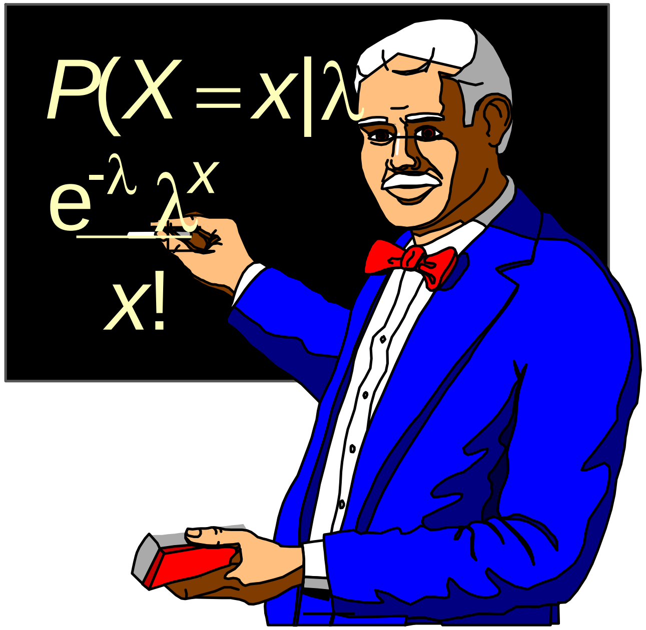
Phys 443
Computational Physics

Fitting

ML & Least Squares Methods



ML Linear Fit

- N experimental points have been taken. Each point is the measurement of a physical quantity y_i, i=1,...,N for N different values of another physical quantity x_i. We make the following assumptions:
  - each measurement of y_i is characterized by a *Gaussian* pdf with a known variance σ^2_i ;
  - the x_i values are assumed to be known with **no or negligible uncertainty**;
  - the y_i measurements are not correlated;
  - we make the hypothesis that the two physics quantities y and x are related by

$$y = \alpha x + \beta$$

- So, a set of measurements: $(x_1, y_1 \pm \sigma_1)$, $(x_2, y_2 \pm \sigma_2)$, … $(x_n, y_n \pm \sigma_n)$ and the points are thought to come from a straight line. Find α and β

The likelihood function is:

$$L = \prod_{i=1}^{n} f(x_i, \alpha, \beta) = \prod_{i=1}^{n} \frac{1}{\sigma_i \cdot \sqrt{2\pi}} e^{-\frac{(y_i - q(x_i, \alpha, \beta))^2}{2\sigma_i^2}} = \prod_{i=1}^{n} \frac{1}{\sigma_i \cdot \sqrt{2\pi}} e^{-\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_i^2}}$$

Maximum Likelihood Method

Suppose we have a set of n measurements:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{\partial}{\partial \alpha}\sum_{i=1}^{n}\left[\ln\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right) - \frac{(y_i-\alpha-\beta x_i)^2}{2\sigma_i^2}\right] = \sum_{i=1}^{n}\left[-\frac{2(y_i-\alpha-\beta x_i)(-1)}{2\sigma_i^2}\right] = 0$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{\partial}{\partial \beta}\sum_{i=1}^{n}\left[\ln\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right) - \frac{(y_i-\alpha-\beta x_i)^2}{2\sigma_i^2}\right] = \sum_{i=1}^{n}\left[-\frac{2(y_i-\alpha-\beta x_i)(-x_i)}{2\sigma_i^2}\right] = 0$$

Assume all σ's are the same for simplicity:

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\alpha - \sum_{i=1}^{n}\beta x_i = 0$$

$$\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n}\alpha x_i - \sum_{i=1}^{n}\beta x_i^2 = 0$$

We now have two equations that are linear in the two unkowns α and β

$$\sum_{i=1}^{n} y_i = n\alpha + \beta\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} y_i x_i = \alpha\sum_{i=1}^{n} x_i + \beta\sum_{i=1}^{n} x_i^2$$

Maximum Likelihood Method

$$\begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_i \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$\alpha = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i x_i \sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \quad \text{and} \quad \beta = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

- EXAMPLE:
  - A trolley moves along a track at constant speed. Suppose the following measurements of the time vs. distance were made. From the data find the best value for the speed (v) of the trolley.

Time t (seconds)	1.0	2.0	3.0	4.0	5.0	6.0
Distance d (mm)	11	19	33	40	49	61

  - Our model of the motion of the trolley tells us that:

$$d = d_0 + vt$$

Maximum Likelihood Method

We want to find v, the slope (β) of the straight line describing the motion of the trolley.
We need to evaluate the sums listed in the above formula:

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{6} t_i = 21\ s$$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{6} d_i = 213\ mm$$

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{6} t_i d_i = 919\ s \cdot mm$$

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{6} t_i^2 = 91\ s^2$$

$$v = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{6\times 919 - 21\times 213}{6\times 91 - 21^2} = 9.9\, mm/s$$

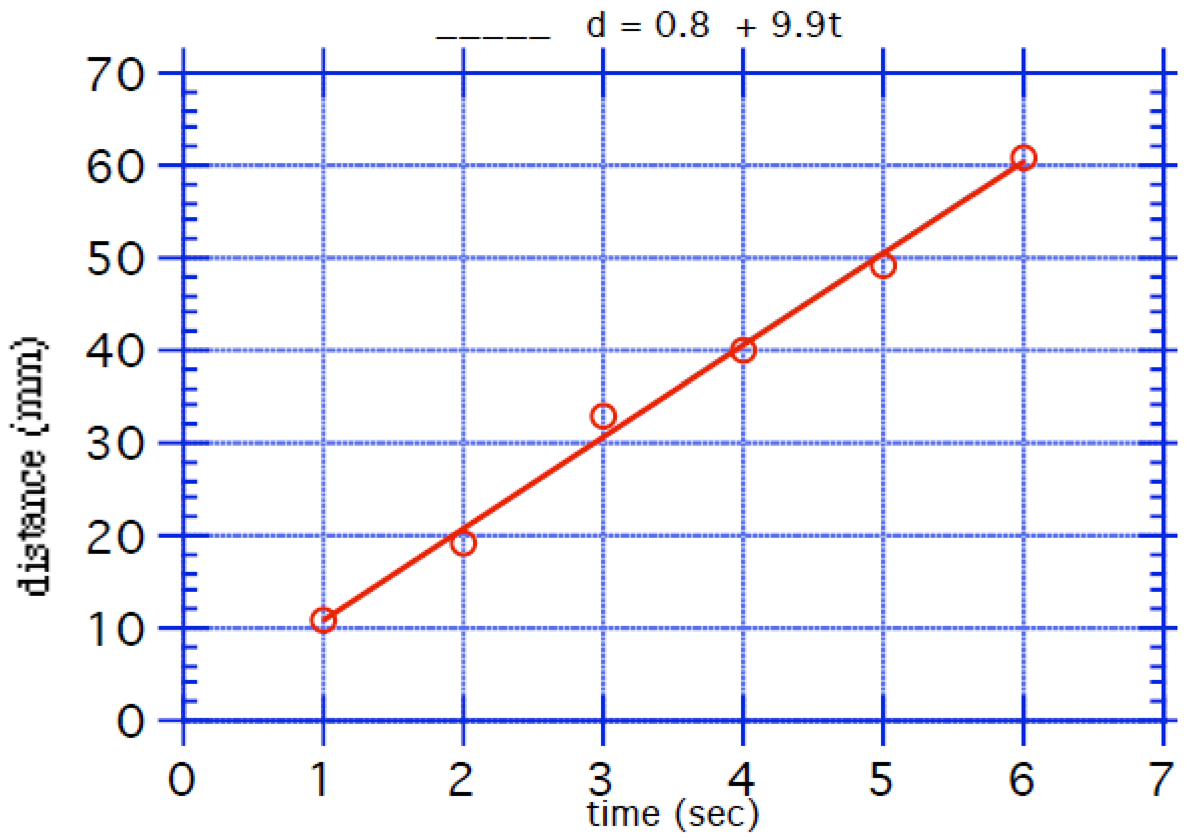
best estimate of the speed

$$d_0 = 0.8\ mm$$

best estimate of the starting point

Maximum Likelihood Method

Exercise



Determining the Slope and Intercept with MLM
σ's are NOT the same

- If all σ's are NOT the same :

$$L=\prod_{i=1}^{n} f(x_i,\alpha,\beta)=\prod_{i=1}^{n}\frac{1}{\sigma_i\sqrt{2\pi}}e^{-\frac{(y_i-q(x_i,\alpha,\beta))^2}{2\sigma_i^2}}=\prod_{i=1}^{n}\frac{1}{\sigma_i\sqrt{2\pi}}e^{-\frac{(y_i-\alpha-\beta x_i)^2}{2\sigma_i^2}}$$

We wish to find the α and β that maximizes the likelihood function L. Thus we need to take some derivatives:

$$\frac{\partial \ln L}{\partial\alpha}=\frac{\partial}{\partial\alpha}\sum_{i=1}^{n}\left[\ln\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)-\frac{(y_i-\alpha-\beta x_i)^2}{2\sigma_i^2}\right]=\sum_{i=1}^{n}\left[-\frac{2(y_i-\alpha-\beta x_i)(-1)}{2\sigma_i^2}\right]=0$$

$$\frac{\partial \ln L}{\partial\beta}=\frac{\partial}{\partial\beta}\sum_{i=1}^{n}\left[\ln\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)-\frac{(y_i-\alpha-\beta x_i)^2}{2\sigma_i^2}\right]=\sum_{i=1}^{n}\left[-\frac{2(y_i-\alpha-\beta x_i)(-x_i)}{2\sigma_i^2}\right]=0$$

$$\frac{\partial \ln L}{\partial\alpha}=\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}-\alpha\sum_{i=1}^{n}\frac{1}{\sigma_i^2}-\beta\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}=0$$

$$\frac{\partial \ln L}{\partial\beta}=\sum_{i=1}^{n}\frac{x_i y_i}{\sigma_i^2}-\alpha\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}-\beta\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}=0$$

$$\Longrightarrow\quad \begin{pmatrix}\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}\\ \sum_{i=1}^{n}\frac{x_i y_i}{\sigma_i^2}\end{pmatrix}=\begin{pmatrix}\sum_{i=1}^{n}\frac{1}{\sigma_i^2} & \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\\ \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$$

- We have to solve the two equations for the two unknowns, α and β . We can get an exact solution since these equations are linear in α and β. Just have to invert a matrix.

Determining the Slope and Intercept with MLM

$$\alpha = \frac{\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} - \sum_{i=1}^{n}\frac{y_i x_i}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} - (\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2})^2} \quad \text{and} \quad \beta = \frac{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i y_i}{\sigma_i^2} - \sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} - (\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2})^2}$$

Let's calculate the error (covariance) matrix for α and β:

$$V_{ij} = -\left(\frac{\partial^2 \ln L}{\partial\alpha_i \partial\alpha_j}\right)^{-1}$$

$$\left.\begin{aligned}
\frac{\partial^2 \ln L}{\partial\alpha^2} &= \frac{\partial}{\partial\alpha}\left(\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2} - \alpha\sum_{i=1}^{n}\frac{1}{\sigma_i^2} - \beta\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\right) = -\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\\
\frac{\partial^2 \ln L}{\partial\beta^2} &= \frac{\partial}{\partial\beta}\left(\sum_{i=1}^{n}\frac{x_i y_i}{\sigma_i^2} - \alpha\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} - \beta\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}\right) = -\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}\\
\frac{\partial^2 \ln L}{\partial\alpha\partial\beta} &= \frac{\partial}{\partial\beta}\left(\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2} - \alpha\sum_{i=1}^{n}\frac{1}{\sigma_i^2} - \beta\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\right) = -\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}
\end{aligned}\right\} \quad V^{-1} = (-1)\begin{pmatrix} \sum_{i=1}^{n}\frac{1}{\sigma_i^2} & \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} \end{pmatrix}$$

Covariance is the measure the strength of the linear relationship between two random variables

Reminder: Inverse of a Matrix

Finding the Inverse of a 2x2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

Step-1 First find what is called the Determinant

This is calculated as ad-bc

Step-2 Then swap the elements in the leading diagonal $\begin{bmatrix} d & b \\ c & a \end{bmatrix}$

Step-3 Then negate the other elements $\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Step-4 Then multiply the Matrix by 1/determinant

$$\frac{1}{ad-cb}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Determining the Slope and Intercept with MLM

$$V = \begin{pmatrix} \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} / D & -\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} / D \\ -\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} / D & \sum_{i=1}^{n} \frac{1}{\sigma_i^2} / D \end{pmatrix} \text{ with } D = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}\right)^2$$

$$\boxed{V = \begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{pmatrix}}$$

Note: We could also derive the variance of α and β just using propagation of errors on the formulas for α and β.

Chi Square Distribution (χ^2) and Least Squares Fitting

Problem: We have some measurements and would like some way to measure how "good" these measurements really are.
Solution: Consider calculating the "χ^2" ("chi-square")

Assume:

- We have a set of measurements $\{x_1, x_2, \dots x_n\}$.
- We know the true value of each x_i ($x_{t1}, x_{t2}, \dots x_{tn}$).
- Obviously the closer the ($x_1, x_2, \dots x_n$)'s are to the ($x_{t1}, x_{t2}, \dots x_{tn}$)'s the better (or more accurate) the measurements.

Can we get more specific? Can we put a number (or probability) on how well they agree?

Assume:

- The measurements are independent of each other.
- The measurements come from a Gaussian distribution.
- Let ($\sigma_1, \sigma_2 \dots \sigma_n$) be the standard deviation associated with each measurement.

Chi Square Distribution (χ^2) and Least Squares Fitting

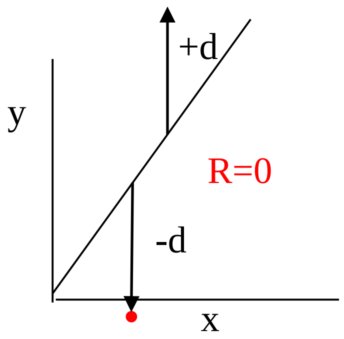
- Consider the following two possible measures of the quality of the data:

$$R \equiv \sum_{i=1}^{n} \frac{x_i - x_{ti}}{\sigma_i} = \sum_{i=1}^{n} \frac{d_i}{\sigma_i}$$

$$\chi^2 \equiv \sum_{i=1}^{n} \frac{(x_i - x_{ti})^2}{\sigma_i^2} = \sum_{i=1}^{n} \frac{d_i^2}{\sigma_i^2}$$

Both measures give zero when the measurements are identical to the true values.

- Which of the above gives more information on the quality of the data?
- Both R and χ^2 are zero if the measurements agree with the true value.
  - R looks good because via the Central Limit Theorem as $n \rightarrow \infty$ the sum $\rightarrow$ Gaussian.
  - **However, χ^2 is better!**



Chi Square Distribution (χ^2) and Least Squares Fitting

One can show (derive) that the probability distribution function for χ^2 is exactly:

$$p(\chi^2, n) = \frac{1}{2^{n/2}\Gamma(n/2)}[\chi^2]^{n/2-1} e^{-\chi^2/2} \quad 0 \le \chi^2 \le \infty$$

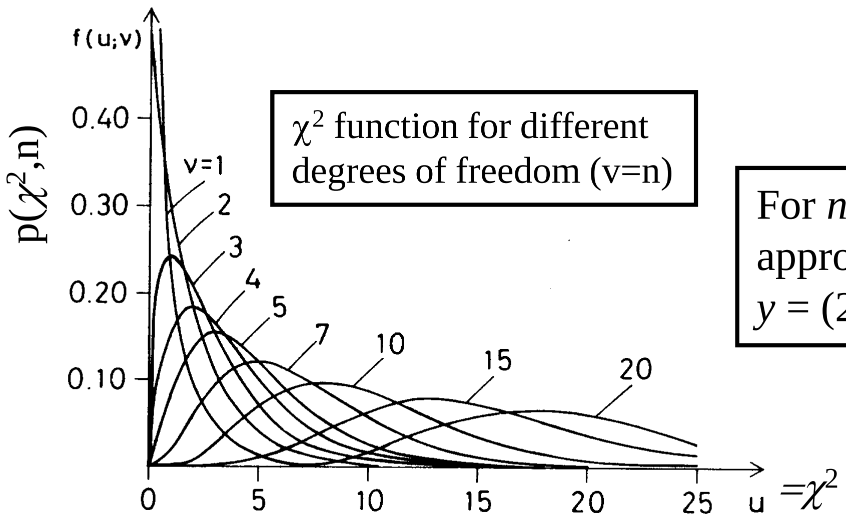
This is called the "Chi Square" (χ^2) probability distribution function.

Γ is the “Gamma Function”.

$$\Gamma(x) \equiv \int_0^\infty e^{-t} t^{x-1} dt \quad x > 0$$
$$\Gamma(n+1) = n! \quad (n = 1,2,3\ldots)$$
$$\Gamma(1/2) = \sqrt{\pi}$$

Chi Square Distribution (χ^2) and Least Squares Fitting

- This is a continuous probability distribution that is a function of two variables:
  χ^2 and n = number of degrees of freedom (DOF).
  ***DOF*=*n* = (# of data points) – (# of parameters calculated from the data points)**



χ^2 function for different degrees of freedom (v=n)

For $n \geq 20$, $P(\chi^2 > y)$ can be approximated using a Gaussian *pdf* with $y = (2\chi^2)^{1/2} - (2n-1)^{1/2}$

Least Squares Fitting

- Suppose we have n data points (x_i, y_i, σ_i).
  - Assume that we know a functional relationship between the points,

$$y = f(x, a, b \ldots)$$

  - Assume that for each y_i we know x_i exactly.
  - The parameters a, b, … are constants that we wish to determine from out data.
- A procedure to obtain a and b is to minimize the following χ^2 with respect to a and b.

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

- This is very similar to the Maximum Likelihood Method.

For the Gaussian case MLM and LS are identical.

- Technically this is a χ^2 distribution only if the y's are from a Gaussian distribution.
- Since most of the time the y's are not from a Gaussian we call it "least squares" rather than χ^2.

Least Squares Fitting

Example: We have a function with one unknown parameter:

$$f(x,b) = 1 + bx$$

Find *b* using the least squares technique.

- We need to minimize the following:

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

- To find the *b* that minimizes the above function, we do the following:

$$\frac{\partial \chi^2}{\partial b} = \frac{\partial}{\partial b} \sum_{i=1}^{n} \frac{[y_i - 1 - bx_i]^2}{\sigma_i^2} = \sum_{i=1}^{n} \frac{-2[y_i - 1 - bx_i]x_i}{\sigma_i^2} = 0$$

$$\sum_{i=1}^{n} \frac{y_i x_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{b x_i^2}{\sigma_i^2} = 0$$

Solving for b we find:

$$b = \frac{\sum_{i=1}^{n} \frac{y_i x_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}}$$

Least Squares Fitting

- Here each measured data point (y_i) is allowed to have a different standard deviation (σ_i).
- The LS technique can be generalized to two or more parameters for simple and complicated (e.g. non-linear) functions.
- One especially nice case is a polynomial function that is linear in the unknowns (a_i):

$$f(x, a_1 \ldots a_n) = a_1 + a_2 x + a_3 x^2 + a_n x^{n-1}$$

- We can always recast problem in terms of solving *n* simultaneous linear equations.
- We use the techniques from linear algebra and invert an *n* x *n* matrix to find the a_i's!

- Example: Given the following data perform a least squares fit to find the value of *b*.

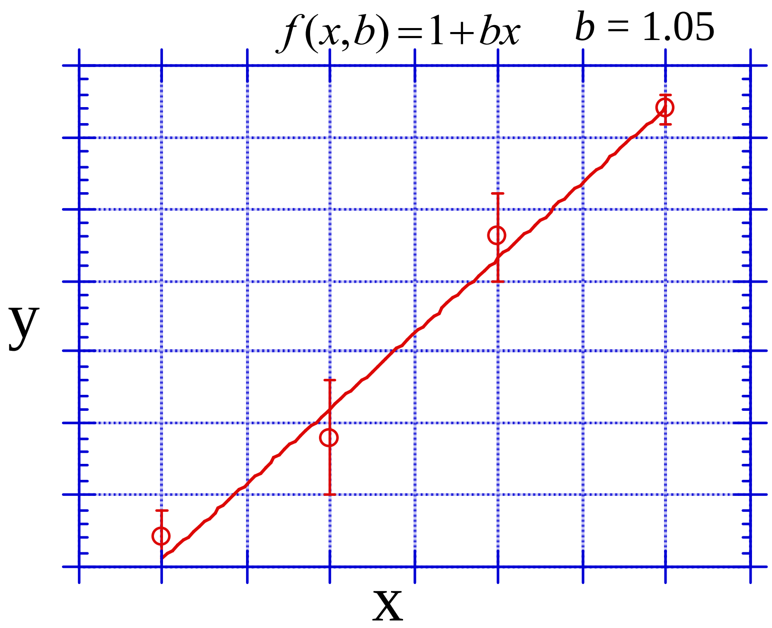
x	1.0	2.0	3.0	4.0
y	2.2	2.9	4.3	5.2
σ	0.2	0.4	0.3	0.1

- Using the above expression for *b* we calculate:

b = 1.05

Least Squares Fitting

- Next is a plot of the data points and the line from the least squares fit:



Least Squares Fitting

- If we assume that the data points are from a Gaussian distribution,
- we can calculate a χ^2 and the probability associated with the fit.

$$\chi^2 = \sum_{i=1}^{n} \frac{[y_i - 1 - 1.05x_i]^2}{\sigma_i^2} = \left(\frac{2.2-2.05}{0.2}\right)^2 + \left(\frac{2.9-3.1}{0.4}\right)^2 + \left(\frac{4.3-4.16}{0.3}\right)^2 + \left(\frac{5.2-5.2}{0.1}\right)^2 = 1.04$$

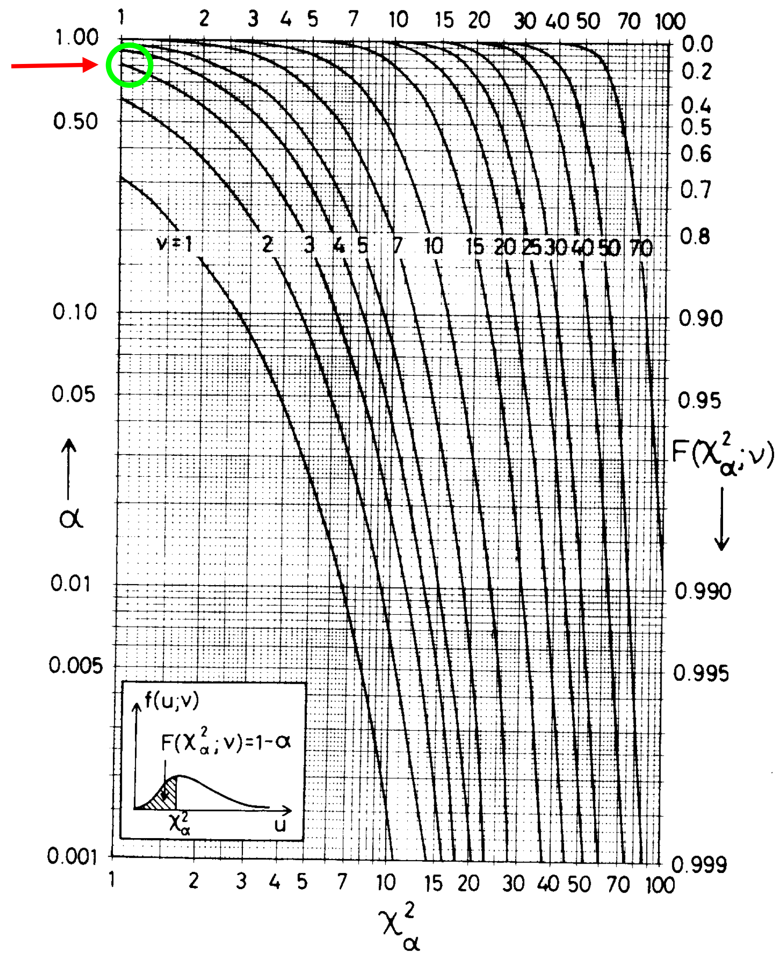
The probability to get $\chi^2 \geq 1.04$ for 3 degrees of freedom $\approx$ 80%.
We call this a "good" fit since the probability is close to 100%.

Thus we say that the probability of getting $\chi^2 \geq 1.04$ with 3 degrees of freedom by chance is about 80%

If however the χ^2 was large (e.g. 15),
the probability would be small ($\approx$ 0.2% for 3 dof).
We would say this was a "bad" fit.

RULE OF THUMB
A "good" fit has χ^2 /dof ≤ 1

TABLE D



1 2 3 4 5 7 10 15 20 30 40 50 70 100
1.00
0.50
0.10
0.05
α
0.01
0.005
0.001
0.0
0.2
0.4
0.5
0.6
0.7
0.8
0.90
0.95
F(χ²α;ν)
0.990
0.995
0.999
ν=1 2 3 4 5 7 10 15 20 25 30 40 50 70
f(u;ν)
F(χ²α;ν)=1-α
χ²α
u
1 2 3 4 5 7 10 15 20 30 40 50 70 100
χ²α

Chi-Square: fit *goodness*

- However, often times the y_i's are NOT from a gaussian *pdf*. In these instances we call this technique "χ^2 fitting" or "Least Squares Fitting".

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

- We can only use a χ^2 probability table when *y* is from a gaussian *pdf*. However, there are many instances where even for non-gaussian *pdf*'s the above sum approximates χ^2 *pdf*.

More on Least Squares Fit (LSQF)

We discussed how we can fit our data points to a linear function (straight line) and get the "best" estimate of the slope and intercept.
However, we did not discuss two important issues:

I) How to estimate the uncertainties on our slope and intercept obtained from a LSQF?

II) How to apply the LSQF when we have a non-linear function?

Estimation of Errors on parameters determined from a LSQF

Assume we have data points that lie on a straight line:

$$y = \alpha + \beta x$$

Assume we have n measurements of the x's and y's.

For simplicity, assume that each y measurement has the same error, σ.

Assume that x is known much more accurately than y.

$\Rightarrow$ ignore any uncertainty associated with x.

More on Least Squares Fit (LSQF)

Previously we showed that the solution for the intercept α and slope β is:

$$\alpha = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \text{ and } \beta = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

Since α and β are functions of the measurements (y_i's) we can use the Propagation of Errors technique to estimate σ_α and σ_β.

$$\sigma_Q^2 = \sigma_x^2 \left(\frac{\partial Q}{\partial x}\right)^2 + \sigma_y^2 \left(\frac{\partial Q}{\partial y}\right)^2 + 2\sigma_{xy}\left(\frac{\partial Q}{\partial x}\right)\left(\frac{\partial Q}{\partial y}\right)$$

Assume that: a) Each measurement is independent of each other ($\sigma_{xy}=0$).
b) We can neglect any error (σ) associated with x.
c) Each y measurement has the same σ, i.e. $\sigma_i=\sigma$.

23

More on Least Squares Fit (LSQF)

$$\sigma_Q^2 = \sigma_x^2\left(\frac{\partial Q}{\partial x}\right)^2 + \sigma_y^2\left(\frac{\partial Q}{\partial y}\right)^2$$

assumption a)

$$\sigma_Q^2 = \sigma_y^2\left(\frac{\partial Q}{\partial y}\right)^2$$

assumption b)

$$\sigma_\alpha^2 = \sum_{i=1}^{n}\sigma_{y_i}^2\left(\frac{\partial \alpha}{\partial y_i}\right)^2 = \sigma^2\sum_{i=1}^{n}\left(\frac{\partial \alpha}{\partial y_i}\right)^2$$

assumption c)

$$\frac{\partial \alpha}{\partial y_i} = \frac{\partial}{\partial y_i}\frac{\sum_{i=1}^{n} y_i \sum_{j=1}^{n} x_j^2 - \sum_{i=1}^{n} x_i y_i \sum_{j=1}^{n} x_j}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{\sum_{j=1}^{n} x_j^2 - x_i \sum_{j=1}^{n} x_j}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

plugging our formula for α

$$\sigma_\alpha^2 = \sigma^2\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{n} x_j^2 - x_i \sum_{j=1}^{n} x_j}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\right)^2 = \sigma^2\sum_{i=1}^{n}\left(\frac{(\sum_{j=1}^{n} x_j^2)^2 + x_i^2(\sum_{j=1}^{n} x_j)^2 - 2x_i\sum_{j=1}^{n} x_j \sum_{j=1}^{n} x_j^2}{(n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)^2}\right)$$

More on Least Squares Fit (LSQF)

$$\sigma_\alpha^2 = \sigma^2 \frac{n(\sum_{j=1}^{n} x_j^2)^2 + \sum_{i=1}^{n} x_i^2 (\sum_{j=1}^{n} x_j)^2 - 2(\sum_{j=1}^{n} x_j)^2 \sum_{j=1}^{n} x_j^2}{(n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)^2} = \sigma^2 \frac{n(\sum_{j=1}^{n} x_j^2)^2 - \sum_{i=1}^{n} x_i^2 (\sum_{j=1}^{n} x_j)^2}{(n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)^2}$$

$$= \sigma^2 \sum_{j=1}^{n} x_j^2 \frac{n\sum_{j=1}^{n} x_j^2 - (\sum_{j=1}^{n} x_j)^2}{(n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)^2}$$

$$\boxed{\sigma_\alpha^2 = \sigma^2 \frac{\sum_{j=1}^{n} x_j^2}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}}$$

variance in the intercept

More on Least Squares Fit (LSQF)

We can find the variance in the slope (β) using exactly the same procedure:

$$\sigma_\beta^2 = \sum_{i=1}^{n} \sigma_{y_i}^2 \left(\frac{\partial \beta}{\partial y_i}\right)^2 = \sigma^2 \sum_{i=1}^{n} \left(\frac{\partial \beta}{\partial y_i}\right)^2 = \sigma^2 \sum_{i=1}^{n} \left(\frac{n x_i - \sum_{j=1}^{n} x_j}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\right)^2$$

$$= \sigma^2 \frac{n^2 \sum_{j=1}^{n} x_j^2 + n(\sum_{j=1}^{n} x_j)^2 - 2n \sum_{i=1}^{n} x_i \sum_{j=1}^{n} x_j}{(n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)^2}$$

More on Least Squares Fit (LSQF)

$$\sigma_\beta^2 = \frac{n\sigma^2}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

variance in the slope

- If we don't know the true value of σ, we can estimate the variance using the spread between the measurements (y_i's) and the fitted values of y:

$$\sigma^2 \approx \frac{1}{n-2}\sum_{i=1}^{n}(y_i - y_i^{fit})^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$$

$n-2$ = number of degree of freedom

= number of data points $-$ number of parameters (α, β) extracted from the data

- If each y_i measurement has a different error σ_i:

$$\sigma_\alpha^2 = \frac{1}{D}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}$$

$$\sigma_\beta^2 = \frac{1}{D}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}$$

$$D = \sum_{i=1}^{n}\frac{1}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} - (\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2})^2$$

weighted slope and intercept

More on Least Squares Fit (LSQF)

The above expressions simplify to the “equal variance” case.
Don't forget to keep track of the “n’s” when factoring out equal σ’s. For example:

$$\sum_{i=1}^{n}\frac{1}{\sigma_i^2}=\frac{n}{\sigma^2} \quad not \quad \frac{1}{\sigma^2}$$

LSQF with non-linear functions:

Example: Decay of a radioactive substance. Fit the following data to find N_0 and τ:

$$N(t) = N_0 e^{-t/\tau}$$

- N represents the amount of the substance present at time t.
- N_0 is the amount of the substance at the beginning of the experiment ($t = 0$).
- τ is the lifetime of the substance.

i	1	2	3	4	5	6	7	8	9	10
t_i	0	15	30	45	60	75	90	105	120	135
N_i	106	80	98	75	74	73	49	38	37	22
$y_i = \ln N_i$	4.663	4.382	4.585	4.317	4.304	4.290	3.892	3.638	3.611	3.091

Least Squares Fitting Example

$$N(t) = N(0)e^{-t/\tau}$$

As written the above *pdf* is not linear in τ. We can turn this into a linear problem by taking the natural log of both sides of the *pdf*.

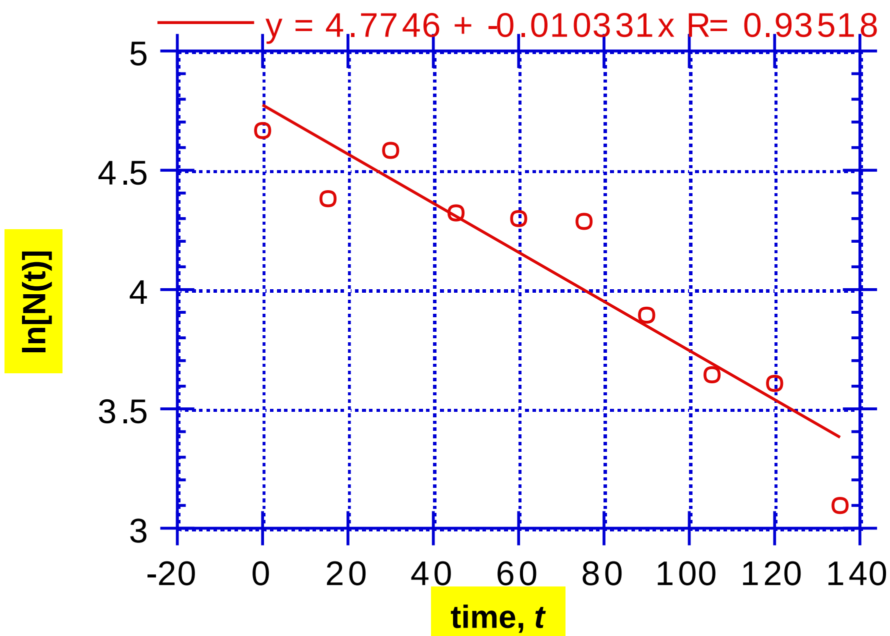
$$\ln(N(t)) = \ln(N(0)) - t/\tau \Rightarrow y = C - Dt$$

LSQF with non-linear functions:

$$D = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{10 \times 2560.41 - 40.773 \times 675}{10 \times 64125 - (675)^2} = -0.01033$$

$$\tau = -1/D = 96.80 \text{ seconds}$$

- The intercept is given by: $C = 4.77 = \ln A$ or $A = 117.9$



LSQF with non-linear functions:

Example: Find the values A and τ taking into account the uncertainties in the data points.

- The uncertainty in the number of radioactive decays is governed by Poisson statistics.
- The number of counts N_i in a bin is assumed to be the average (μ) of a Poisson distribution: $\mu = N_i$ = Variance
- The variance of y_i(= ln N_i) can be calculated using propagation of errors:

$$\sigma_y^2 = \sigma_N^2 (\partial y/\partial N)^2 = (N)(\partial \ln N/\partial N)^2 = (N)(1/N)^2 = 1/N$$

- The slope and intercept from a straight line fit that includes uncertainties in the data points:

$$\alpha = \frac{\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} - \sum_{i=1}^{n}\frac{x_i y_i}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} - (\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2})^2} \text{ and } \beta = \frac{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i y_i}{\sigma_i^2} - \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} - (\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2})^2}$$

If all the σ's are the same then the above expressions are identical to the unweighted case.

α=4.725 and β = -0.00903

τ = -1/β = -1/0.00903 = 110.7 sec

LSQF with non-linear functions:

- To calculate the error on the lifetime (t), we first must calculate the error on b:

$$\sigma_\beta^2 = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} - (\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2})^2} = \frac{652}{652 \times 2684700 - (33240)^2} = 1.01 \times 10^{-6}$$

$$\sigma_\tau^2 = \sigma_\beta^2 (\partial\tau / \partial\beta)^2 \Rightarrow \sigma_\tau = \sigma_\beta \left(1/\beta^2\right) = \frac{1.005 \times 10^{-3}}{(9.03 \times 10^{-3})^2} = 12.3$$

The experimentally determined lifetime is: $\tau = 110.7 \pm 12.3$ sec.

Least Squares Fitting Example

We can calculate the χ^2 to see how "good" the data fits an exponential decay distribution:

For this problem: $\ln A = 4.725 \rightarrow A = 112.73$ and $\tau = 110.7$ sec

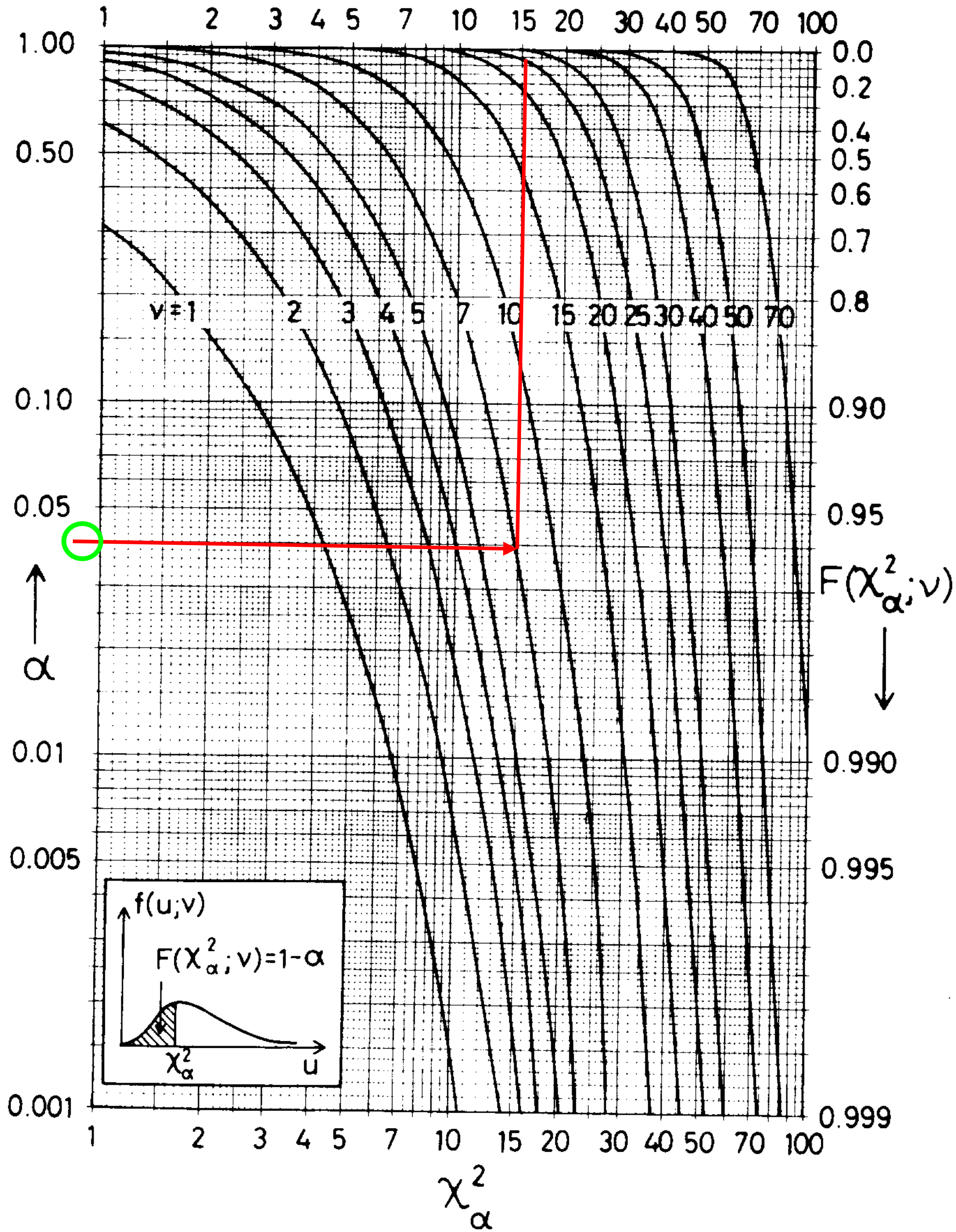
$$\chi^2 = \sum_{i=1}^{10} \frac{(N_i - Ae^{-t_i/\tau})^2}{\sigma_i^2} = \sum_{i=1}^{10} \frac{(N_i - Ae^{-t_i/\tau})^2}{Ae^{-t_i/\tau}} = 15.6$$

Poisson approximation

A good fit has $\chi^2/\text{DOF} \leq 1$

The chi sq per dof is 1.96
The chi sq prob. is 4.9 %

This is not such a good fit since the probability is only ~4.9%.



1 2 3 4 5 7 10 15 20 30 40 50 70 100
1.00
0.50
0.10
0.05
0.01
0.005
0.001
0.0
0.2
0.4
0.5
0.6
0.7
0.8
0.90
0.95
0.990
0.995
0.999
ν = 1 2 3 4 5 7 10 15 20 25 30 40 50 70
α
$F(\chi^2_\alpha; \nu)$
$f(u;\nu)$
$F(\chi^2_\alpha; \nu) = 1-\alpha$
χ^2_α
u
χ^2_α

Least Squares Fitting: quadratic function

- Let's now allow a curved line of polynomial form

$$y(x) = a + bx + cx^2 + dx^3 + \ldots$$

which is the curve we want to fit to the data.

- For simplicity, let's consider a second-degree polynomial (quadratic). The chi-square for this situation is

$$\chi^2 = \sum\left[\frac{y_i - y(x)}{\sigma_i}\right]^2 = \sum\left[\frac{1}{\sigma_i}\left(y_i - a - bx - cx^2\right)\right]^2$$

- Following exactly the same approach as before, we end up with three equations in three unknowns (the parameters a, b and c):

$$\frac{\partial}{\partial a}\chi^2 = \frac{\partial}{\partial a}\sum\left[\frac{1}{\sigma_i}\left(y_i - a - bx_i - cx_i^2\right)\right]^2 = -2\sum\left[\frac{1}{\sigma_i^2}\left(y_i - a - bx_i - cx_i^2\right)\right] = 0,$$

$$\frac{\partial}{\partial b}\chi^2 = \frac{\partial}{\partial b}\sum\left[\frac{1}{\sigma_i}\left(y_i - a - bx_i - cx_i^2\right)\right]^2 = -2\sum\left[\frac{x_i}{\sigma_i^2}\left(y_i - a - bx_i - cx_i^2\right)\right] = 0,$$

$$\frac{\partial}{\partial c}\chi^2 = \frac{\partial}{\partial c}\sum\left[\frac{1}{\sigma_i}\left(y_i - a - bx_i - cx_i^2\right)\right]^2 = -2\sum\left[\frac{x_i^2}{\sigma_i^2}\left(y_i - a - bx_i - cx_i^2\right)\right] = 0.$$

Least Squares Fitting: quadratic function

- The solution, then, can be found from the same determinant technique we used before, except now we have 3 x 3 determinants:

$$a = \frac{1}{\Delta}\begin{vmatrix} \sum \frac{y_i}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \\ \sum \frac{x_i y_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} & \sum \frac{x_i^3}{\sigma_i^2} \\ \sum \frac{x_i^2 y_i}{\sigma_i^2} & \sum \frac{x_i^3}{\sigma_i^2} & \sum \frac{x_i^4}{\sigma_i^2} \end{vmatrix}, \qquad b = \frac{1}{\Delta}\begin{vmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{y_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i y_i}{\sigma_i^2} & \sum \frac{x_i^3}{\sigma_i^2} \\ \sum \frac{x_i^2}{\sigma_i^2} & \sum \frac{x_i^2 y_i}{\sigma_i^2} & \sum \frac{x_i^4}{\sigma_i^2} \end{vmatrix}$$

$$c = \frac{1}{\Delta}\begin{vmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} & \sum \frac{y_i}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} & \sum \frac{x_i y_i}{\sigma_i^2} \\ \sum \frac{x_i^2}{\sigma_i^2} & \sum \frac{x_i^3}{\sigma_i^2} & \sum \frac{x_i^2 y_i}{\sigma_i^2} \end{vmatrix}, \quad \text{where } \Delta = \begin{vmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} & \sum \frac{x_i^3}{\sigma_i^2} \\ \sum \frac{x_i^2}{\sigma_i^2} & \sum \frac{x_i^3}{\sigma_i^2} & \sum \frac{x_i^4}{\sigma_i^2} \end{vmatrix}$$

- You can see that extending to arbitrarily high powers is straightforward

Example

Superimpose the 2nd order polnomial to the given data

$$p(t) = \alpha + \beta t + \gamma t^2$$

Minimizing

$$F(\alpha, \beta, \gamma) = \sum_{i=1}^{10} (\alpha + \beta t_i + \gamma t_i^2 - y_i)^2$$

$$\frac{\partial F}{\partial \alpha} = \frac{\partial F}{\partial \beta} = \frac{\partial F}{\partial \gamma} = 0$$

Calendar year	Computational year t_i	Temperature deviation y_i
1991	1	0.29
1992	2	0.14
1993	3	0.19
1994	4	0.26
1995	5	0.28
1996	6	0.22
1997	7	0.43
1998	8	0.59
1999	9	0.33
2000	10	0.29

The global annual mean temperature deviation measured in ^{o}C for years 1991-2000

Example

$$\frac{\partial F}{\partial \alpha} = 2\sum_{i=1}^{10}\left(\alpha + \beta t_i + \gamma t_i^2 - y_i\right) = 0$$

$$10\alpha + \left(\sum_{i=1}^{10} t_i\right)\beta + \left(\sum_{i=1}^{10} t_i^2\right)\gamma = \sum_{i=1}^{10} y_i$$

$$\frac{\partial F}{\partial \beta} = 2\sum_{i=1}^{10}\left(\alpha + \beta t_i + \gamma t_i^2 - y_i\right) t_i = 0$$

$$\left(\sum_{i=1}^{10} t_i\right)\alpha + \left(\sum_{i=1}^{10} t_i^2\right)\beta + \left(\sum_{i=1}^{10} t_i^3\right)\gamma = \sum_{i=1}^{10} y_i t_i$$

Example

$$\frac{\partial F}{\partial \gamma} = 2\sum_{i=1}^{10} \left(\alpha + \beta t_i + \gamma t_i^2 - y_i\right) t_i^2 \ = \ 0$$

$$\left(\sum_{i=1}^{10} t_i^2\right) \alpha + \left(\sum_{i=1}^{10} t_i^3\right) \beta + \left(\sum_{i=1}^{10} t_i^4\right) \gamma \ = \ \sum_{i=1}^{10} y_i t_i^2$$

$$\sum_{i=1}^{10} t_i = 55, \qquad \sum_{i=1}^{10} t_i^2 = 385, \quad \sum_{i=1}^{10} t_i^3 = 3025,$$

$$\sum_{i=1}^{10} t_i^4 = 25330, \quad \sum_{i=1}^{10} y_i = 3.12, \quad \sum_{i=1}^{10} t_i y_i = 20,$$

$$\sum_{i=1}^{10} t_i^2 y_i = 138.7,$$

Example

Which leads to linear system

$$\begin{pmatrix} 10 & 55 & 385 \\ 55 & 385 & 3025 \\ 385 & 3025 & 25330 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 3.12 \\ 20 \\ 138.7 \end{pmatrix}$$

Solving the linear system

$$\begin{aligned} \alpha &\approx -0.4078, \\ \beta &\approx 0.2997, \\ \gamma &\approx -0.0241. \end{aligned}$$

$$F(t) = -0.4078 + 0.2997t - 0.0241t^2$$

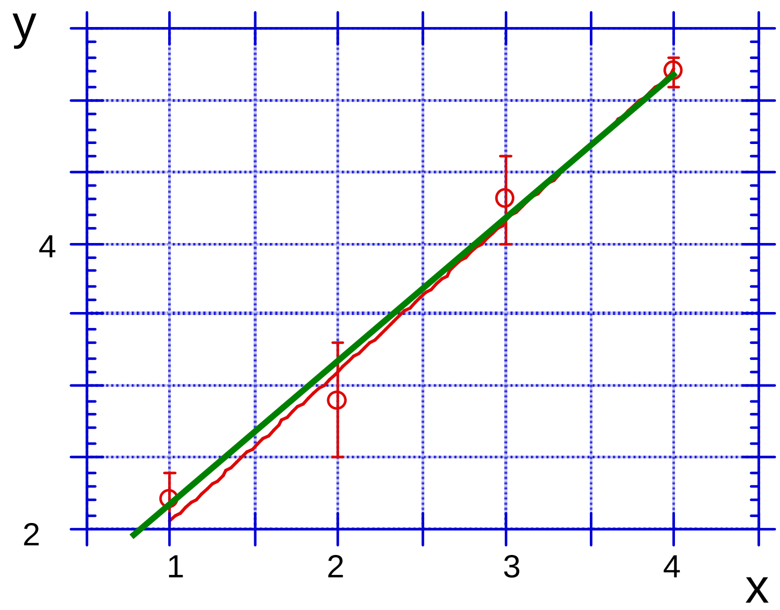
In Class Exercise

Write a python code which does the straigth line fit to the given data.

- Calculate fit parameters (α and β) using Least Squares
- Calculate error on the fit parameters
- Superimose the line on the data points
- Calculate the χ^2 of the fit, is it good/bad fit ?

x	1.0	2.0	3.0	4.0
y	2.2	2.9	4.3	5.2
σ	0.2	0.4	0.3	0.1

In Class Exercise



χ^2=0.65 for 2 degrees of freedom
P(χ^2>0.65 for 2 dof) = 72%   ''good fit''

Back up

The Error on the Mean

Question: If we have a set of measurements of the same quantity:

$$x_1 \pm \sigma_1 \;\; x_2 \pm \sigma_2 \ldots x_n \pm \sigma_n$$

What's the best way to combine these measurements?

How to calculate the variance once we combine the measurements?

Assuming Gaussian statistics, the Maximum Likelihood Method says combine the measurements as:

$$x = \frac{\sum_{i=1}^{n} x_i / \sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}$$

weighted average

If all the variances ($\sigma^2{}_1 = \sigma^2{}_2 = \sigma^2{}_3 ..$) are the same:

$$x = \frac{1}{n} \sum_{i=1}^{n} x_i$$

unweighted average

The variance of the weighted average can be calculated using propagation of errors:

The Error on the Mean

The variance of the weighted average can be calculated using propagation of errors:

$$\sigma_x^2 = \sum_{i=1}^{n} \left[\frac{\partial}{\partial x_i} x \right]^2 \sigma_i^2$$

And the derivatives are:

$$\frac{\partial}{\partial x_i} x = \frac{1/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}$$

$$\sigma_x^2 = \sum_{i=1}^{n} \left[\frac{\partial}{\partial x_i} x \right]^2 \sigma_i^2 = \sum_{i=1}^{n} \frac{1/\sigma_i^4}{\left[\sum_{i=1}^{n} 1/\sigma_i^2 \right]^2} \sigma_i^2 = \frac{1}{\left[\sum_{i=1}^{n} 1/\sigma_i^2 \right]^2} \sum_{i=1}^{n} 1/\sigma_i^2 = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

$$\boxed{\sigma_x^2 = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}}$$

σ_x=error in the weighted mean

The Error on the Mean

If all the variances are the same: $\sigma_x^2 = \dfrac{1}{\sum_{i=1}^{n} 1/\sigma_i^2} = \dfrac{\sigma^2}{n}$

The error in the mean (σ_x) gets smaller as the number of measurements (n) increases.
Don't confuse the error in the mean (σ_x) with the standard deviation of the distribution (σ)!
If we make more measurements:

The standard deviation (σ) of the (gaussian) distribution remains the same,
BUT the error in the mean (σ_x) decreases

The Error on the Mean

Example: Two experiments measure the mass of the proton:

Experiment 1 measures $m_p = 950 \pm 25$ MeV

Experiment 2 measures $m_p = 936 \pm 5$ MeV

Using just the average of the two experiments we find:

$$m_p = (950+936)/2 = 943\text{MeV}$$

Using the weighted average of the two experiments we find:

$$m_p = (950/25^2 + 936/5^2)/(1/25^2 + 1/5^2) = 936.5\text{MeV}$$

and the variance:

$$\sigma^2 = 1/(1/25^2 + 1/5^2) = 24\ \text{MeV}^2$$

$$\sigma = 4.9\ \text{MeV}$$

Since experiment 2 was more precise than experiment 1 we would expect the final result to be closer to experiment 2's value.

$$m_p = (936.5 \pm 4.9)\ \text{MeV}$$