Least-Square Data Fitting

data fitting = art worth serious study (not here)

seen: interpolate within a table

now: least-squares, “best” fit to data (search, matrices)

Rubin H Landau

With

Sally Haerer

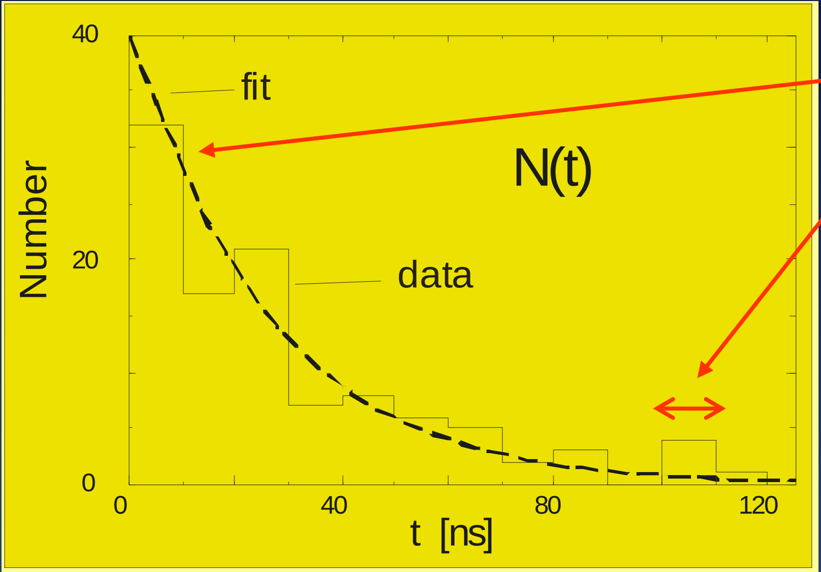
Computational Physics for Undergraduates

BS Degree Program: Oregon State University

“Engaging People in Cyber Infrastructure”

Support by EPICS/NSF & OSU

Problem: Fitting Exponential Decay

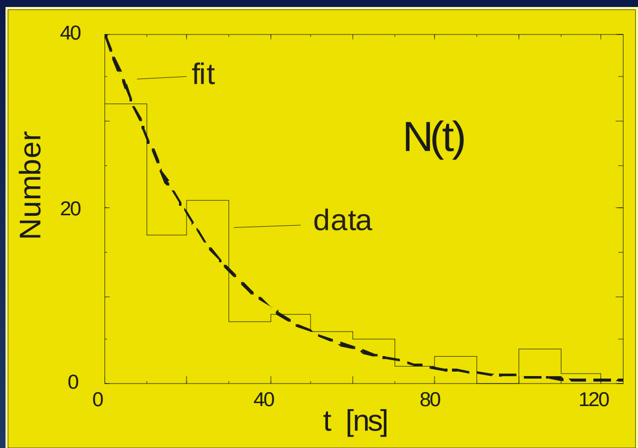


- Number $=\Delta$ N = # of π meson decays
- Time in Δ **t = 10** ns "bins"
- Curve = theoretical fit (your problem)
  - deduce π meson lifetime τ
  - tabulated τ **= 2.6 10^{-8} s**

- Theory: spontaneous decay (stochastic)
- Model: Exponential decay **(Δ t $\rightarrow$ 0)**

$$\frac{dN(t)}{dt} = \frac{dN(0)}{dt} e^{-t/\tau} \qquad (1)$$

Method: Least-Squares Fitting 1



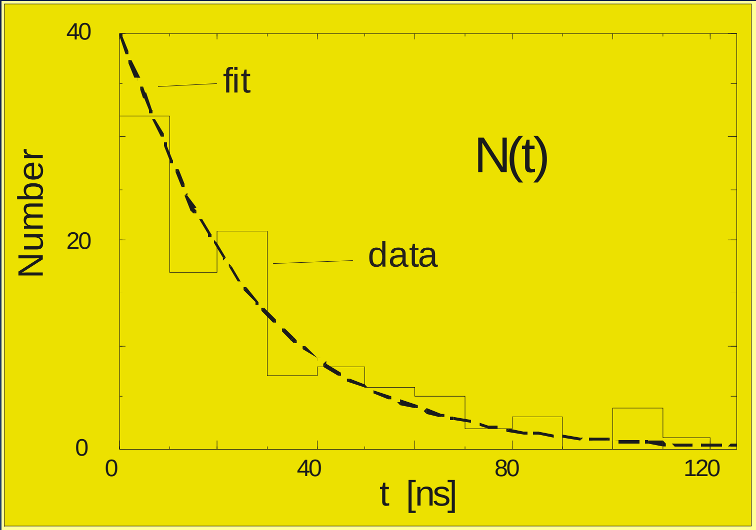
- Meaning "good" fit to data?
- Statistics = big subject (see refs)
- Three points to remember
  1. if errors, theory not pass through all
  2. if theory wrong, "best" fit terrible
  3. most best fits via search

Theory

$$\frac{dN(t)}{dt} = \frac{dN(0)}{dt} e^{-t/\tau} \qquad (1)$$

- ***y = g (x)*** describes data
- ***g (x) = g (x, a_1, a_2, … a_M) = g (x, $\vec{a}$)***
  - ***a_1, a_2, … a_M*** = parameters, part of theory
  - ***a_1*** = lifetime τ, ***a_2***= initial rate ***dN(0)/dt***

Least-Squares Fitting 2



$$\frac{dN(t)}{dt} = \frac{dN(0)}{dt}\, e^{-t/\tau} \qquad (1)$$

1. $\boldsymbol{y = g(x, a_1, a_2, \ldots\ a_M)}$ = Eq.(1) describes data
2. Given: N_D data values
3. $\boldsymbol{(x_i,\ y_i \pm \sigma_i),\ \ i = 1,\ N_D = \#}$ data points
4. $\boldsymbol{y_i}$ = independent variable: $\boldsymbol{t}$
5. $\boldsymbol{x_i}$ = dependent variable: $\Delta\, \boldsymbol{N(t)}$
6. $\pm\sigma_i$ = uncertainty ("error") in $\boldsymbol{y_i}$

χ^2 = Measure of Goodness (the "square")

$$\chi^2 \stackrel{\text{def}}{=} \sum_{i=1}^{N_D} \left[\frac{y_i - g(x_i, \vec{a})}{\sigma_i}\right]^2 \qquad (2)$$

- χ**2** = summed squares deviations of data from ***g (x)***
- $\Rightarrow$ smaller χ**2** $\Rightarrow$ better fit
- **1 /** σ_i^2 = weighting $\Rightarrow$ large errors contribute least
- Least-squares = "best" fit
- "Fit" = adjust a_i $\Rightarrow$ minimum χ**2** (the "least")
- Determine parameters in theory
- χ**2** $\cong$ $N_D - M_P$ **= #** degrees freedom, good
- Good fit: misses ~ 1/3 points
- χ**2 = 0** $\Rightarrow$ theory passes thru all data points

Equations for Best Fit

$$\chi^2 \stackrel{\text{def}}{=} \sum_{i=1}^{N_D} \left[\frac{y_i - g(x_i, \vec{a})}{\sigma_i} \right]^2 \qquad (1)$$

Solve for a_m that make $\chi 2$ a minimum:

$$\frac{\partial \chi^2}{\partial a_m} = 0 \qquad (m = 1, M_P) \qquad (2)$$

$$\Rightarrow \quad \sum_{i=1}^{N_D} \frac{[y_i - g(x_i)]}{\sigma_i^2} \frac{\partial g(x_i)}{\partial a_m} = 0 \qquad (3)$$

- $\Rightarrow$ ***M_P*** simultaneous equations (nonlinear) in ***a_m***
- Work them out for each case
- Solution: trial-and-error search in ***M_P*** dimensions
- Need check: χ**2** = minimum, global minimum
- Different starting values $\Rightarrow \Delta$ min

E.G. Linear Least-Square Fit

$$g\,(x;\vec{a}) = a_1 + a_2 x \tag{4}$$

- Fitted function $g\,(x, \vec{a}\,)$ linear in parameter $\boldsymbol{a_i}$
- Here straight line too!
- Simple (analytic) solution ("Linear regression")
- M_P = 2 parameters: slope $\boldsymbol{a_2}$, $\boldsymbol{y}$ intercept $\boldsymbol{a_1}$
- *N.B.* still $\boldsymbol{N_D} \geq \boldsymbol{M_P = 2}$ of data points to fit
  - Unique solution: More data than unknowns
- Straight line (4) $\Rightarrow$ 2 derivatives for min χ**2**:

$$\frac{\partial g(x_i)}{\partial a_1} = 1, \qquad \frac{\partial g(x_i)}{\partial a_2} = x_i \tag{5}$$

- Solve χ**2** minimization equations, ...

E.G. Linear Least-Square Fit

$$g(x;\vec{a}) = a_1 + a_2 x \qquad (6)$$

Solve χ2 minimization equations (algebra):

$$a_1 = \frac{S_{xx}S_y - S_x S_{xy}}{\Delta}, \qquad a_2 = \frac{S S_{xy} - S_x S_y}{\Delta} \qquad (7)$$

$$S = \sum_{i=1}^{N_D} \frac{1}{\sigma_i^2}, \qquad S_x = \sum_{i=1}^{N_D} \frac{x_i}{\sigma_i^2}, \qquad (8)$$

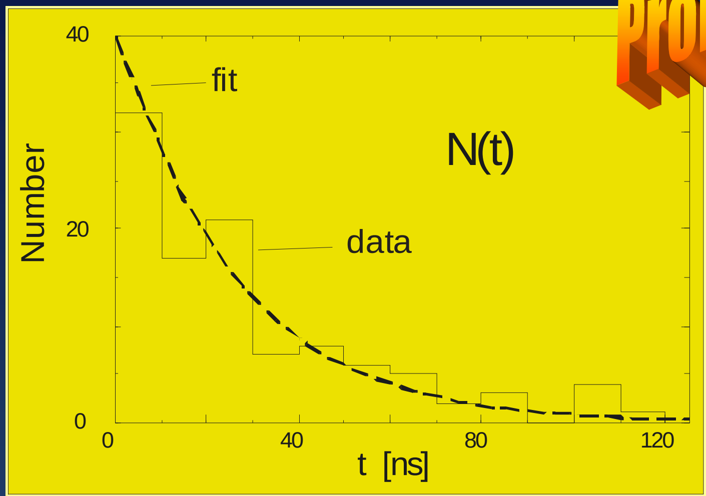
$$S_{xy} = \sum_{i=1}^{N_D} \frac{x_i y_i}{\sigma_i^2}, \qquad \Delta = S S_{xx} - S_x^2 \qquad (9)$$

Variance = uncertainty in parameters

$$\sigma_{a_1}^2 = \frac{S_{xx}}{\Delta}, \qquad \sigma_{a_2}^2 = \frac{S}{\Delta} \qquad (10)$$

Good Time for a Break

Assessment: Fit to Exponential Decay



Problem

1. Fit exponential decay law to data in Fig
2. Find best values for *dN/dt(0)* and τ
3. Judge how good the fit is

Implement

1. Construct table **(Δ N_i, Δ t_i)** from figure
2. Estimate error σ_i by "eye" or $\sqrt{N_i}$
3. Fit the semilog plot

(1) $$\frac{dN(t)}{dt} = \frac{dN(0)}{dt} e^{-t/\tau}$$

$$\ln\left|\frac{\Delta N(t)}{\Delta t}\right| \simeq -\frac{1}{\tau}\Delta t + \ln\left|\frac{\Delta N_0}{\Delta t}\right| \qquad (2)$$

$$y = ax + b \qquad (3)$$

4. Plot best fit with data and compare

Extension: Linear Quadratic Fit

$$g(x) = a_1 + a_2 x + a_3 x^2 \tag{1}$$

- Recall: *linear* fit ⇒ linear in parameters $\boldsymbol{a_i}$ not straight line!
- Nonlinear: $\mathbf{g(x) = (a_1 + a_2 x)\, e^{a\, x}}$
- Parabola: min χ**2** ⇒ **3** simultaneous equations

$$\sum_{i=1}^{N_D} \frac{[y_i - g(x_i)]}{\sigma_i^2} \frac{\partial g(x_i)}{\partial a_1} = 0, \qquad \frac{\partial g}{\partial a_1} = 1, \tag{2}$$

$$\sum_{i=1}^{N_D} \frac{[y_i - g(x_i)]}{\sigma_i^2} \frac{\partial g(x_i)}{\partial a_2} = 0, \qquad \frac{\partial g}{\partial a_2} = x, \tag{3}$$

$$\sum_{i=1}^{N_D} \frac{[y_i - g(x_i)]}{\sigma_i^2} \frac{\partial g(x_i)}{\partial a_3} = 0, \qquad \frac{\partial g}{\partial a_3} = x^2. \tag{4}$$

Solution: Linear Quadratic Fit

$$g(x) = a_1 + a_2 x + a_3 x^2 \tag{1}$$

3 eqs, 3 unknowns

$$\begin{aligned} S a_1 + S_x a_2 + S_{xx} a_3 &= S_y, && (2)\\ S_x a_1 + S_{xx} a_2 + S_{xxx} a_3 &= S_{xy}, && (3)\\ S_{xx} a_1 + S_{xxx} a_2 + S_{xxxx} a_3 &= S_{xxy} && (4) \end{aligned}$$

$$e.g.\quad S_{xy} = \sum_{i=1}^{N_D} \frac{x_i y_i}{\sigma_i^2} \tag{5}$$

JamaFit.java

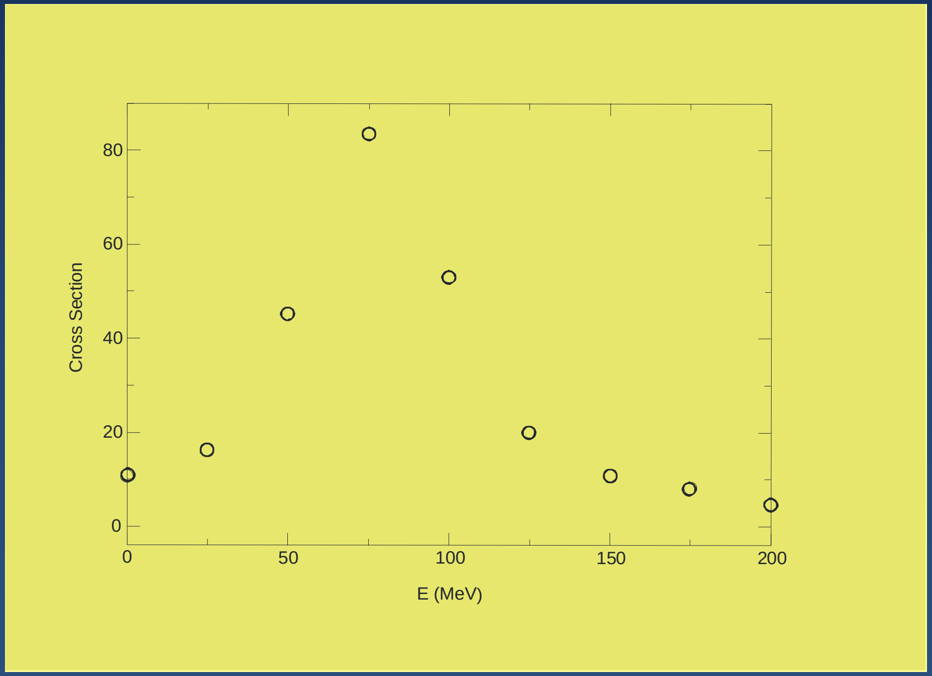
$$[\alpha]\vec{a} = \vec{\beta} \tag{6}$$

$$[\alpha] = \begin{bmatrix} S & S_x & S_{xx} \\ S_x & S_{xx} & S_{xxx} \\ S_{xx} & S_{xxx} & S_{xxxx} \end{bmatrix} \tag{7}$$

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \qquad \vec{\beta} = \begin{bmatrix} S_y \\ S_{xy} \\ S_{xxy} \end{bmatrix} \tag{8}$$

Nonlinear Fit to Cross Section (challenge)

i	1	2	3	4	5	6	7	8	9
E_i	0	25	50	75	100	125	150	175	200
$f(E_i)$	10.6	16.0	45.0	83.5	52.8	19.9	10.8	8.25	4.70
$+-\sigma_i$	9.34	17.9	41.5	85.5	51.5	21.5	10.8	6.29	4.14



$$f(E) = \frac{f_r}{(E - E_r)^2 + \Gamma^2/4}$$

- 3 parameters (OK)
- Non linear in *E* (hard)
- Newton-Raphson Search