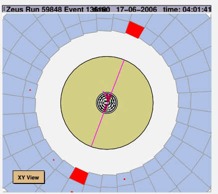
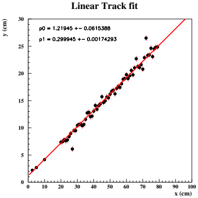
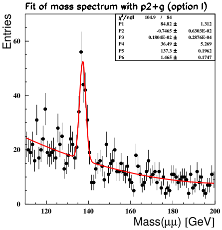
Least Squares Fits

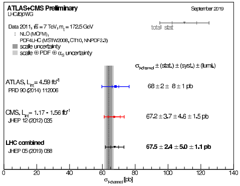
Aachen Online Statistics School, March 13, 2023

Olaf Behnke, DESY









- Least squares fits are a workhorse in data analysis
- In its most simple form: fitting to data $y_i \pm \sigma_i$ measured at positions x_i known model $f(x_i; \vec{a})$ depending on fit parameters $\vec{a}$, by minimising

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

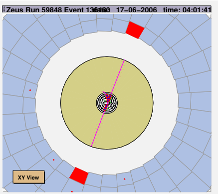
- Ideal tool if measurements have known gaussian uncertainties
- HEP Examples: Track fits, s+b fits to binned mass distributions *(not optimal tool!)* and combining data

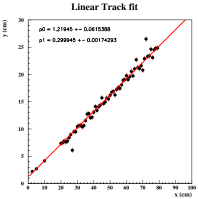
Quiz

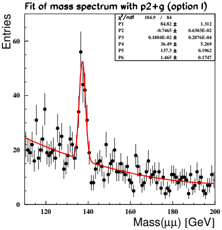
What is the maximum number of fit parameters that have been fitted with Least Squares in HEP (and in which application)?

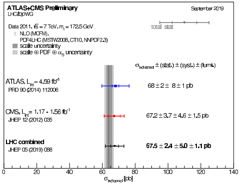
Outline

- Combining measurements
- χ^2 as a measure of goodness-of-fit
- Linear and non-linear fits (straight line, circle, mass peak fit)

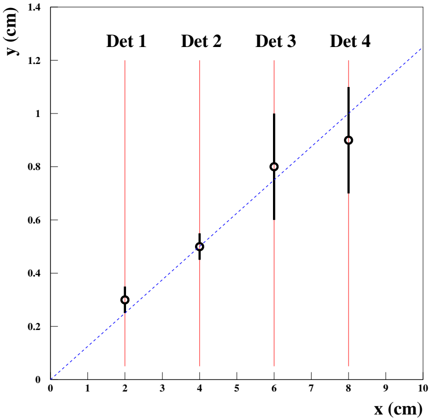








χ^2 fit - Heuristic motivation



- n-measurements $y_i \pm \sigma_i$ at fixed x_i
- Model: $y = f(x, a)$ here: $y = ax$
- **How to determine *a*?**
  $\Rightarrow$ Idea: for correct a one expects: $|y_i - f(x_i, a)| \lessapprox \sigma_i$

Min. $\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f(x_i, a))^2}{\sigma_i^2}$ turns out to be good & practical method!

χ^2 fit - Minimisation

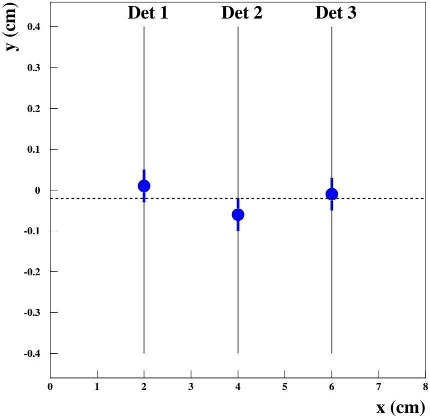
$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - f(x_i, a))^2}{\sigma_i^2}$$

Task: find Minimum w.r.t a

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

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

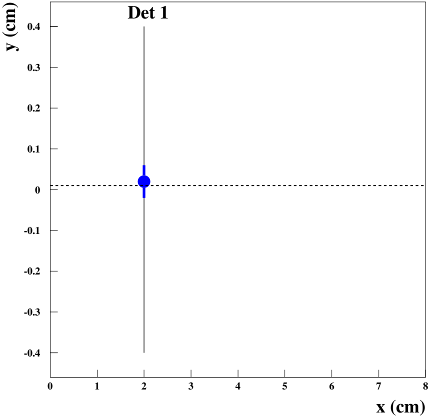
Fit of a constant function (no x dependence)



- Determine vertical position of horizontally flying particle
- 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 function (one measurement)



- **"Idiot example"** of single measurement $y_1 \pm \sigma_1$

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

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

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

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

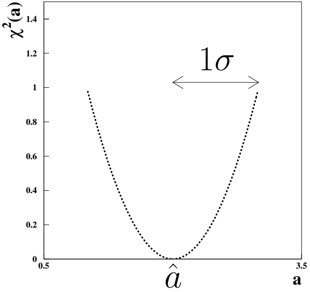
Fit of a constant function (one measurement)

Likelihood $L \sim \exp\left[-\frac{(a-\hat{a})^2}{2\sigma_{\hat{a}}^2}\right]$ with $\chi^2 = \frac{(a-\hat{a})^2}{\sigma_{\hat{a}}^2}$

$\Rightarrow L \sim e^{-\chi^2/2}$ and $\chi^2 = -2ln(L)$

Max. L $\equiv$ Min. χ^2 (holds for fitting to measurements with known gaussian uncertainties)

Retrieve $\sigma_{\hat{a}}^2$ from : $\frac{1}{\sigma_{\hat{a}}^2} = \frac{1}{2}\frac{d^2\chi^2}{da^2}_{|a=\hat{a}}$ or from $\chi^2(\hat{a} \pm \sigma_{\hat{a}}) - \chi^2(\hat{a}) = 1$

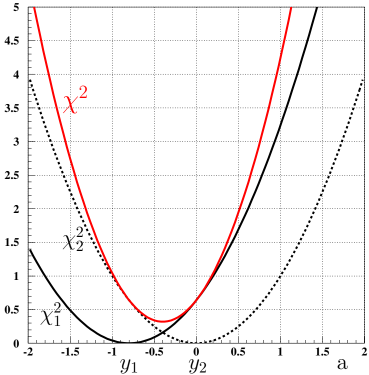


Note: These are the two standard error determination methods for χ^2 fits!
For generalised $\tilde{\chi}^2 = -2\,ln(L)$, the second method is more reliable for non-gaussian L, why?

Fit of a constant function - *n* measurements

Likelihood for observed measurements y_i as function of true value a:

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



- χ^2 is sum of individual $\chi_i^2 = \frac{(y_i-a)^2}{\sigma_i^2}$
- The sum of parabolas is another parabola
- Averaging can be done graphically!

Fit of a constant function - many measurements

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

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

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

'Hesse matrix'
(for one par. a number)

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

$\Rightarrow$ *Latter term can be interpreted as Bayesian posterior density for true* $\boldsymbol{a}$*, using flat prior: Gaussian with center* $\boldsymbol{\hat{a}}$ *and width* $\boldsymbol{\sigma = H^{-1/2}}$

Averaging several measurements

n measurements $y_i \pm \sigma_i$:

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

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

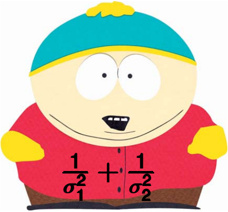
Result

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

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

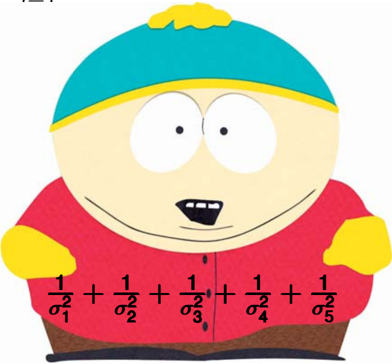
Role of Hesse matrix - illustrated for weighted average

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



***H* "grows" with each measurement**

$\longrightarrow$



H is "counting the Fisher information" from the measurements

Finally $\sigma_{\hat{a}}^2 = cov(\hat{a}) = H^{-1}$

Note: all this holds also for LSQ fits with many parameters

Averaging - reformulated

Single measurements contribute with **weight** $G_i = \frac{1}{\sigma_i^2}$; $G_s := \sum_{i=1}^{n} G_i$;

Fit result

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

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

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

$\Rightarrow$ **Least square fitting is a clever mapping of measurements to fit-parameters and applying error propagation!**

Averaging - battle of weight schemes

Generalised averaging result

$$\hat{a} = \frac{1}{G_s} \cdot \sum_{i=1}^{n} G_i y_i \quad \text{with} \quad G_s := \sum_{i=1}^{n} G_i$$

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

Average $y_1 = 12 \pm 1$ and $y_2 = 8 \pm 3$

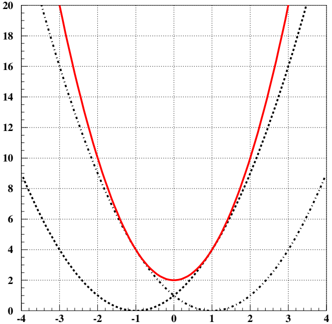
1. $G_i = 1 \Rightarrow \hat{a} = 10.; \quad \sigma_{\hat{a}} \approx 1.6$
2. $G_i = 1/\sigma_i \Rightarrow \hat{a} = 11.1; \quad \sigma_{\hat{a}} \approx 1.05$
3. $G_i = 1/\sigma_i^2 \Rightarrow \hat{a} = 11.6; \quad \sigma_{\hat{a}} \approx 0.95$

Least squares wins

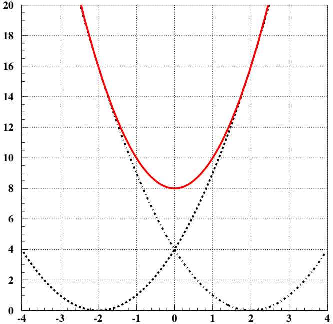
Graphical averaging of two measurements - Exercise

All input measurements have uncertainty $\sigma = 1$
Shown are χ^2 curves for two measurements and their sum (red)

Case 1: $y_1 = -1$, $y_2 = 1$



Case 2: $y_1 = -2$, $y_2 = 2$



- How do the LSQ results for $\hat{a}$ and $\sigma_{\hat{a}}$ differ for the two cases?
- Homework: proof that in general $\chi^2_{min} = \chi^2(\hat{a}) = \frac{(y_1 - y_2)^2}{\sigma_1^2 + \sigma_2^2}$, where σ_1 and σ_2 denote the y_1 and y_2 unc.

Consistency of measurements

Recall Likelihood decomposition for averaging *n* measurements:

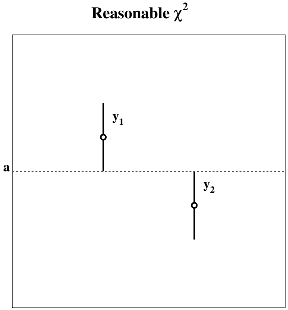
$$\Rightarrow L(y_1, y_2, ..., y_n|a) \propto \underbrace{e^{-\frac{\chi^2(\hat{a})}{2}}}_{\text{Fit consistency}} \cdot \underbrace{e^{-\frac{1}{2} H \cdot (\hat{a}-a)^2}}_{\text{Parameter info}}$$

Now lets have a closer look at the first term

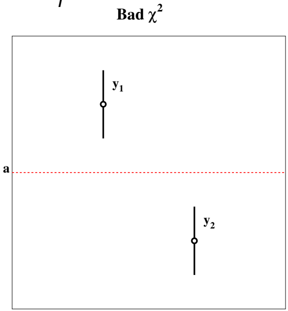
Consistency of measurements

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

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



$\chi^2 = 2$



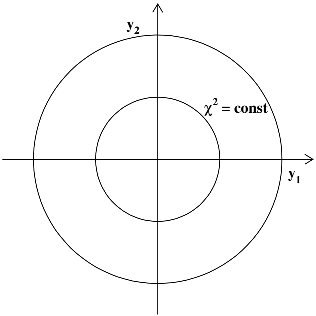
$\chi^2 = 8$

$\Rightarrow \chi^2$ is a measure of consistency
But how should χ^2 be distributed?

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

$$f(y_1, y_2) \propto e^{-y_1^2/2} e^{-y_2^2/2} = e^{-r^2/2} = e^{-\chi^2/2}$$

Probability to find value between r and $r + dr$
$\Rightarrow$ enhanced by space factor $2\pi r$



Finally

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

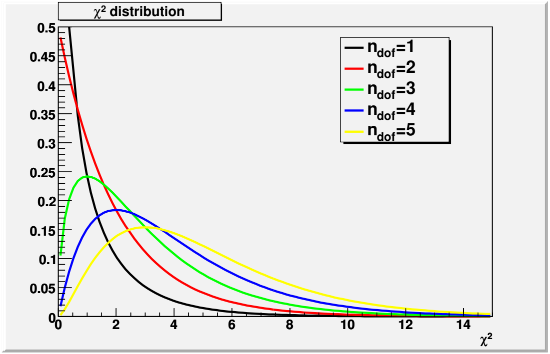
$\rightarrow$ introduces χ^2-distribution for $z = \chi^2$ and two dimensions (ndf=2):
$f(z, 2) = \frac{1}{2} e^{-z/2}$

χ^2 function for n degrees of freedom

→ 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:

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

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

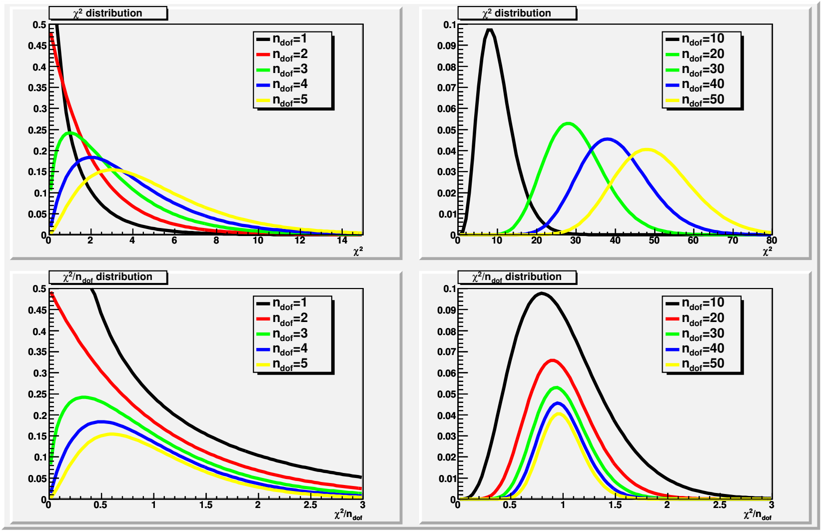
$V(\chi^2) = 2n; \; \sigma(\chi^2) = \sqrt{2n}$

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

$V(\chi^2/n) = 2$

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

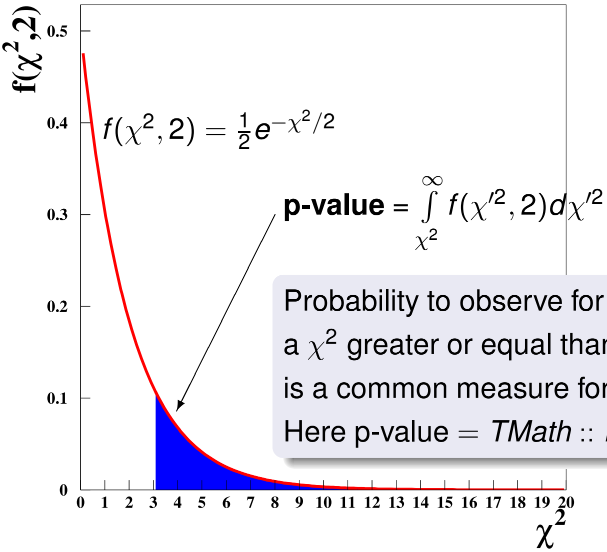
χ^2 function for various n



χ^2 distr.

χ^2/n distr.

$f(\chi^2, 2)$ function and p-value

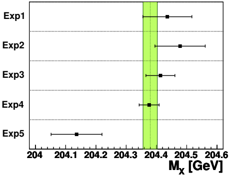


Probability to observe for repeated experiments a χ^2 greater or equal than the current one is a common measure for consistency

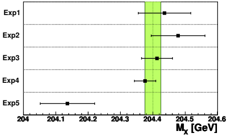
Here p-value $=$ $TMath :: Prob(3, 2) = 22\%$

Question: Is a $\chi^2/ndf = 1.2$ showing reasonable consistency?

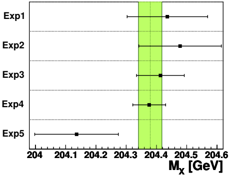
World averages



$\chi^2_{min} = 10.8$, $n_{dof} = 4$
p-value of χ^2_{min} = 0.029



Taking out Experiment 5:
$\chi^2_{min} = 1.7$, $n_{dof} = 3$, p-value = 0.64
"Outlier rejection", is this allowed?



Scaling all errors by $s = \sqrt{\chi^2_{min}/n_{dof}} = 1.64$
$\chi^2_{min} = n_{dof} = 4$, p-value = 0.4
Standard procedure by Particle Data group
$\rightarrow$ "destroying" the hard work of many experimentalists, but what can one do?

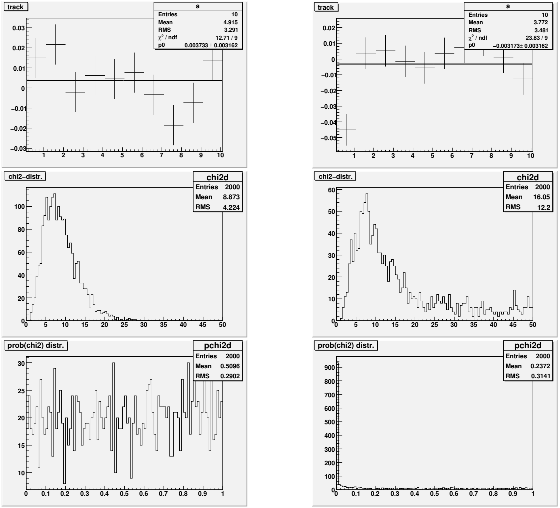
Toy simulations of p0 “track fits” through 10 data points

No outliers | Random 10% outl.

Exemplary fit

χ^2 distribution for 2000 fits

$TMath :: Prob(\chi^2, 9)$ distribution



$\Rightarrow \chi^2$ and its p-value value highly sensitive to outliers!

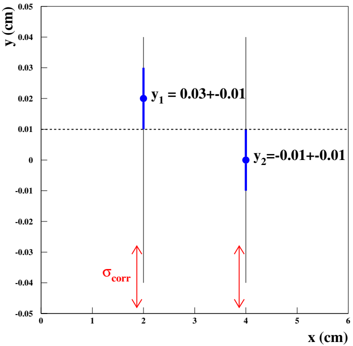
Most general LSQ fit case

- y_i, y_j correlated measurement with cov. V_{ij}
- Use vectors
  $\vec{y}^t = (y_1, y_2, .., y_n)$ and $\vec{f}(\vec{a})^t = (f(x_1), f(x_2), .., f(x_n))$
- m fit-parameters $\vec{a}$

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

Example averaging two correlated measurements y_1, y_2

Measure vertical track position in two detector layers with global position uncertainty:



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

Min. χ^2 solution: $a = 0.01 \pm \sqrt{\frac{0.01^2}{2} + \sigma_{corr}^2}$

Linear LSQ fits

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

Linear model $\vec{y}: \; = A\,\vec{a}$, A is called design matrix

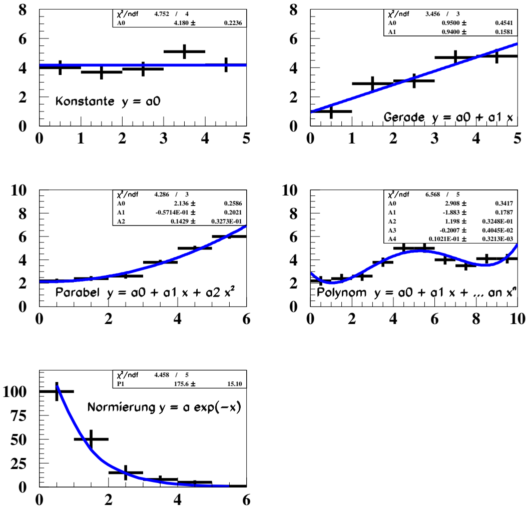
Example constant: $y = a_0; \quad \rightarrow \vec{a} = (a_0); \quad A = \begin{pmatrix} 1 \\ .. \\ 1 \end{pmatrix}$

Example parabola: $y = a_0 + a_1 x + a_2 x^2$

$$\rightarrow \vec{a}^t = (a_0, a_1, a_2); \quad A = \begin{pmatrix} 1 & x_1 & x_1^2 \\ .. & & \\ 1 & x_n & x_n^2 \end{pmatrix}$$

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

Examples for linear least squares fits



- Function can be highly non-linear in x

Linear χ^2 fit solution

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

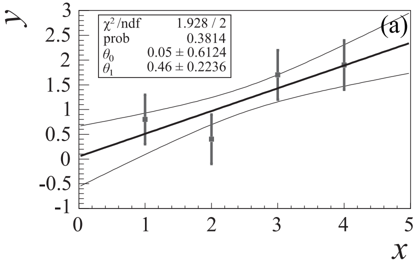
$$\text{Min. } \chi^2 \rightarrow \frac{d\chi^2}{d\vec{a}^{\,t}} = 0 = -2A^t V^{-1} \vec{y} + 2A^t V^{-1} A\vec{a}$$

Normal equations:

$$\begin{aligned} \hat{\vec{a}} &= (A^t V^{-1} A)^{-1} A^t V^{-1} \vec{y} \\ &= H^{-1} A^t V^{-1} \vec{y} \qquad \text{with } H = (A^t V^{-1} A) = \frac{1}{2} \frac{d^2\chi^2}{d\vec{a}^2} \\ Cov(\hat{\vec{a}}) &= H^{-1} \end{aligned}$$

Powerful & simple linear algebra to solve fit!

Straight line fit



$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - \theta_0 - x_i\,\theta_1)^2}{\sigma^2}$$

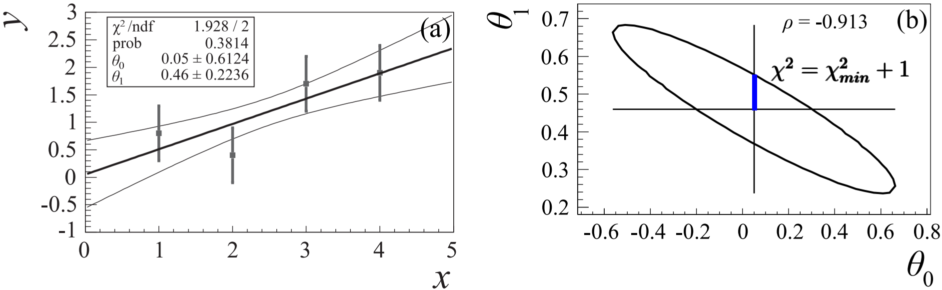
$$\Leftrightarrow \chi^2 = (\vec{y} - A\vec{\theta})^T\, V^{-1}\, (\vec{y} - A\vec{\theta}),$$

with $\vec{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$; $A = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}$; $V = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}$

Solution with normal equations:

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

Straight line fit - Fit parameter uncertainties



$$V(\hat{\vec{\theta}}) = \frac{\sigma^2}{NV[x]} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$$

- Uncertainty of slope $\theta_1 \sim 1/\sqrt{V[x]}$ – lever arm matters!
- Negative correlation coefficient $\rho = \frac{V_{01}}{\sqrt{V_{00}V_{11}}} = \frac{-\overline{x}}{\sqrt{\overline{x^2}}} = -0.913$

  $\Leftrightarrow$ Raising θ_0 can be compensated by lowering θ_1
- Fixing θ_0 to 0.05 $\Rightarrow$ reduces θ_1 uncert. by factor $\sqrt{1-\rho^2} = 0.4$

Non linear least squares fits (one parameter example)

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

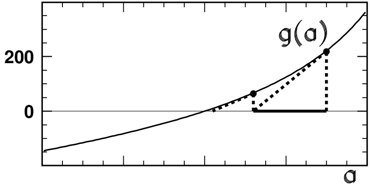
Now $f(x_i, a)$ depends **non-linearly** on a, examples:

$$f(x, a) = tan(ax), \quad ln(ax), \quad a\,exp(-ax)$$

Find min. χ^2 by solving for $g = \frac{d\chi^2}{da} = 0$ with **Newton steps**:

$$a_{m+1} = a_m - \frac{g(a_m)}{g'(a_m)}$$

iteration index m



In Appendix: example of a circle fit (transverse track trajectory)

Highly non-linear mass peak fit, $x = m = m_{\mu^+\mu^-}$

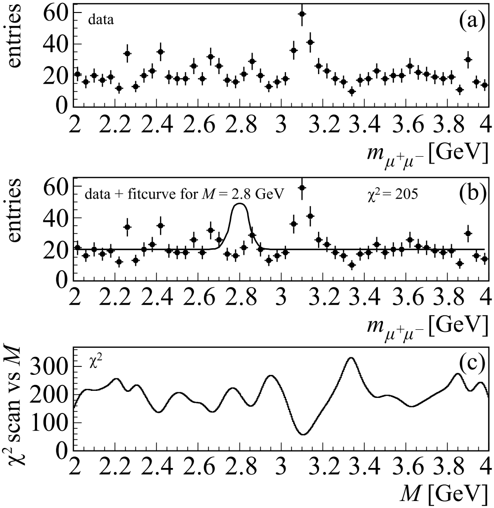
Fit to observed event counts k_i

$$f(m; M) = B + S \cdot \exp\left[\frac{(m-M)^2}{2\sigma^2}\right]$$

- B known background
- S: predicted Signal strength
- σ: known detector resolution
- M: unknown mass of particle

Use Neyman-χ^2 that assumes $\sigma_{k_i} = \sqrt{k_i}$ and **scan χ^2 vs M**

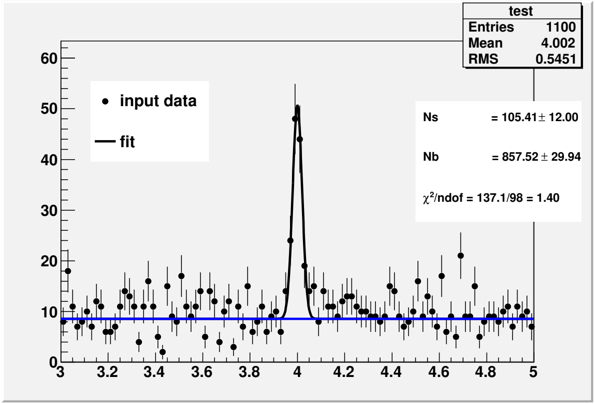
$$\chi^2 = \sum_{bin\, i} \frac{[k_i - f_i(m; M)]^2}{k_i}$$



- Many local χ^2 minima, danger to get caught there
- reasonable χ^2 of 47 (ndof = 49) only at global χ^2_{min} near J/ψ mass

Binned mass peak fit: Neyman χ^2

- Fit to event counts k_i in bin i
- Fit function f=g+p0; $f_i = \int_{bin\,i} f\,dm$:



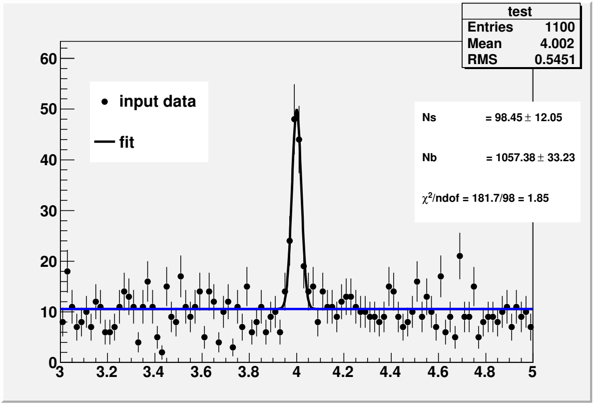
Neyman χ^2 estimator:
Assumes $\sigma_{k_i} = \sqrt{k_i}$

$$\chi^2 = \sum_i \frac{(f_i - k_i)^2}{k_i}$$

Bias:

$$\sum_i \hat{f}_i = \sum_i k_i - \chi^2$$

Bins with $k_i < f_i$ pull fit down, because assumed uncertainty $\sigma_i = \sqrt{k_i}$ is too small!



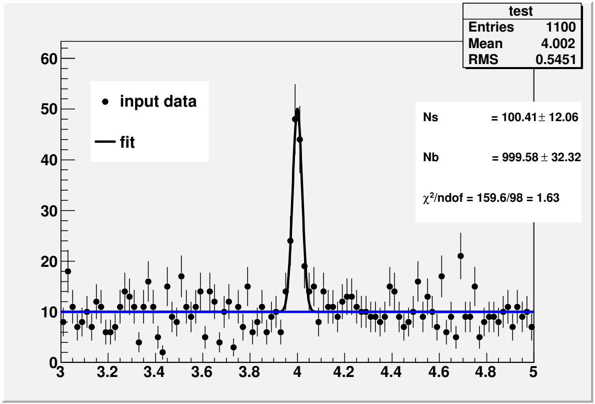
Pearson χ^2 estimator:
Assumes $\sigma_{k_i} = \sqrt{f_i}$

$$\chi^2 = \sum_i \frac{(f_i - k_i)^2}{f_i}$$

Bias:

$$\sum_i \hat{f}_i = \sum_i k_i + \chi^2/2$$

Increasing f_i in denominator of χ^2 terms decreases χ^2!



Poisson based likelihood:

$$L = \prod e^{-f_i} f_i^{k_i} / k_i!$$

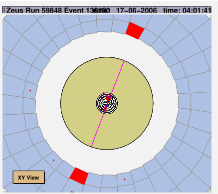
$$\tilde{\chi}^2 = -2 ln(L)$$

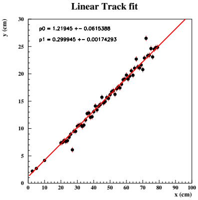
No Bias:

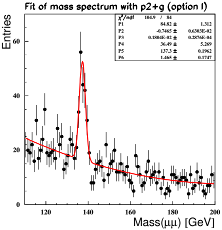
$$\sum_i \hat{f}_i = \sum_i k_i$$

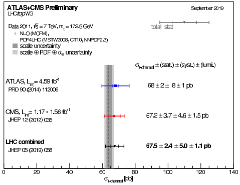
And the winner is Maximum Likelihood

Summary







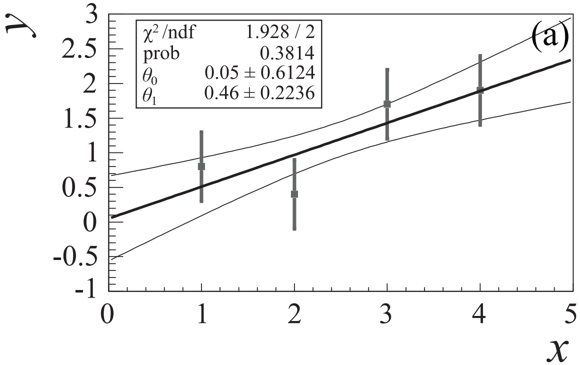


- Least squares fit is an essential parameter estimation tool
- Ideal for fits to measurements with known gaussian uncertainties
- Min. χ^2 values provide important GOF-test
- Many more LSQ fit applications than discussed today, e.g.:
  - **Alignment**, fit with **Millepede** positions of $\sim$40k CMS tracker modules
  - **Kinematic constraint fits** (see http://www-library.desy.de/preparch/books/BloLoBuch.pdf)
  - **Unfolding of differential cross sections** see https://arxiv.org/abs/1611.01927 and https://indico.desy.de/indico/event/22731/session/5/contribution/24/material/slides/0.pdf

Literature

- Roger Barlow: “Statistics, A Guide To The Use Of Statistical Methods In The Physical Sciences” Wiley & Sons, 1994
- Olaf Behnke, Kevin Kröninger, Gregory Schott and Thomas Schörner Sadenius: “Data Analysis in-High-Energy-Physics” Wiley & Sons, 2013
- Glen Cown: “Statistical Data Analysis”, Oxford Science Publications, 1997
- Fred James: “Statistical Methods in Experimental Physics”, 2nd edition, World Scientific, 2006
- Louis Lyons: ”Statistics for Nuclear and Particle Physicists”, Cambridge University Press, 1986

Appendix



Central straight-line fit defines **best position estimate** $\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x$ and two lines $\hat{y} \pm \sigma_{\hat{y}}$ a 68% **enclose central confidence region**, with $\sigma_{\hat{y}}$ from error propagation:

$$\sigma_{\hat{y}} = \sqrt{\left(\frac{\partial \hat{y}}{\partial \theta_0}\right)^2 V_{00} + \left(\frac{\partial \hat{y}}{\partial \theta_1}\right)^2 V_{11} + 2\frac{\partial \hat{y}}{\partial \theta_0}\frac{\partial \hat{y}}{\partial \theta_1} V_{01}} = \sqrt{V_{00} + x^2 V_{11} + 2x V_{01}}\,.$$

LSQ Straight line fit: Data $y_i(x_i) \Rightarrow$ parameters $\theta_0, \theta_1 \Rightarrow$ Trajectory $y(x)$

- Fit the curvature κ of a track flying through perpendicular magnetic field