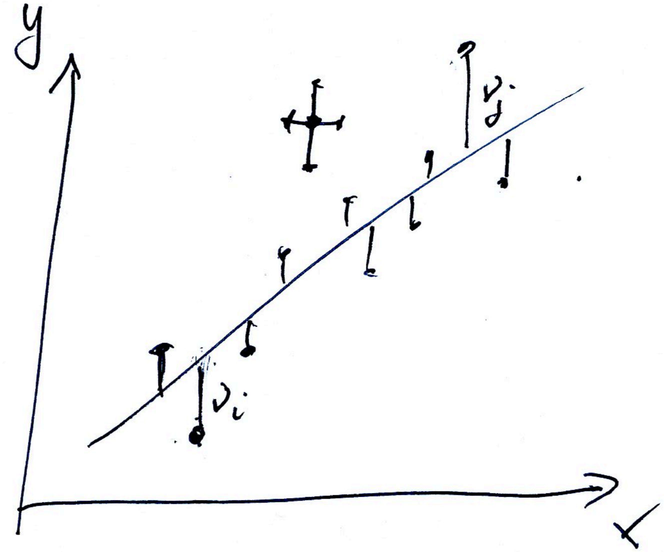
Regression

Linear regression (x_i, y_i)



$$y_i = a x_i + b$$

Assumption: $\boxed{\begin{array}{l} x_i \text{ are precise} \\ y_i \text{ have errors} \end{array}}$

independent

$$y_i = a x_i + b + \nu_i$$

$$\langle \nu_i \rangle = 0$$

$$\langle \nu_i^2 \rangle = \sigma^2$$

$$\langle \nu_i \nu_j \rangle = \sigma^2 \delta_{ij}$$

Later ν_i normal

$$P(\nu_i) = \frac{e^{-\frac{\nu_i^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

$$y_i - a x_i - b = \nu_i$$

$$\langle \nu_i \nu_j \rangle - \langle \nu_i \rangle \langle \nu_j \rangle = 0$$

$$\langle \nu_i \rangle = 0$$

$$\min \left[\sum_i \nu_i^2 \right] = \boxed{\chi^2 = \sum_i \nu_i^2 = \sum_i \left[y_i - a x_i - b \right]^2}$$

$$\chi^2 = \sum_i \left[y_i^2 + a^2 x_i^2 + b^2 - 2a\cdot x_i y_i - 2b y_i + 2ab x_i \right] =$$

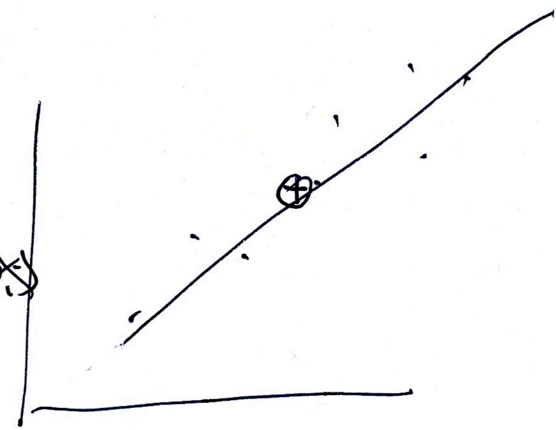
$$\chi^2 = \sum_i y_i^2 + a^2 \sum x_i^2 + b^2 \sum_i 1 - 2a \sum_i x_i y_i - 2b \sum y_i + 2ab \sum x_i$$

$$\frac{1}{N}\chi^2 = \langle y^2 \rangle + a^2 \langle x^2 \rangle + b^2 - 2a \langle xy \rangle - 2b \langle y \rangle + 2ab \langle x \rangle$$

$$\frac{1}{N}\frac{\partial \chi^2}{\partial a} = 0 = \cancel{2}a \langle x^2 \rangle - \cancel{2} \langle xy \rangle + \cancel{2}b \langle x \rangle = 0$$

$$\frac{1}{N}\frac{\partial \chi^2}{\partial b} = 0 = \cancel{2}b - \cancel{2} \langle y \rangle + \cancel{2}a \langle x \rangle = 0 \qquad a\langle x \rangle + b = \langle y \rangle$$

$$a \langle x^2 \rangle + b \langle x \rangle = \langle xy \rangle$$



$$a x_i + b = y_i - \nu_i \Rightarrow a\langle x \rangle + b = \langle y \rangle \quad \cancel{\langle \nu_i \rangle}$$

$$a\langle x_i^2 \rangle + b x_i = x_i y_i - \underbrace{\nu_i x_i}_{\langle \nu_i x_i \rangle = \cancel{\langle \nu_i \rangle} \langle x_i \rangle}$$

$$a\langle x\rangle + \beta = \langle y\rangle \longrightarrow \beta = \langle y\rangle - a\langle x\rangle$$

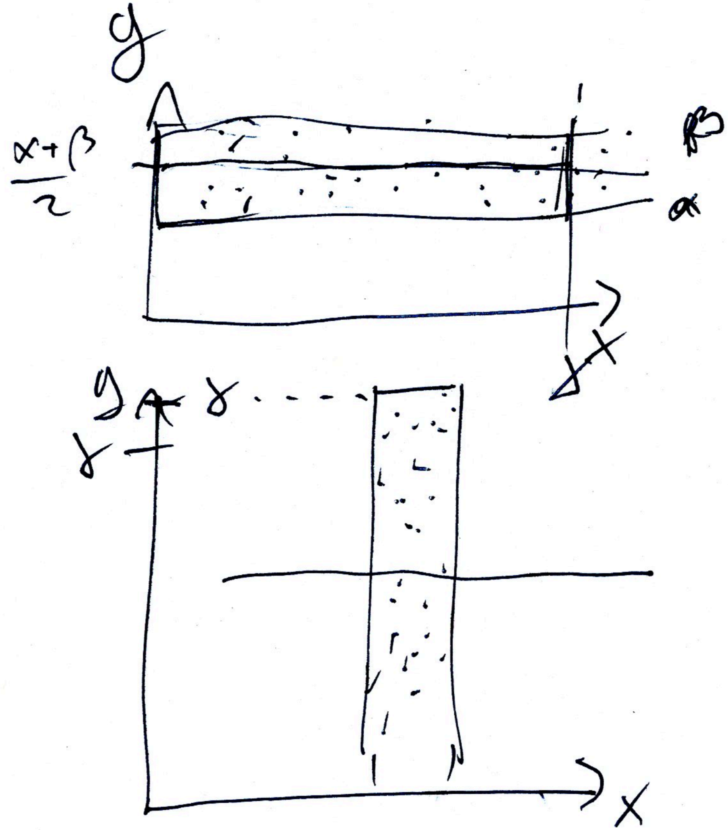
$$a\langle x^2\rangle + \beta\langle x\rangle = \langle xy\rangle$$

$$\langle x, y\rangle = \langle x\rangle\langle y\rangle$$

$$\underline{a\langle x^2\rangle} + \left[\langle y\rangle - \underline{a\langle x\rangle}\right]\langle x\rangle = \langle xy\rangle$$

$$a\underbrace{\left[\langle x^2\rangle - \langle x\rangle^2\right]}_{Var(x)} = \underbrace{\langle xy\rangle - \langle x\rangle\langle y\rangle}_{Cov(x,y)}$$

$$\boxed{a = \frac{Cov(x,y)}{Var(x)}}$$



χ^2 What is the meaning of χ^2?

$$\frac{1}{\sigma^2}\chi^2 = \sum\left(\frac{\nu_i^2}{\sigma_i^2}\right) = \qquad \mathrm{Var}\left(\frac{\nu_i}{\sigma}\right) = 1$$

If ν_i Gaussian: $$P(\nu_i) = \frac{e^{-\nu_i^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \qquad P\left(\frac{\nu_i}{\sigma}\right) = \frac{e^{-\frac{t_i^2}{2}}}{\sqrt{2\pi}}$$

$$= P(\nu_1)P(\nu_2)P(\nu_3)\dots P(\nu_N) = P_N(\nu_1,\dots \nu_N) \qquad t_i = \frac{\nu_i}{\sigma_i}$$

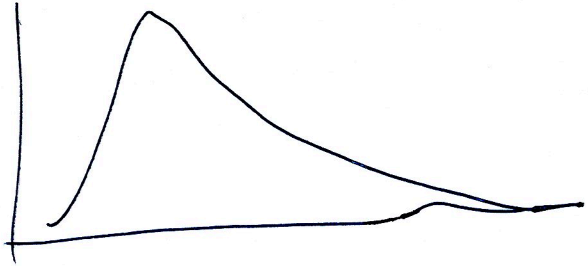
$$P_N = \frac{e^{-\frac{t_1^2}{2}}\,e^{-\frac{t_2^2}{2}}\dots e^{-\frac{t_N^2}{2}}}{(2\pi)^{N/2}} = \frac{e^{-\frac{1}{2}\sum t_i^2}}{(2\pi)^{N/2}}$$ likelihood fn

$\chi^2 = -$log likelihood

$$\boxed{-\ln P = +\underbrace{\frac{N}{2}\ln(2\pi)} + \underbrace{\frac{1}{2}\sum_i t_i^2}} = \chi^2 + c$$

$$\langle\chi^2\rangle = \left\langle \sum_{i=1}^{N}\frac{[y_i - \alpha x_i - \beta]^2}{\sigma^2}\right\rangle = \sum_{i=1}^{N}\underbrace{\langle t_i^2\rangle}_{=1} = N$$

Chi-square distribution



$$\langle \chi_N^2 \rangle = N$$

Fit line to $(x_i, y_i) \Rightarrow a, b \Rightarrow \frac{1}{\sigma^2}\sum (y_i - a x_i - b)^2 \Rightarrow N$

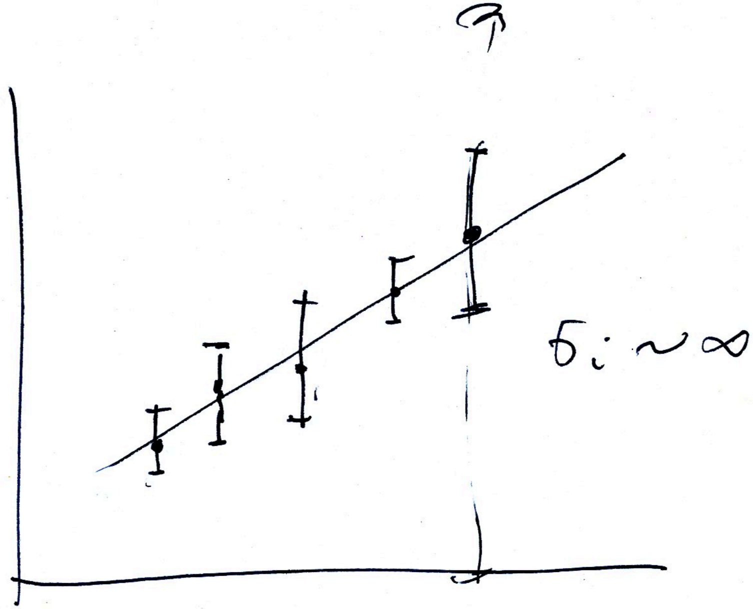
Consistency check on our model.

$\boxed{\sigma^2 = \frac{1}{N}\chi^2}$ estimate of the error.

x_i, y_i, σ_i

Normal

$$\cancel{\chi^2_L} \quad y_i - a x_i - b - \sigma_i \nu_i = 0$$



$$\chi^2 = \sum \nu_i^2$$

$$\nu_i = \frac{(y_i - a x_i - b)}{\sigma_i}$$

$$\frac{1}{\sigma^2} \to 0$$

$$\chi^2 = \sum \underbrace{\frac{1}{\sigma_i^2}}_{w_i} (y_i - a x_i - b)^2$$

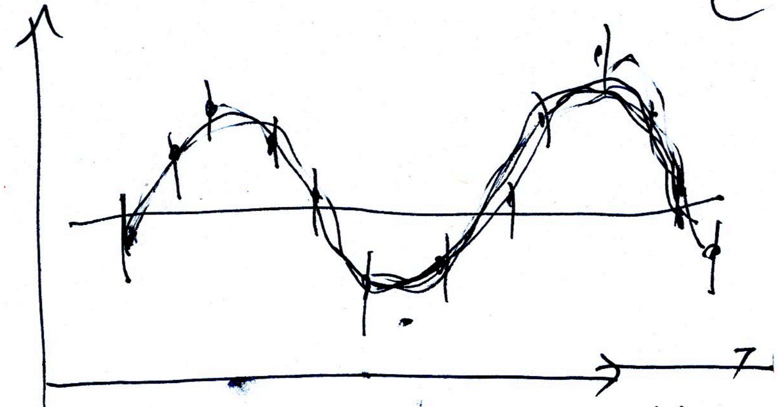
residual

$$= \sum w_i (y_i - a x_i - b)^2$$

Weighted Least Squares.

$\frac{1}{\sigma_i^2}$ information content

$$\boxed{\text{inverse variance weighting}}$$



$$y = A\cdot\sin(2\pi t + \varphi) + c.$$

$$= A\cdot\sin(2\pi t)\cos\varphi + A\cos(2\pi t)\sin\varphi$$

$$= \underbrace{[A\cos\varphi]}_{a}\cdot\sin(2\pi t) + \underbrace{[A\sin\varphi]}_{b}\cos(2\pi t)$$

$$y_i = a\cdot\sin(2\pi x_i) + b\cos(2\pi x_i) + c. + \nu_i.$$

$$\nu_i = [y_i - a\sin(2\pi x) - b\cos(2\pi x) - c]$$

$$\chi^2 = \sum \nu_i^2 = \sum[y_i - a\overset{f_1(x)}{\underline{\cos(2\pi x)}} - b\overset{f_2(x)}{\sin(2\pi x)} - c]^2 = f(a,b,c)$$

$$\cancel{a^2 + b^2 + c^2} \ldots$$

$$\frac{\partial\chi^2}{\partial a} = \sum 2[y_i - a f_1(x_i) - b f_2(x_i) - c](-f_1(x_i)) = 0$$

$$\frac{\partial\chi^2}{\partial b} = \sum \ldots (-f_2()) = 0$$

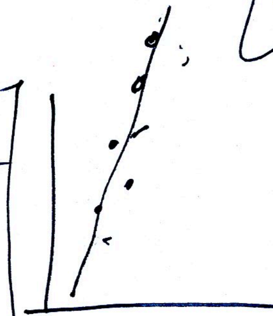
$$\frac{\partial\chi^2}{\partial c} \quad [\quad\quad](-1) = 0$$

Linear model

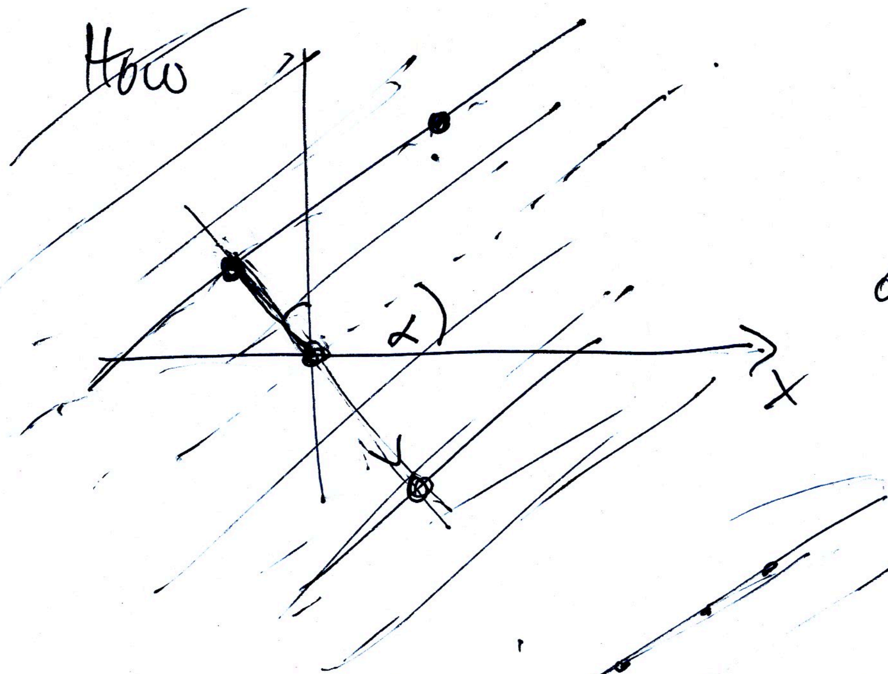
$$y = \sum_k a_k f_k(x)$$

$$y = ax + b$$

$$x = a'y + b'$$



How



$0 \le \alpha \le \pi$

1 point, angle

↓

signed distance, direction

Parametric equation of a line.