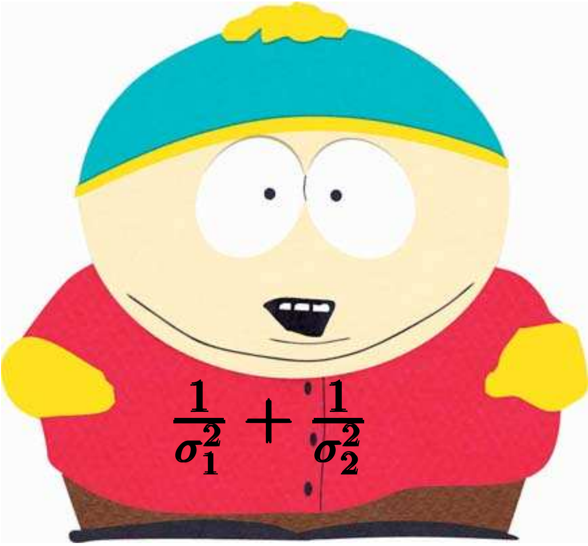
$$\boxed{\begin{aligned} \sigma_{\hat{a}}^2 &= \sum_{i=1}^{n} \left(\frac{d\hat{a}}{dy_i}\right)^2 \cdot \sigma_i^2 = \sum_{i=1}^{n} \left(\frac{G_i}{G_s}\right)^2 \cdot \sigma_i^2 \\ &= \frac{1}{G_s^2} \cdot \sum_{i=1}^{n} G_i = \frac{1}{G_s} = \frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2} \end{aligned}}$$

$\Rightarrow$ Corollar: least square fitting is nothing else than a clever mapping of measurements to the fitparameters and obtaining fitparameter uncertainties using error propagation

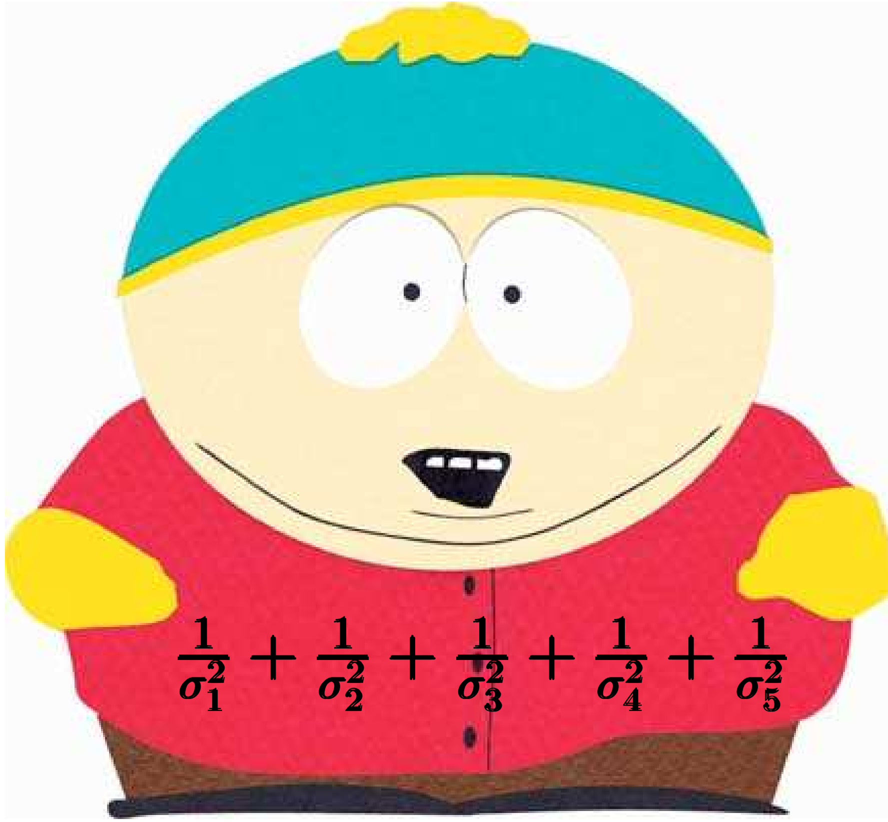
The role of the Hesse matrix

illustrated for weighted average (just a number)

$$H = \frac{1}{2}\frac{d^2\chi^2}{da^2} = \sum_{i=1}^{n}\frac{1}{\sigma_i^2}$$



H "grows" with each measurement $\longrightarrow$

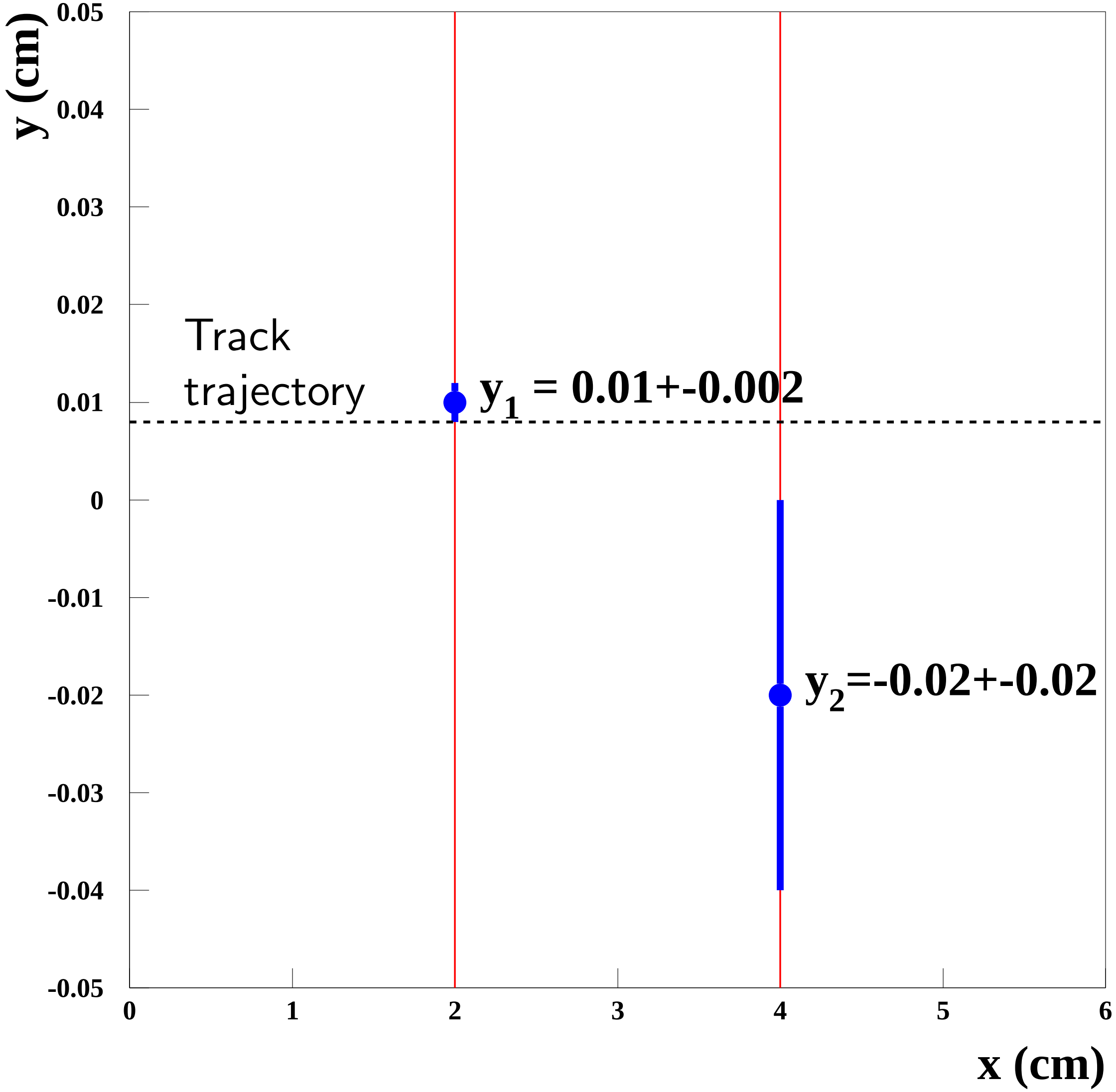


H is "counting the information" from the measurements

Finally $V = H^{-1}$

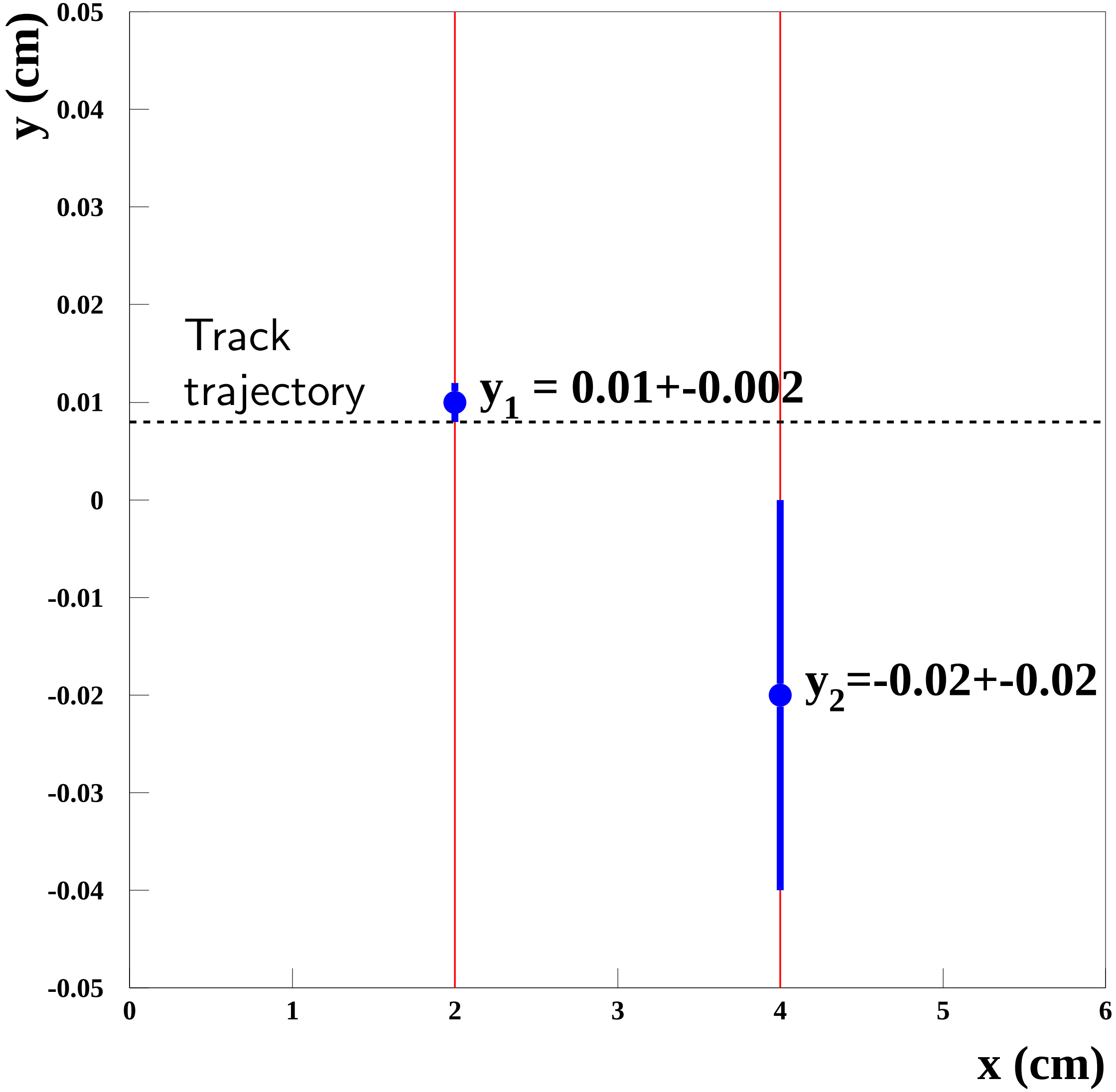
Note: all this holds also for fits with many parameters

Mini exercise weighted average



Weighted average of two measurements:

Mini exercise weighted average



Weighted average of two measurements:

$$\hat{y} = \frac{1}{\frac{1}{0.002^2} + \frac{1}{0.02^2}} \left(\frac{0.01}{0.002^2} + \frac{-0.02}{0.02^2} \right) = 0.0097$$

$$\sigma_{\hat{y}} = \sqrt{\frac{1}{\frac{1}{0.002^2} + \frac{1}{0.02^2}}} = 0.00199$$

Mini summary of what we have learnt

One parameter fits:

- **Least square expression for independent measurements:**
  $\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f(x_i,a))^2}{\sigma_i^2}$
  $\Rightarrow$ **get estimator** $\hat{a}$ **from minimum** $\chi^2 \Leftrightarrow d\chi^2/da_{|a=\hat{a}} = 0$
- **True physics parameters have a definite value, so true probability densities exist only for the measurements, fitting means estimating (inverse) probability densities for the true parameters**
- $\frac{1}{\sigma_{\hat{a}}^2} = \frac{1}{2}\frac{d^2\chi^2}{da^2}_{|a=\hat{a}}$ **(general relation)**
- $\chi^2(\hat{a} \pm \sigma_{\hat{a}}) = 1$ **(general relation)**
- **Averaging several measurements can be easily done graphically by adding individual** χ^2 **parabolas**
- **Least square fitting is nothing else than clever mapping of measurements to fitparameters; errors of fitparameters can be obtained from simple errorpropagation**
- **The Hesse matrix** $H = \frac{1}{2}\frac{d^2\chi^2}{da^2}_{|a=\hat{a}}$ **"counts the information" from the measurements**

Consistency of measurements

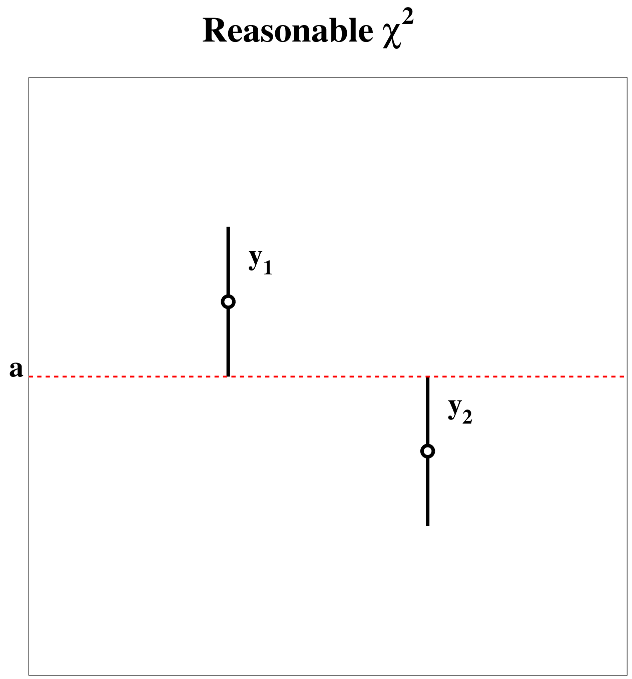
Recall "inverse probability density" for averaging n measurements:

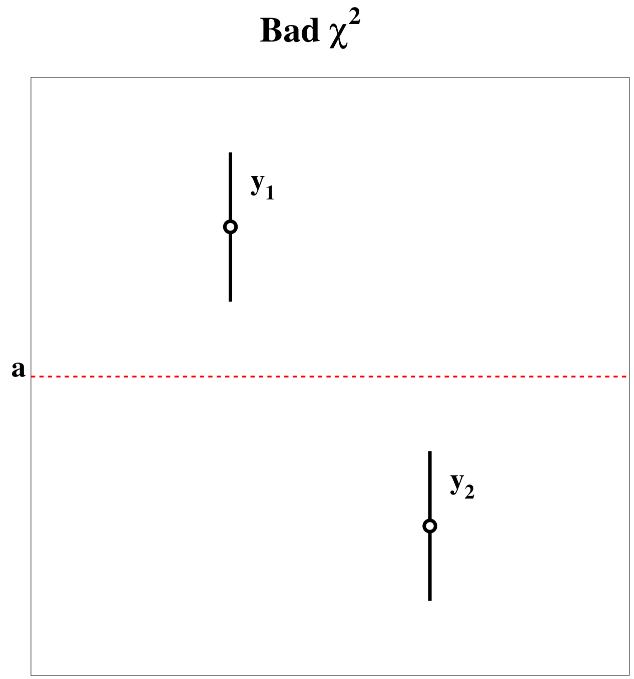
$$\Rightarrow p(y_1, y_2, ..., y_n|a) \propto \underbrace{e^{-\frac{\chi^2(\hat{a})}{2}}}_{\text{Fit consistency}} \cdot \underbrace{e^{-\frac{1}{2}H\cdot(\hat{a}-a)^2}}_{\text{gaussian density}}$$

Now lets have a closer look at the first term

Consistency of measurements

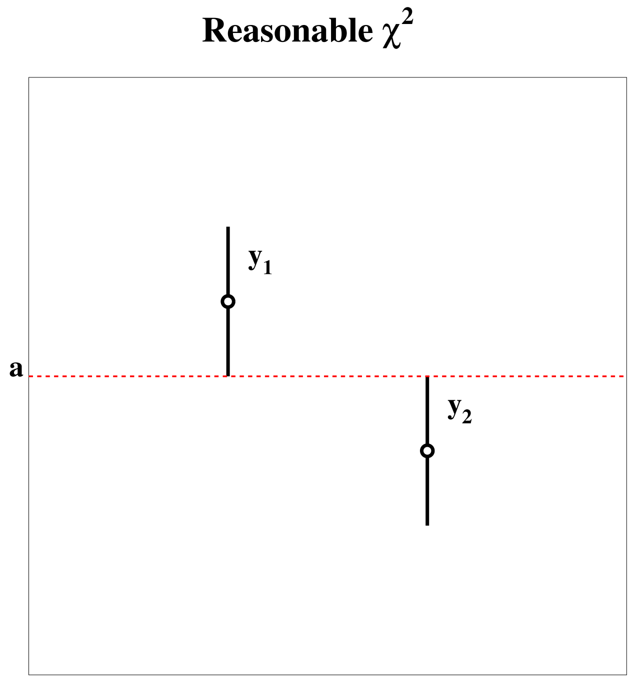
Example: Two measurements $y_1 \pm \sigma_1$ and $y_2 \pm \sigma_2$; the true value a be known, are the measurements consistent with a?:

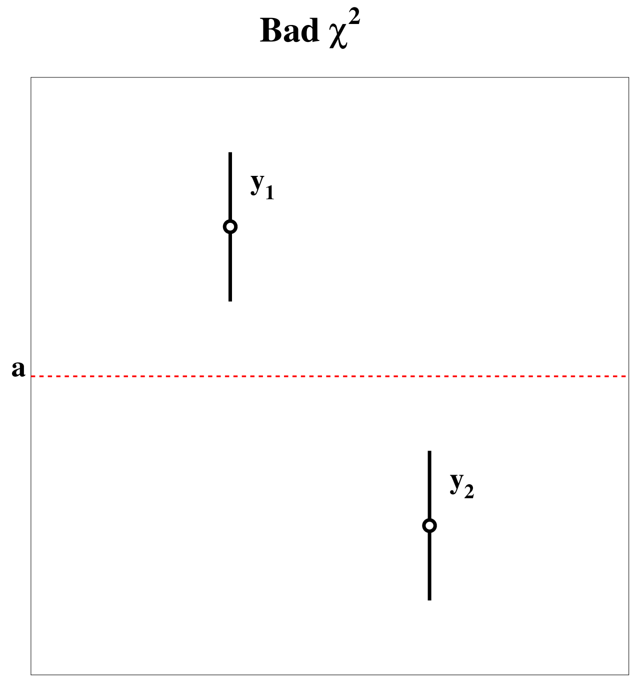




Consistency of measurements

Example: Two measurements $y_1 \pm \sigma_1$ and $y_2 \pm \sigma_2$; the true value a be known, are the measurements consistent with a?:





$$\chi^2 = 2 \qquad\qquad \chi^2 = 8$$

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

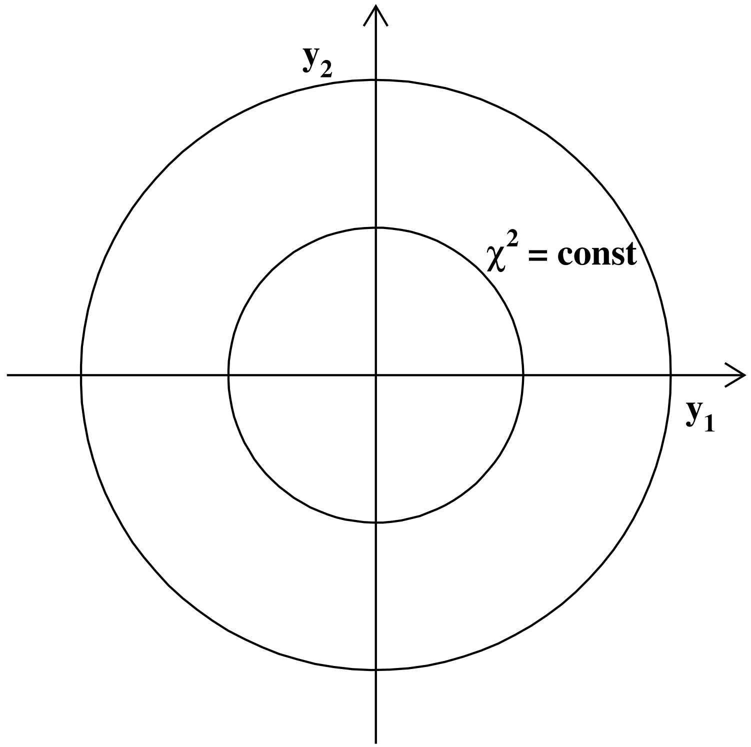
But how should χ^2 be distributed?

χ^2 for two measurements and known true value

Expected density for (y_1, y_2) (simple case $a = 0; \sigma_1 = \sigma_2 = 1$):

$$f(y_1, y_2) = \frac{1}{2\pi} e^{-y_1^2/2} e^{-y_2^2/2} = \frac{1}{2\pi} e^{-r^2/2}$$

with $r = \sqrt{y_1^2 + y_2^2} = \sqrt{\chi^2}$

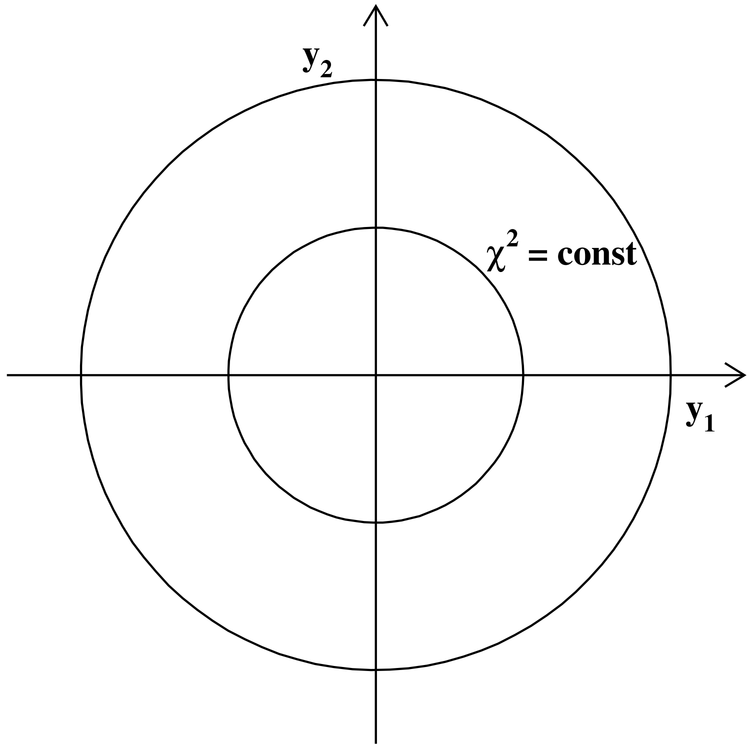


χ^2 for two measurements and known true value

Expected density for (y_1, y_2) (simple case $a = 0; \sigma_1 = \sigma_2 = 1$):

$$f(y_1, y_2) = \frac{1}{2\pi} e^{-y_1^2/2} e^{-y_2^2/2} = \frac{1}{2\pi} e^{-r^2/2}$$

with $r = \sqrt{y_1^2 + y_2^2} = \sqrt{\chi^2}$



Probability to find value between r and $r + dr$:

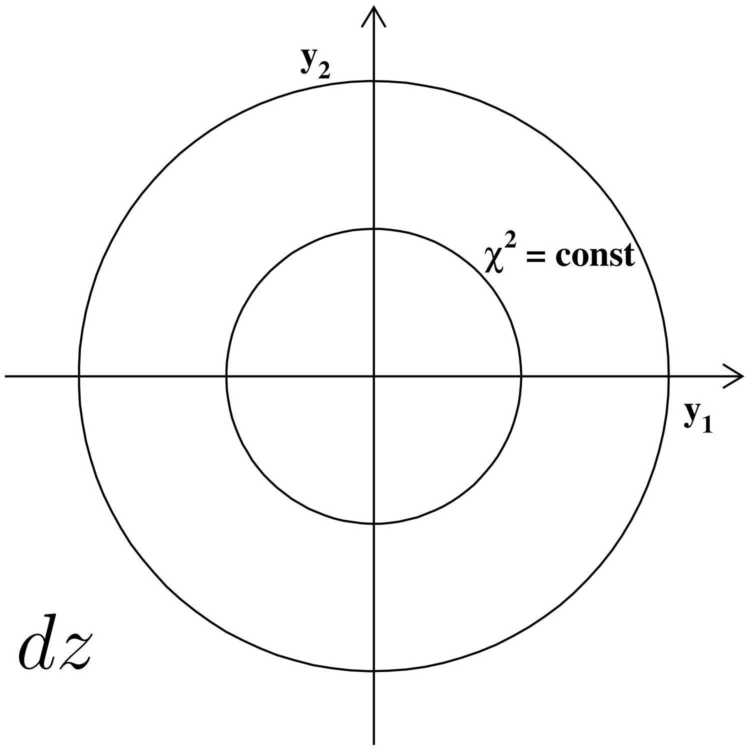
$$f(r)\,dr = \frac{2\pi r}{2\pi} e^{-r^2/2}\,dr = r e^{-r^2/2}\,dr$$

χ^2 for two measurements and known true value

Expected density for (y_1, y_2) (simple case $a = 0; \sigma_1 = \sigma_2 = 1$):

$$f(y_1, y_2) = \frac{1}{2\pi} e^{-y_1^2/2} e^{-y_2^2/2} = \frac{1}{2\pi} e^{-r^2/2}$$

with $r = \sqrt{y_1^2 + y_2^2} = \sqrt{\chi^2}$



Probability to find value between r and $r + dr$:

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

$$z = r^2 : \rightarrow f(z)\,dz = f(r)\frac{dr}{dz} dz = \frac{1}{2} e^{-z/2}\,dz$$

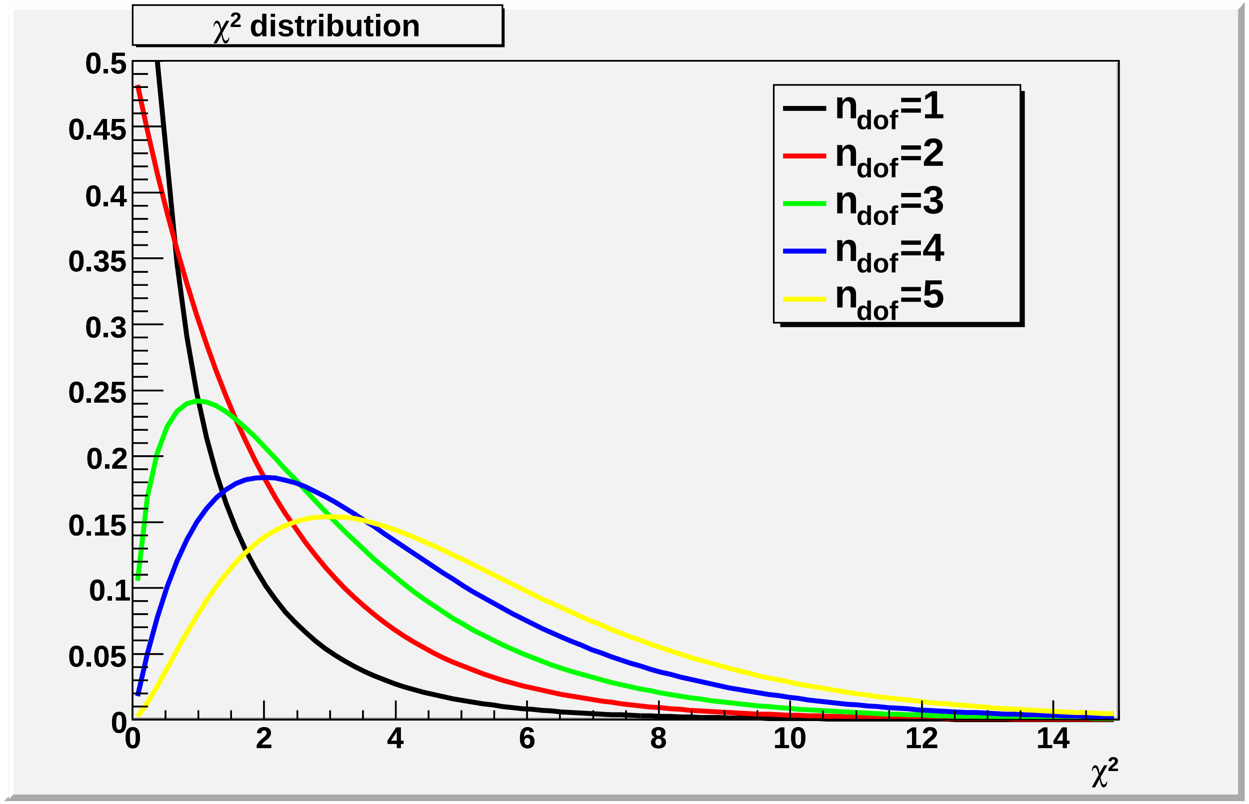
$\rightarrow$ introduces χ^2-distribution for $z = \chi^2$ and two dimensions (ndf=2): $f(z, 2) = \frac{1}{2} e^{-z/2}$

χ^2-function for n degrees of freedom

$\rightarrow$ maps the χ^2 in n dimensions into probability density for χ^2

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

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



Properties:

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

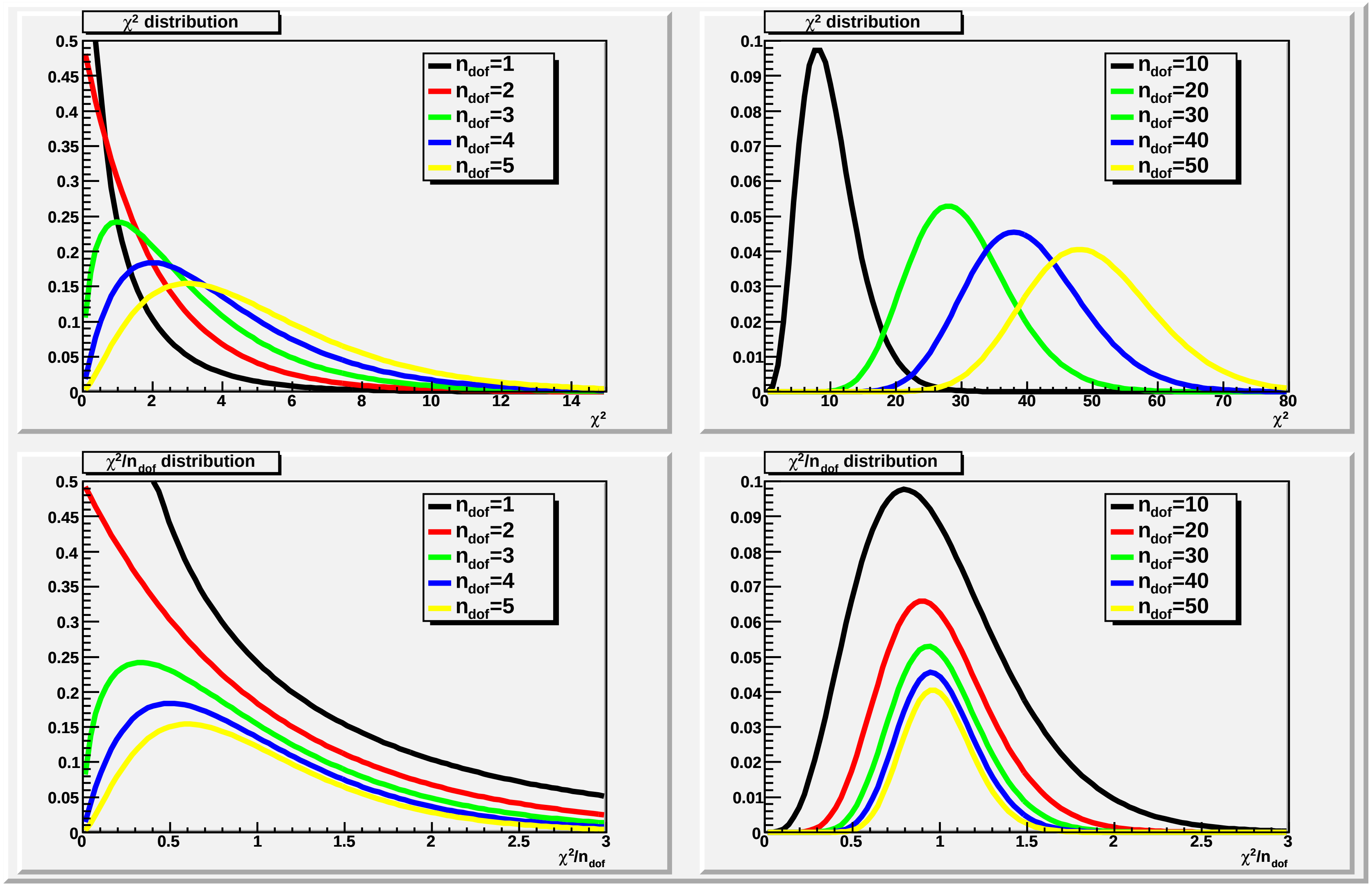
$\langle \chi^2 \rangle = n$

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

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

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

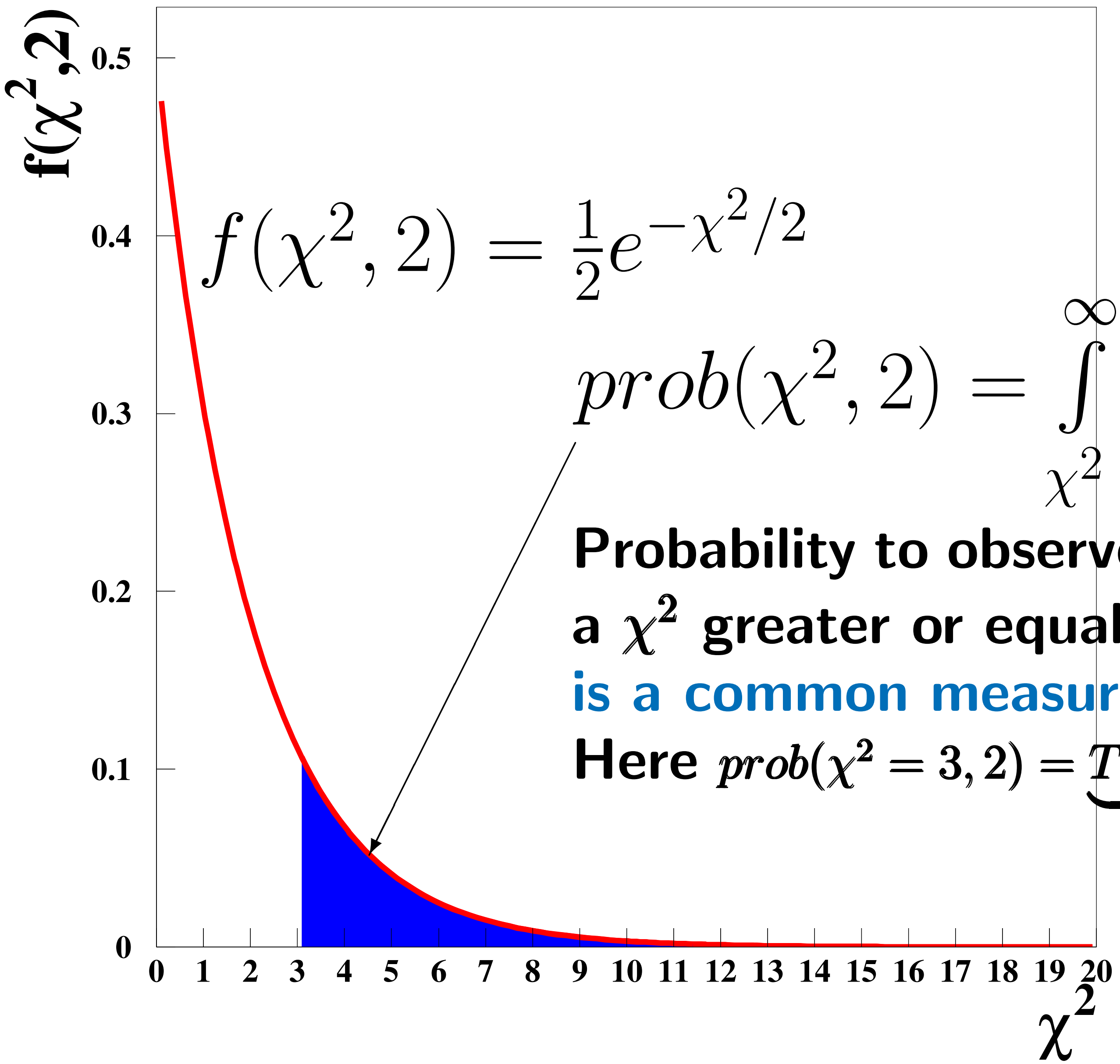
χ^2 distributions for various n



χ^2 distr.

χ^2/n distr.

$f(\chi^2, 2)$ function and $prob(\chi^2, 2)$



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

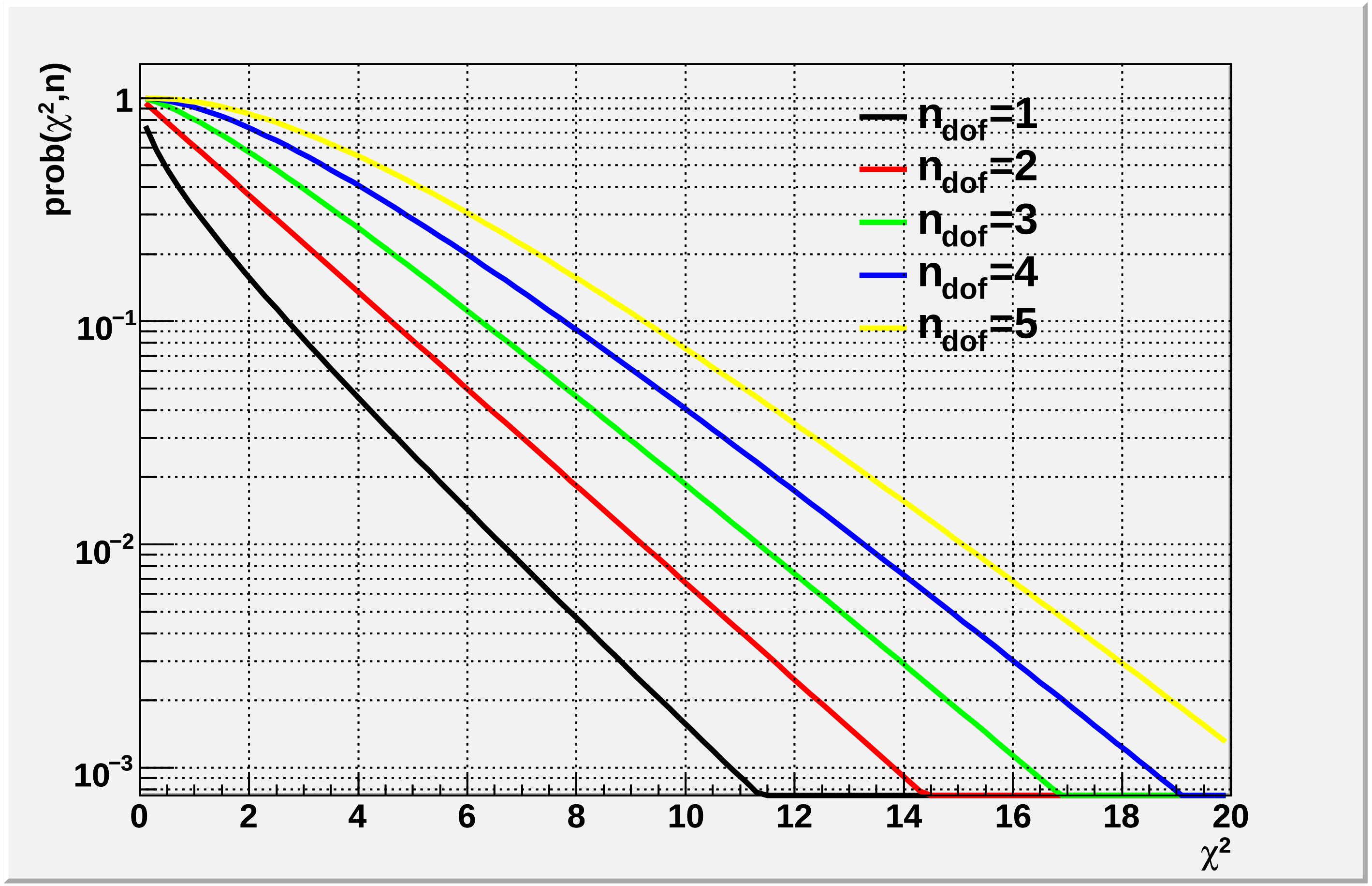
$$prob(\chi^2, 2) = \int_{\chi^2}^{\infty} f(\chi'^2, 2) d\chi'^2$$

Probability to observe for repeated experiments a χ^2 greater or equal than the current one
is a common measure for consistency

Here $prob(\chi^2 = 3, 2) = \underbrace{TMath :: Prob(3, 2)}_{\text{ROOT Function}} = 22\%$

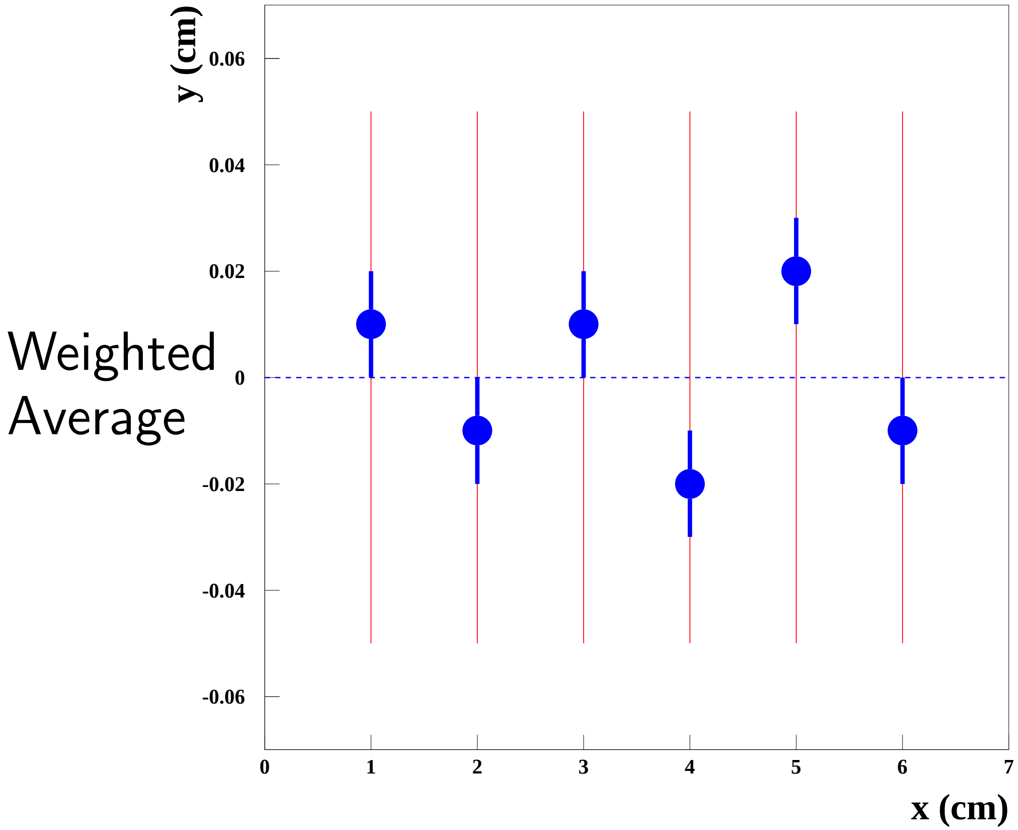
$prob(\chi^2, n)$-function for n degrees of freedom

$$prob(\chi^2, n) = \int_{\chi^2}^{\infty} f(\chi'^2, n) d\chi'^2 = \frac{1}{\Gamma(n/2)} \cdot \int_{\chi^2/2}^{\infty} dt\, e^{-t}\, t^{n/2-1}$$



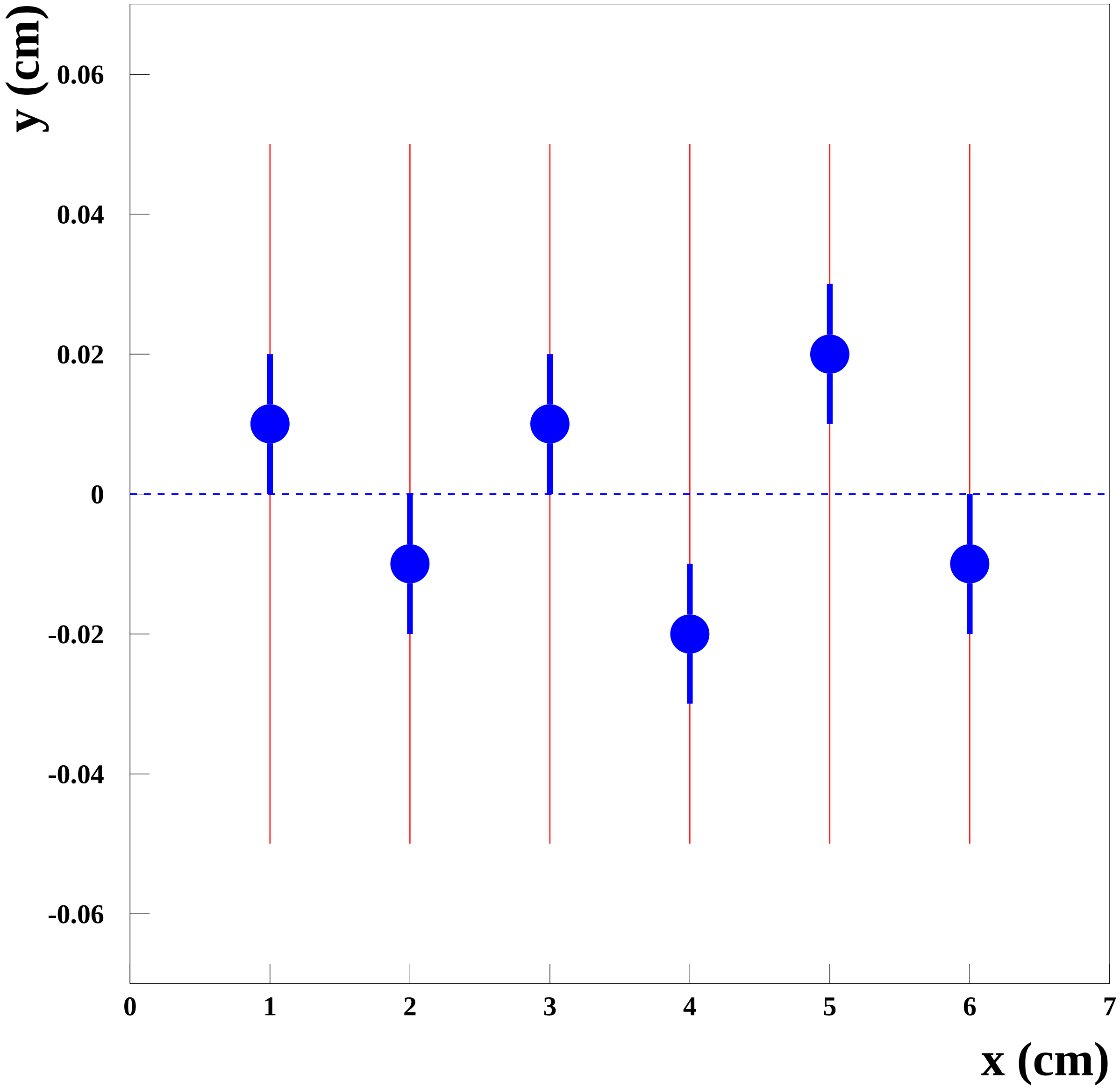
Note: for repeated experiments expect the observed values of $prob(\chi^2, n)$ to be flatly distributed over interval $[0, 1]$

χ^2 for averaging measurements



The figure shows the result of a fit of a constant using n measurements. When repeating the fit many times the resulting χ^2_{min} distribution should follow a χ^2 distribution with $n - 1$ degrees of freedom. One degree of freedom is sacrificed to determine the weighted average. A prove for this (for $n = 2$) is given in the appendix.

Mini exercise χ^2 and probability



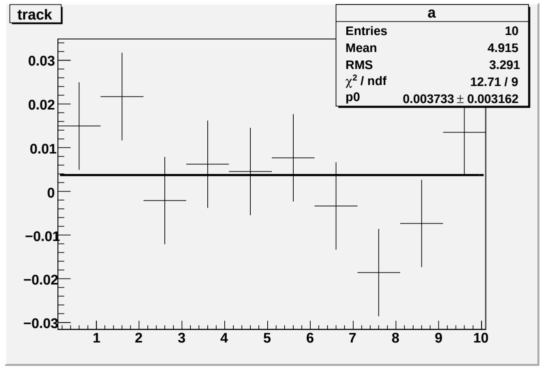
The figure shows the result of a fit of a constant. Determine the total χ^2 (from reading the figure) and the χ^2- probability.

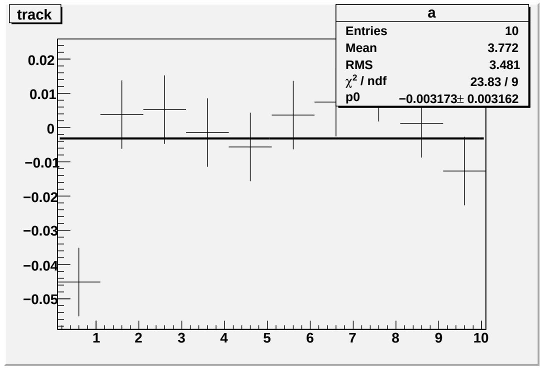
Toy simulations of constant fits through 10 data points

Fits with problems: outliers

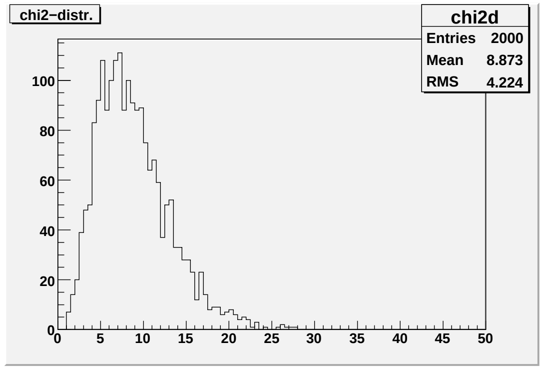
No outliers | 10% random outliers (10σ)

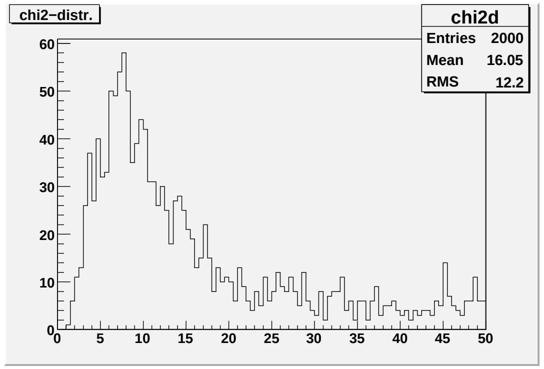
Exemplary fit



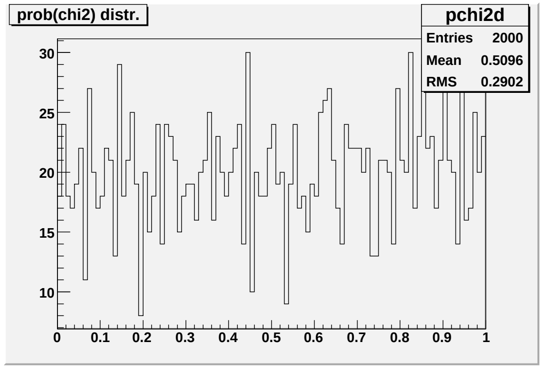


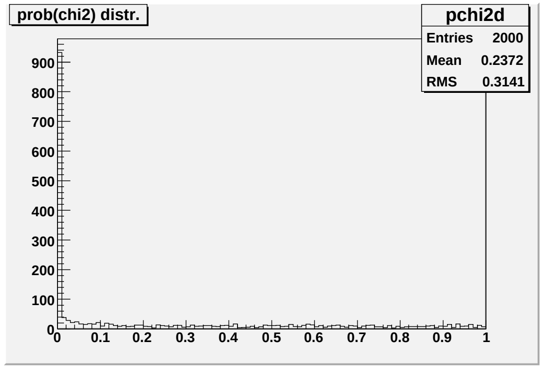
χ^2_{min} distribution for 2000 experiments





$prob(\chi^2_{min}, 9)$ distribution for 2000 experiments

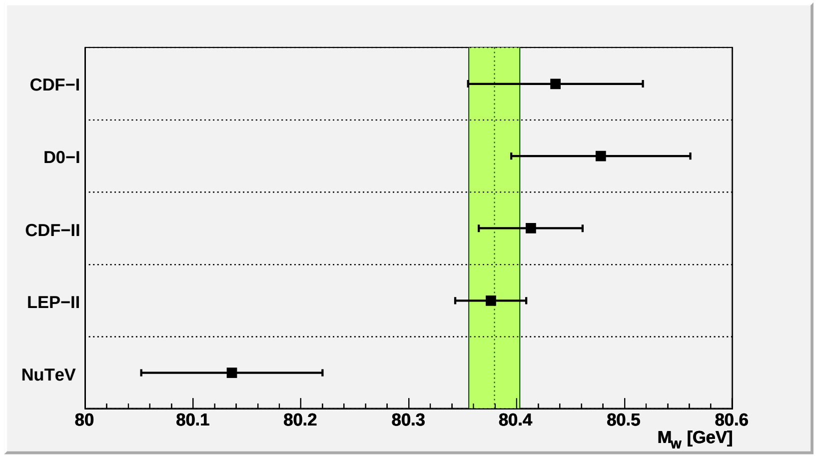




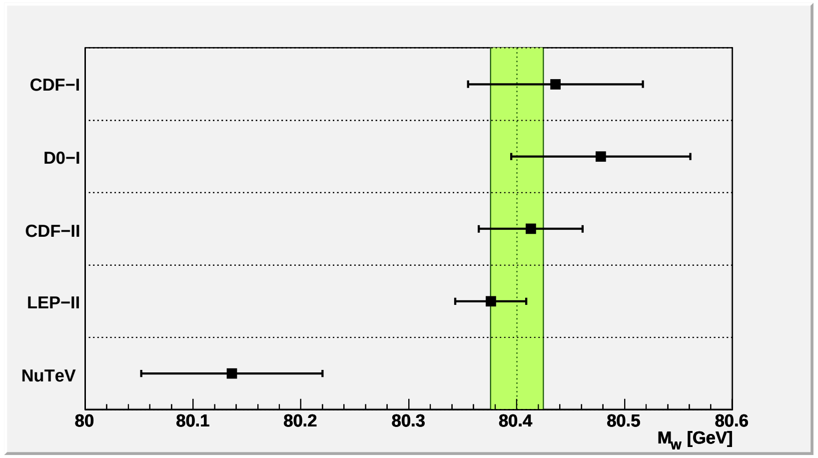
$\Rightarrow \chi^2_{min}$ and $prob(\chi^2_{min}, n_{dof})$ highly sensitive to wrong measurements

World average of W boson mass

or how to arrive at a good χ^2

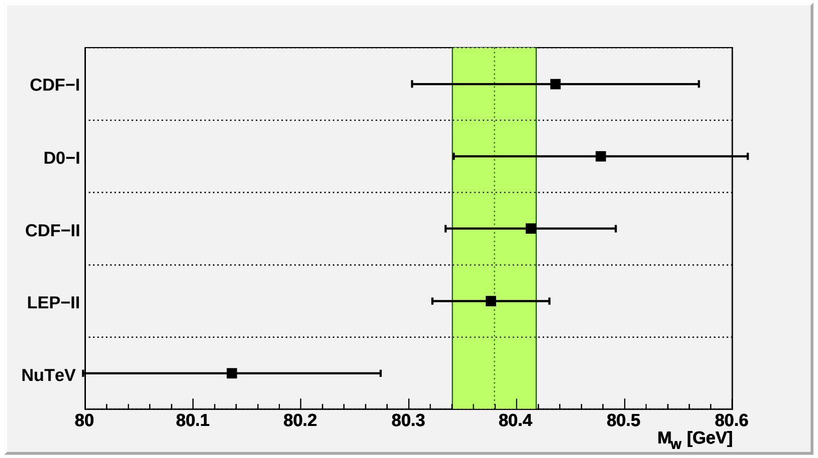


$\chi^2_{min} = 10.8$, $n_{dof} = 4$, probability $= 0.029$



Taking out NuTeV result:
$\chi^2_{min} = 1.7$, $n_{dof} = 3$, probability $= 0.64$

"Outlier rejection", is this allowed?



Scaling all errors by $S = \sqrt{\chi^2_{min}/n_{dof}} = 1.64$
$\chi^2_{min} = 4.$, $n_{dof} = 4$, probability $= 0.4$

Standard procedure by Particle Data group
→ "destroying" the hard work of many experimentalists

Mini summary of what we have learnt

- **The** χ^2_{min} **of a fit is a consistency check**
- **Expect** $\chi^2_{min}/n_{dof} \sim 1$ **for good fits**
- **if** χ^2_{min}/n_{dof} **significantly larger than one then suspect**
  - **data could contain outliers or errors are (generally) underestimated**
  - **the fitfunction might not be the correct model for the data**
- **for repeated experiments (e.g. many track fits) expect for good fits**
  - **mean value of** χ^2_{min}/n_{dof} **distribution** $\rightarrow 1$
  - **and flat** $prob(\chi^2_{min}, n_{dof})$ **distribution in interval [0,1]**

If time allows:

Average 10 measurements with noise:

with Root Macro p0toyf.C

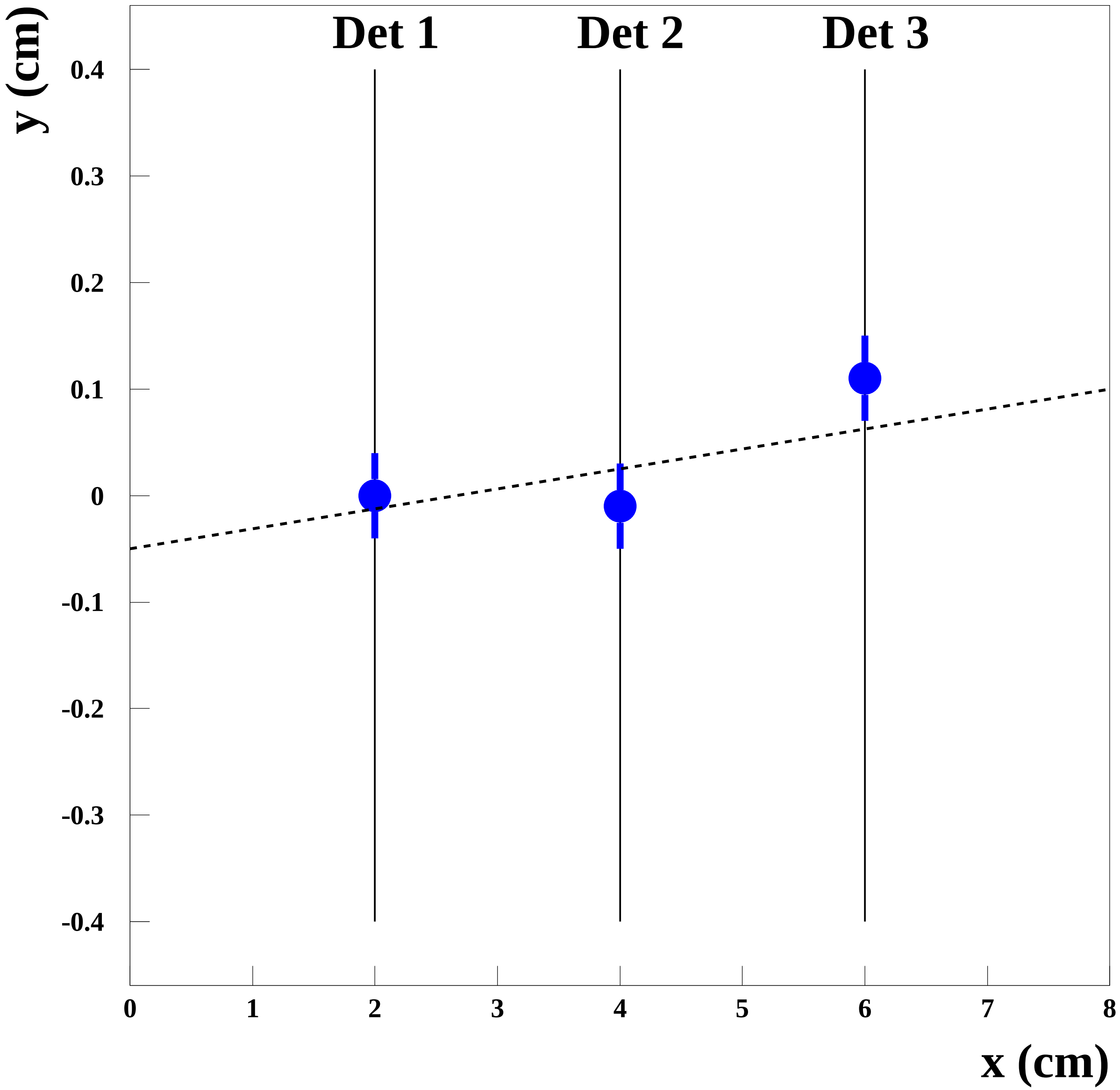
Note: Macro available at
http://www.desy.de/~obehnke/stat/school_mar10/p0toyf.C

Task instructions available at
http://www.desy.de/~obehnke/stat/school_mar10/compueb_p0toyf.pdf

Lecture part 2

- General solution for linear least square fits (normal equations)
- Straight line fits

Our study object



The main trajectories we will study in this workshop are straight lines:

$$y_i = a_0 + a_1 x_i$$

This is a classical linear least square fit problem.

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

Linear model $\vec{y} := A\,\vec{a}$, $\vec{a}$ vector of m fitparameters $\begin{pmatrix} a_1 \\ . \\ a_m \end{pmatrix}$

Example: $y = a_0;$

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

Linear model $\vec{y} :\, = A\,\vec{a}$, $\vec{a}$ vector of m fitparameters $\begin{pmatrix} a_1 \\ . \\ a_m \end{pmatrix}$

Example: $y = a_0; \quad \rightarrow \vec{a} = (a_0); \quad A = \begin{pmatrix} 1 \\ .. \\ 1 \end{pmatrix}$

In general: $A = A(\vec{x})$, but no dependence on $\vec{a}$

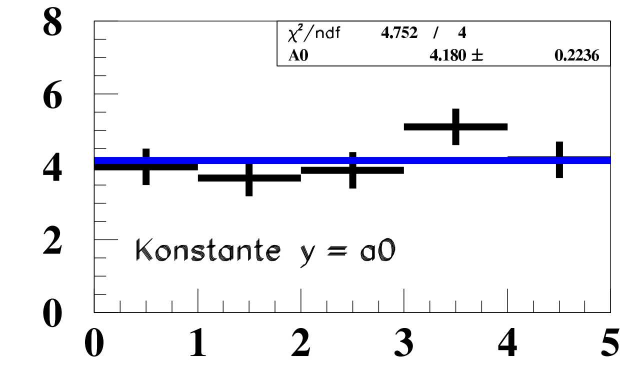
"Master formula": $$\boxed{\chi^2 = (\vec{y} - A\,\vec{a})^t\, V^{-1}\, (\vec{y} - A\,\vec{a})}$$

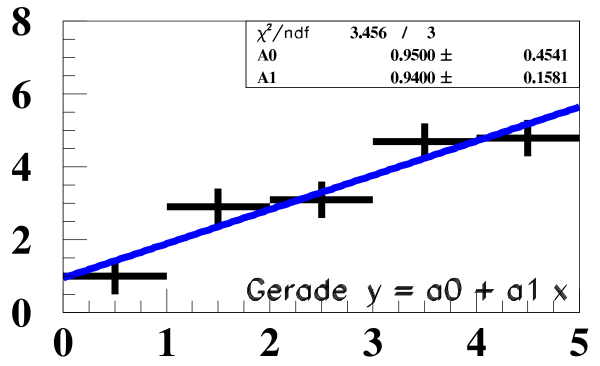
$\longrightarrow$ to be minimised w.r.t $\vec{a}$

$\longrightarrow$ obtain estimators $\hat{\vec{a}}$ and covariance matrix $V_{\hat{\vec{a}}}$

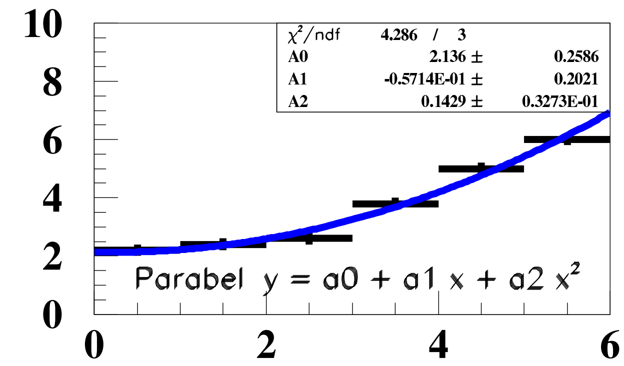
Examples for linear least square fits

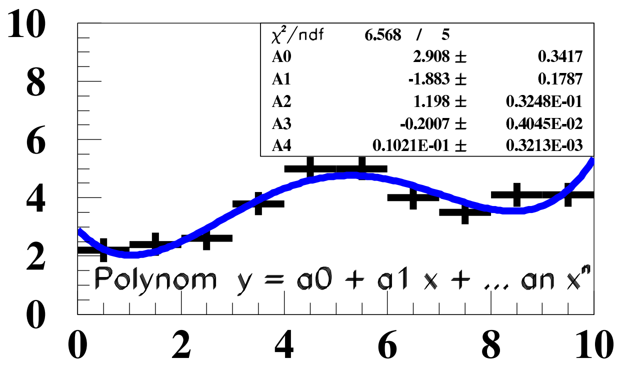
Linear means that y depends linearly on the fitparameters a_i.

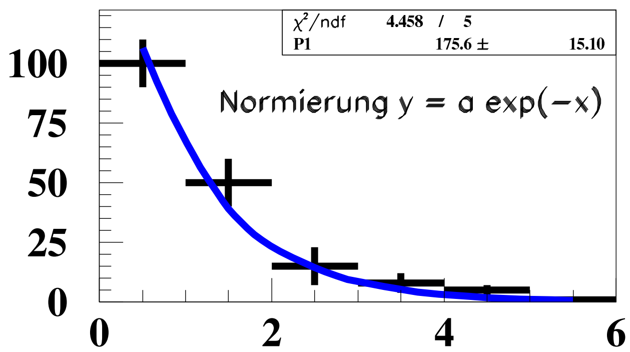




$$\vec{a} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}; \; A = \begin{pmatrix} 1 & x_1 \\ . & . \\ 1 & x_n \end{pmatrix}$$







$$\vec{a} = (a); \; A = \begin{pmatrix} e^{-x_1} \\ .. \\ e^{-x_n} \end{pmatrix}$$

$\leftarrow$ **Watch out: function can be highly non-linear in x**

General solution via normal equations

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

Min. $\chi^2 \rightarrow \frac{d\chi^2}{d\vec{a}^t} = -2A^t V^{-1} \vec{y} + 2A^t V^{-1} A \vec{a} = 0$

General solution via normal equations

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

Min. $\chi^2 \longrightarrow \dfrac{d\chi^2}{d\vec{a}^t} = -2A^t V^{-1} \vec{y} + 2A^t V^{-1} A \vec{a} = 0$

Solution:

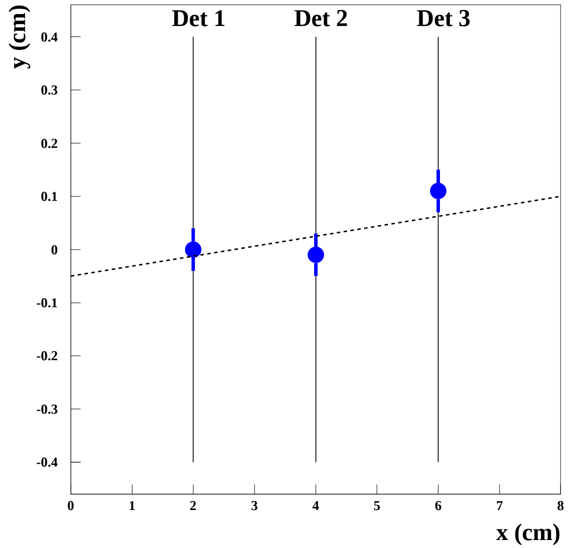
$$\hat{\vec{a}} = (A^t V^{-1} A)^{-1} A^t V^{-1} \vec{y}$$
$$= H^{-1} A^t V^{-1} \vec{y}$$
$$= U A^t V^{-1} \vec{y} \text{ with } U = H^{-1} = Cov(\hat{\vec{a}})$$

Normal equations:

Powerful & simple linear algebra to solve fit!

Straight line fit through n detector layers

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

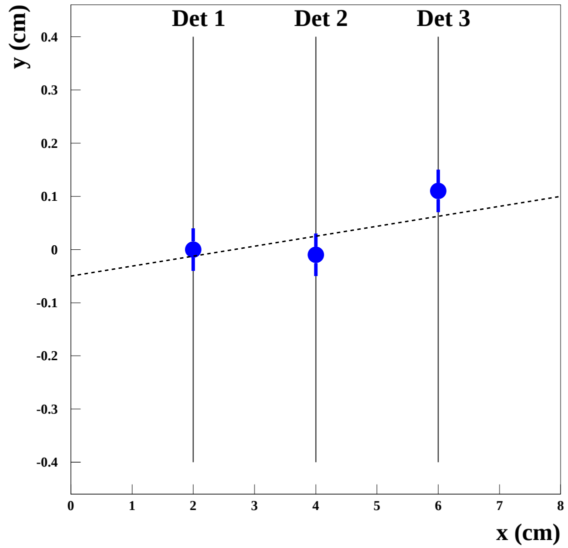


$$\vec{y} = A\vec{a};\;\; \vec{a} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix};\; A = \begin{pmatrix} 1 & x_1 \\ . & . \\ 1 & x_n \end{pmatrix};\; A^t = \begin{pmatrix} 1 & .. & 1 \\ x_1 & .. & x_n \end{pmatrix};\; V = \begin{pmatrix} \sigma^2 & & 0 \\ & . & \\ 0 & & \sigma^2 \end{pmatrix}$$

Apply normal equations:

$$\hat{\vec{a}} \;=\; (A^t V^{-1} A)^{-1} A^t V^{-1} \vec{y} = \sigma^2 (A^t A)^{-1} \cdot \frac{1}{\sigma^2} A^t \cdot \vec{y} \;=\; (A^t A)^{-1} A^t \cdot \vec{y}$$

Straight line fit through n detector layers



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

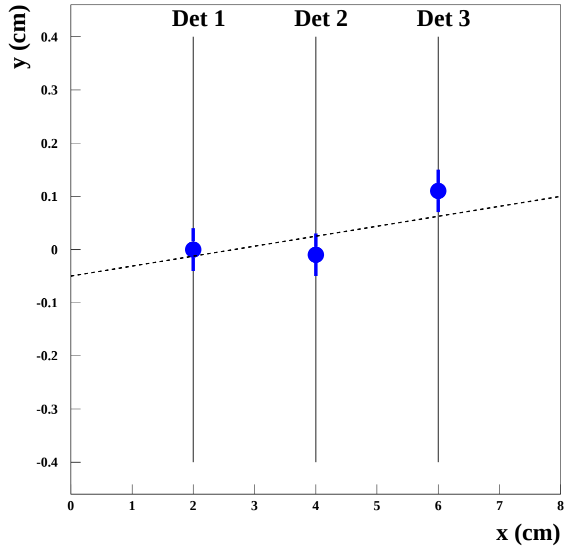
$$\vec{y} = A\vec{a};\ \ \vec{a} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix};\ A = \begin{pmatrix} 1 & x_1 \\ . & . \\ 1 & x_n \end{pmatrix};\ A^t = \begin{pmatrix} 1 & .. & 1 \\ x_1 & .. & x_n \end{pmatrix};\ V = \begin{pmatrix} \sigma^2 & & 0 \\ & . & \\ 0 & & \sigma^2 \end{pmatrix}$$

Apply normal equations:

$$\hat{\vec{a}} = (A^t V^{-1} A)^{-1} A^t V^{-1} \vec{y} = \sigma^2 (A^t A)^{-1} \cdot \frac{1}{\sigma^2} A^t \cdot \vec{y} = (A^t A)^{-1} A^t \cdot \vec{y}$$

$$= \begin{pmatrix} \sum_i 1 & \sum_i x_i \\ . & . \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix} = \begin{pmatrix} N & N\overline{x} \\ N\overline{x} & N\overline{x^2} \end{pmatrix}^{-1} \cdot \begin{pmatrix} N\overline{y} \\ N\overline{xy} \end{pmatrix}$$

Straight line fit through n detector layers



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

$$\vec{y} = A\vec{a};\ \ \vec{a} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix};\ A = \begin{pmatrix} 1 & x_1 \\ . & . \\ 1 & x_n \end{pmatrix};\ A^t = \begin{pmatrix} 1 & .. & 1 \\ x_1 & .. & x_n \end{pmatrix};\ V = \begin{pmatrix} \sigma^2 & & 0 \\ & . & \\ 0 & & \sigma^2 \end{pmatrix}$$

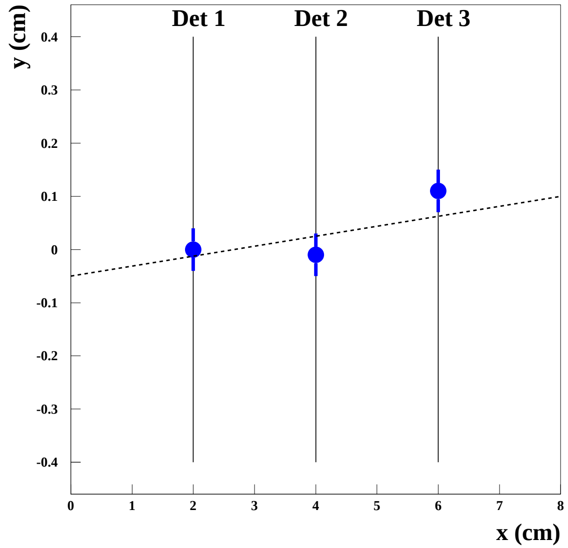
Apply normal equations:

$$\hat{\vec{a}} = (A^tV^{-1}A)^{-1}A^tV^{-1}\vec{y} = \sigma^2(A^tA)^{-1}\cdot\frac{1}{\sigma^2}A^t\cdot\vec{y} = (A^tA)^{-1}A^t\cdot\vec{y}$$

$$= \begin{pmatrix} \sum_i 1 & \sum_i x_i \\ . & . \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix} = \begin{pmatrix} N & N\overline{x} \\ N\overline{x} & N\overline{x^2} \end{pmatrix}^{-1} \cdot \begin{pmatrix} N\overline{y} \\ N\overline{xy} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \overline{y} \\ \overline{xy} \end{pmatrix} = \frac{1}{\overline{x^2} - \overline{x}^2}\begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}\begin{pmatrix} \overline{y} \\ \overline{xy} \end{pmatrix} = \frac{1}{V[x]}\cdot\begin{pmatrix} \overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy} \\ -\overline{x}\,\overline{y} + \overline{xy} \end{pmatrix}$$

Straight line fit through n detector layers



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

$$\vec{y} = A\vec{a};\ \ \vec{a} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix};\ A = \begin{pmatrix} 1 & x_1 \\ . & . \\ 1 & x_n \end{pmatrix};\ A^t = \begin{pmatrix} 1 & .. & 1 \\ x_1 & .. & x_n \end{pmatrix};\ V = \begin{pmatrix} \sigma^2 & 0 \\ & . \\ 0 & \sigma^2 \end{pmatrix}$$

Apply normal equations:

$$\hat{\vec{a}} = (A^tV^{-1}A)^{-1}A^tV^{-1}\vec{y} = \sigma^2(A^tA)^{-1}\cdot\frac{1}{\sigma^2}A^t\cdot\vec{y} = (A^tA)^{-1}A^t\cdot\vec{y}$$

$$= \begin{pmatrix} \sum_i 1 & \sum_i x_i \\ . & . \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix} = \begin{pmatrix} N & N\overline{x} \\ N\overline{x} & N\overline{x^2} \end{pmatrix}^{-1} \cdot \begin{pmatrix} N\overline{y} \\ N\overline{xy} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \overline{y} \\ \overline{xy} \end{pmatrix} = \frac{1}{\overline{x^2} - \overline{x}^2}\begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}\begin{pmatrix} \overline{y} \\ \overline{xy} \end{pmatrix} = \frac{1}{V[x]}\cdot\begin{pmatrix} \overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy} \\ -\overline{x}\,\overline{y} + \overline{xy} \end{pmatrix}$$

$$U = \begin{pmatrix} \sigma^2_{\hat{a}_0} & cov(\hat{a}_0, \hat{a}_1) \\ cov(\hat{a}_0, \hat{a}_1) & \sigma^2_{\hat{a}_1} \end{pmatrix} = (A^tV^{-1}A)^{-1} = \frac{\sigma^2}{NV[x]}\begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$$

Mini exercise: straight line track-fit

The covariance formula

$$\begin{pmatrix} \sigma^2_{\hat{a}_0} & cov(\hat{a}_0, \hat{a}_1) \\ cov(\hat{a}_0, \hat{a}_1) & \sigma^2_{\hat{a}_1} \end{pmatrix} = \frac{\sigma^2}{NV[x]} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$$

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 N 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 σ reduced by a factor two

Computer exercise straight line trajectory fit

Physics example: A muon track is measured in four layers of streamer tube detectors at x positions of 4., 5., 6. and 7. (in cm), with a measurement precision for y of 0.5 cm. The goal is to determine its trajectory assuming a straight line.

Macro StraightLineFit.C, accessible at
$\boldsymbol{http://www.desy.de/\tilde{}obehnke/stat/gean10/StraightLineFit.C}$

fits a straight line track trajectory through four measured points.

- Steering parameters in the macro:
  - $xmin, xmax$ = Interval of the trajectory displayed
- Output:
  - Histogram $data$ (it's of the type TGraphErrors)
  - Plots are drawn of the
    - fitted histogram with error bands
    - error ellipse of the two fitparameters

Tasks:

a) Run the macro as it is by .x StraightLineFit.C and fill the fit results for $p0$, $p1$, their errors and correlation into the table below

b) Precision of trajectory: Evaluate (by eye) from the shown error bands at which point roughly the trajectory is known best and with which precision (fill the results in the table below)

c) Precision of extrapolated trajectory: Evaluate the precision of the extrapolated trajectory at $x = 100$ (Hint: Change $xmax$ to large value and run the macro again)

d) Effect of shift of x coordinate origin: Shift all four $xVal$ points in the macro (simply by overwriting by hand) by a constant value -5.5, set $xmin = -4.$ and $xmax = 4.$ and run the macro again. Fill the fit results in the table. Can you explain why the correlation of $p0$ and $p1$ has changed?

e) Apply a very precise vertex constraint at the origin: Change N to 5 and add a new first point to the measurement points list with $xVal = 0.0$, $xErr = 0.0$, $yVal = 0.0$ and $yErr = 0.0001$ (just by hand). Run the macro again and write down the fitted results in the table. How much are the parameter errors reduced by adding this extra point?

Task results sheet

	Straight line fit trough four points
Task a)	$p0 =$ $p1 =$ corr $=$
Task b)	x-best precision $=$ y-error $=$
Task c)	y-error$(x = 100) =$
Task d)	Shifting all x values by -5.5: $p0 =$ $p1 =$ corr $=$
Task e)	Adding vertex constraint at $x = 0$: $p0 =$ $p1 =$ corr $=$

Mini summary of what we have learnt

- **Linear least square problems:** $\vec{y} = A\vec{a}$**,**
  $\rightarrow$ y **is a linear function of the fitparameters** $\vec{a}$ **but can be a linear or nonlinear function of the continuous parameter** x**.**

- **The normal equations are a powerful tool to solve linear least square fit problems** $\hat{\vec{a}} = (A^t V^{-1} A)^{-1} A^t V^{-1} \vec{y}, \;\; cov(\hat{\vec{a}}) = (A^t V^{-1} A)^{-1}$

- **Straight line fits are a typical application and there are many others (e.g. parabolas, higher order polynoms, etc.)**

Appendix

Content:

- Proof that χ^2_{min} for averaging two measurements follows χ^2-distribution with one degree of freedom
- Linear least square fits: Covariance matrix of fit parameters

χ^2 for two measurements with unknown true value

$$\chi^2_{min} = \frac{(y_1-\hat{a})^2}{\sigma_1^2} + \frac{(y_2-\hat{a})^2}{\sigma_2^2}; \quad \hat{a} = \frac{1}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}} \cdot \left(\frac{y_1}{\sigma_1^2}+\frac{y_2}{\sigma_2^2}\right) = \frac{G_1y_1+G_2y_2}{G_1+G_2} \quad \text{(with } G_i := 1/\sigma_i^2\text{)}$$

$$\begin{aligned}
\Rightarrow \chi^2_{min} &= G_1 \cdot \left(y_1 - \frac{(G_1y_1+G_2y_2)}{G_1+G_2}\right)^2 + G_2 \cdot \left(y_2 - \frac{(G_1y_1+G_2y_2)}{G_1+G_2}\right)^2 \\
&= G_1 \cdot \left(\frac{(G_2y_1-G_2y_2)}{G_1+G_2}\right)^2 + G_2 \cdot \left(\frac{(G_1y_2-G_1y_1)}{G_1+G_2}\right)^2 \\
&= \frac{G_1G_2^2}{(G_1+G_2)^2}(y_1-y_2)^2 + \frac{G_2G_1^2}{(G_1+G_2)^2}(y_1-y_2)^2 \\
&= \frac{G_1G_2(G_1+G_2)}{(G_1+G_2)^2} \cdot (y_1-y_2)^2 = \frac{G_1 \cdot G_2}{G_1+G_2} \cdot (y_1-y_2)^2 \\
&= \frac{1}{1/G_1+1/G_2} \cdot (y_1-y_2)^2 = \frac{1}{\sigma_1^2+\sigma_2^2} \cdot (y_1-y_2)^2
\end{aligned}$$

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

$\rightarrow \chi^2 = \Delta^2$ follows 1-dim 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

Linear least square fits: Covariance Matrix

Proof that Covariance matrix U of fit parameters $\hat{\vec{a}}$ is given by $U = H^{-1}$

Use Normal Equations:

$$\hat{\vec{a}} = B\vec{y} \quad \text{with} \quad B = H^{-1}A^tV^{-1}$$

Then apply errorpropagation:

$$\begin{aligned} \rightarrow U &= BVB^t = H^{-1}A^tV^{-1}VV^{-1}AH^{-1} \\ &= H^{-1}A^tV^{-1}AH^{-1} = H^{-1}HH^{-1} = H^{-1} \end{aligned}$$