mean μ

std. dev. σ

$$\mathrm{Prob}(x \mid \mu, \sigma) \propto e^{-(x-\mu)^2/2\sigma^2}$$

$$\int_{-\infty}^{\infty} \mathrm{Prob} = 1$$

$$\Rightarrow \mathrm{PDF}(x \mid \mu, \sigma) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$$

$$\text{Joint PDF}(x_i \mid \mu_i, \sigma_i) = \prod_i \frac{e^{-(x_i-\mu_i)^2/2\sigma_i^2}}{\sqrt{2\pi\sigma_i^2}} \quad \text{iff } x_i \text{ are indep.}$$

$i = 1 \ldots N$

$\Pi \Rightarrow \Sigma$

$$\log(\mathrm{PDF}) = \sum -(x_i-\mu_i)^2/2\sigma_i^2 - \frac{1}{2}\sum \ln(2\pi\sigma_i^2)$$

$$-2\log(\mathrm{PDF}) = \sum \frac{(x_i-\mu_i)^2}{\sigma_i^2} + \sum \ln(2\pi\sigma_i^2)$$

$$\chi^2 \equiv \sum \frac{(x_i - \mu_i)^2}{\sigma_i^2} = -2 \ln(PDF) + \text{norm}\ldots$$

model 1 give μ_i

model 2 gives μ_i'

relative like of model 1 vs. model 2.

$$\frac{PDF_1}{PDF_2} \Rightarrow \log\left(\frac{PDF_1}{PDF_2}\right) = \log(PDF_1) - \log(PDF_2)$$
$$= -\frac{1}{2}(\chi_1^2 - \chi_2^2)$$

$\chi_2^2 = \chi_1^2 + 10$ $\quad e^{-\frac{1}{2}(\chi_1^2 - \chi_2^2)} = e^{-5} \Rightarrow 0.006 \Rightarrow$ model 2 is disfavore

$\delta\chi^2 = 0.5$ $\quad e^{-\frac{1}{2}(0.5)}$ $\quad e^{-1/4}$

$= .75$ $\quad .78$ $\quad$ model 2 is ok

$$\chi^2 = \sum \frac{(x_i - \mu_i)^2}{(\sigma_i)^2} \qquad \mu_i = A_i(m)$$

$$\Rightarrow \sum \frac{(x_i - A_i(m))^2}{\sigma_i^2}$$

$$(x_1 - A_1(m),\ x_2 - A_2(m))^T \begin{pmatrix} \frac{x_1 - A_1}{\sigma_1^2} \\ \frac{x_2 - A_2}{\sigma_2^2} \\ \vdots \end{pmatrix}$$

$$(\vec{x} - \vec{A}(m))^T N^{-1} (x - A(m)) \qquad N_{ii} = \sigma_i^2$$

$$N = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \sigma_3^2 \\ & & & \ddots \end{pmatrix}$$

$$= \sum \frac{(x_i - A_i(m))^2}{\sigma_i^2}$$

$$N^{-1}(x - A(m))$$

$$= \begin{pmatrix} \frac{x_1 - A_1(m)}{\sigma_1^2} \\ \frac{x_2 - A_2(m)}{\sigma_2^2} \\ \vdots \end{pmatrix}$$

$$N^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & \\ 0 & \frac{1}{\sigma_2^2} & \\ & & \ddots \end{pmatrix}$$

identical to what we started with

in Linear algebra

where d are data

$$\chi^2 = (d - A(m))^T N^{-1} (d - A(m))$$

$$\langle d \rangle = A(m) \Rightarrow \langle d \rangle = Am$$

$$\chi^2 = (d - Am)^T N^{-1} (d - Am)$$ if linear

max likelihood:

$$\frac{\partial \chi^2}{\partial m} = 0 \qquad \frac{\partial \chi}{\partial m} = -A^T N^{-1} (d - Am) + (d - Am)^T N^{-1} (-A)$$

$$= -2 A^T N^{-1} (d - Am) = 0$$

$$\Rightarrow A^T N^{-1} d - A^T N^{-1} A m = 0$$

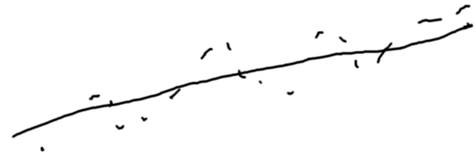
$$\boxed{A^T N^{-1} A m = A^T N^{-1} d}$$

$$m = (A^T N^{-1} A)^{-1} (A^T N^{-1} d)$$

$\langle d_i \rangle = \sum_j c_j x_i^j \Rightarrow$ polynomial

$d_i = \sum_j c_j x_i^j + n_i$

$\langle d_0 \rangle = c_0 + x_0 c_1 + x_0^2 c_2 + \dots$



Am

$$m = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \end{pmatrix}$$

$$A = \begin{pmatrix} x_0^0 & x_0^1 & x_0^2 & x_0^3 & \dots \\ x_1^0 & x_1^1 & x_1^2 & x_1^3 & \dots \\ \vdots \end{pmatrix} \overset{m}{\begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ \vdots \end{pmatrix}} = c_0 + c_1 x_0 + c_2 x_0^2 + \dots \Rightarrow \langle d_0 \rangle$$

rows of A = # of data points

columns of A = # model parameters

$\dim(A) = n_{data} \times n_{par}$ $\quad \dim(N) = n_{data} \times n_{data}$

$\dim(m) = n_{par}$

$\dim(d) = n_{data}$

$(A^T N^{-1} A m = \underbrace{A^T N^{-1} d}_{n_{par} \times 1})$ $\quad A^{-1}$ doesn't exist unless $n_{par} = n_{data}$

$\underbrace{(A^T N^{-1} A)}_{n_{dat} \times n_{par}}$

$m = (A^T N^{-1} A)^{-1} (A^T N^{-1} d)$

$n_{par} \times n_{data} \cdot n_{data} \cdot n_{par}$

$\Rightarrow n_{par} \times n_{par}$

$$\underbrace{A^T N^{-1} A}_{\text{symm.}} \Rightarrow \underbrace{V}_{\text{eigenvals/vectors}} \Lambda V^T$$

$$(A^T N^{-1} A)^{-1} = (V \Lambda V^T)^{-1} \Rightarrow V^{T-1} \Lambda^{-1} V^{-1}$$

$$(ABC)^{-1} = C^{-1} B^{-1} A^{-1}$$

But $V^T = V^{-1}$

$V^{T-1} = V$

$$= (A^T N^{-1} A)^{-1} \Rightarrow V \Lambda^{-1} V^T$$

$\Lambda_{small} \ll \Lambda_{large}$ roundoff error will be bad!!