ADA05 - 9am Thu 22 Sep 2022

Optimal Scaling

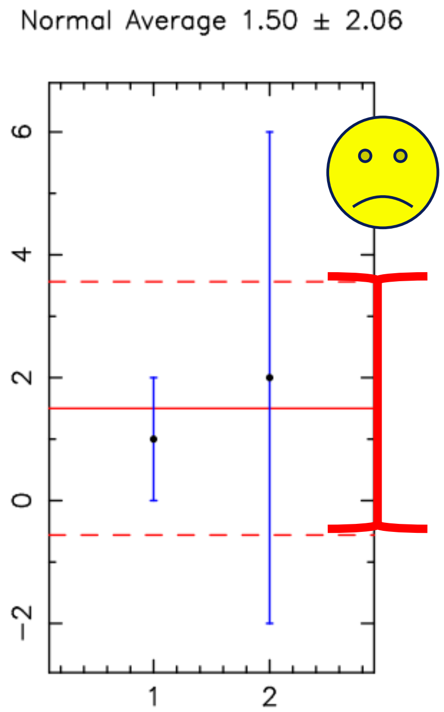
Fitting models by minimizing χ^2
Parameter uncertainty from $\Delta\chi^2=1$

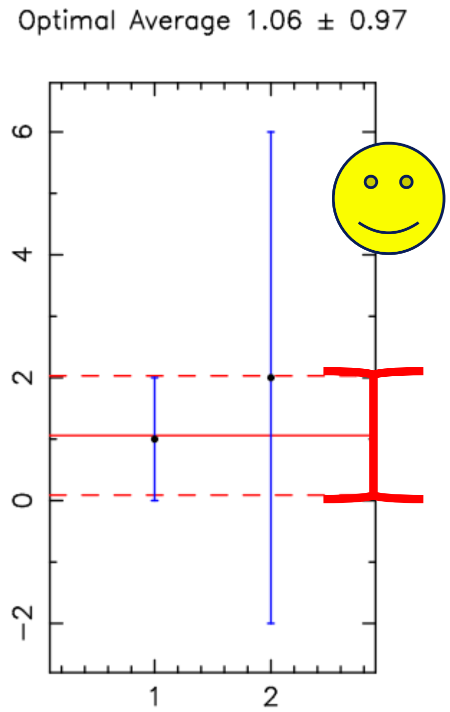
Dancing χ^2 Landscape
χ^2_{min} and $\Delta\chi^2$
Degrees of Freedom

Review: Sample Mean vs Optimal Average

$$\bar{X} \equiv \frac{1}{N} \sum_i X_i$$

$$\sigma^2\left(\bar{X}\right) = \frac{1}{N^2} \sum_i \sigma_i^{\,2}$$





$$\hat{X} \equiv \frac{\sum_i X_i / \sigma_i^2}{\sum_i 1/\sigma_i^2}$$

$$\sigma^2\left(\hat{X}\right) = \frac{1}{\sum_i 1/\sigma_i^2}$$

Equal weights:
Poor data degrades the result.

Better to ignore “bad” data.

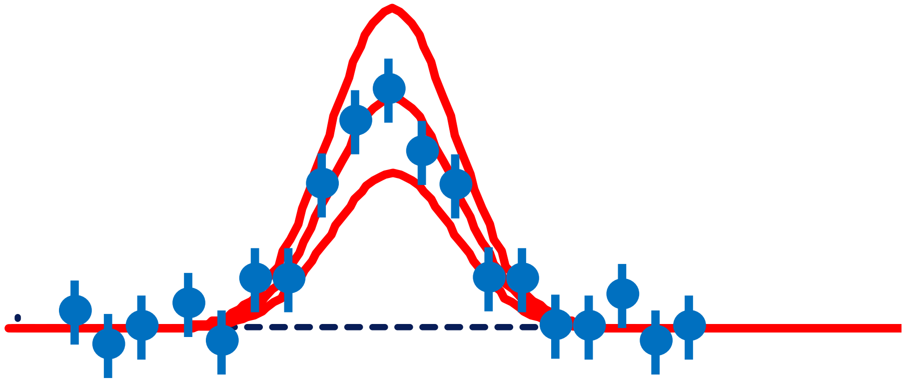
Information lost.

Optimal weights:
New data always improves the result.

Use ALL the data, but with appropriate **1 / Variance** weights.

Must have good error bars.

Measuring a Feature



A = area under the curve,

e.g. flux of a star, strength of a spectral line.

Assume (for now) zero background, known pattern.

Model: $\mu_i \equiv \langle X_i \rangle = A\, P_i \qquad \mathrm{Cov}\left[X_i, X_j \right] = \sigma_i^2\, \delta_{i\,j}$

How to measure A ?

Simple method: **Integrate the Data**:

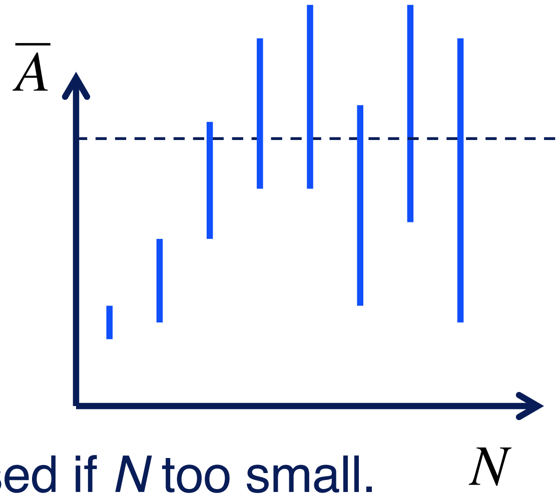
$$\bar{A} \equiv \sum_{i=1}^{N} X_i$$

$$\langle \bar{A} \rangle = A \sum_{i=1}^{N} P_i \qquad \sigma^2\left[\bar{A} \right] = \sum_{i=1}^{N} \sigma_i^2$$

If P_i = fraction of photons in pixel i ,

$$\sum_{i=1}^{N} P_i = 1$$

Dilemma:
How many data points to include ?



Biased if N too small.

Noisy if N too large.

Can we do better? Yes, if the pattern *P* is known.

Optimal Scaling of a Pattern

Scale the pattern P_i by factor A to fit the data.

1: Construct independent unbiased estimates.

2: Optimal average, with $1/\sigma^2$ weights.

Data: $X_i \pm \sigma_i$

Model: $\mu_i \equiv \langle X_i \rangle = A\, P_i$

Pattern: P_i

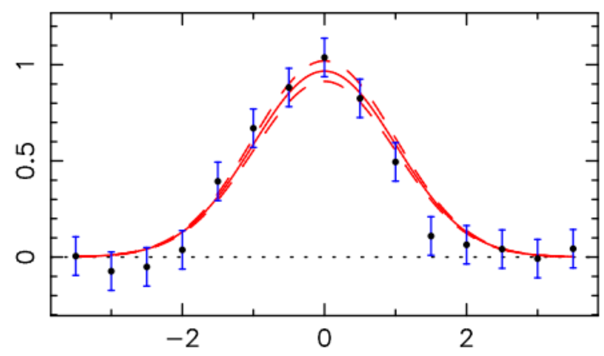
$$A_i \equiv X_i / P_i \quad \text{unbiased:} \quad \langle A_i \rangle = A \qquad \text{Cov}\left[A_i, A_j\right] = \left(\frac{\sigma_i}{P_i}\right)^2 \delta_{ij}$$

$$\text{Optimal average:} \;\; w_i = 1/\text{Var}\left[A_i\right] = \left(P_i / \sigma_i\right)^2$$

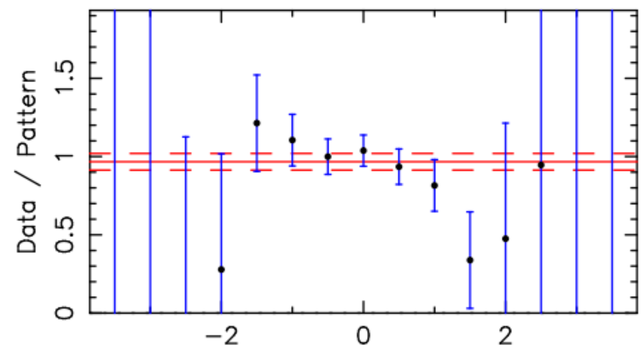
$$\hat{A} = \frac{\sum_i w_i\, A_i}{\sum_i w_i} = \frac{\sum_i \left(\frac{P_i}{\sigma_i}\right)^2 \left(\frac{X_i}{P_i}\right)}{\sum_i \left(\frac{P_i}{\sigma_i}\right)^2} = \frac{\sum_i X_i\, P_i / \sigma_i^2}{\sum_i P_i^2 / \sigma_i^2}$$

$$\text{Var}\left[\hat{A}\right] = \frac{\sum_i \text{Var}\left[X_i\right] \left(P_i / \sigma_i^2 \right)^2}{\left(\sum_i P_i^2 / \sigma_i^2\right)^2} = \frac{1}{\sum_i P_i^2 / \sigma_i^2}$$

Optimal Scaling 0.97 ± 0.05



Optimal Average 0.97 ± 0.05

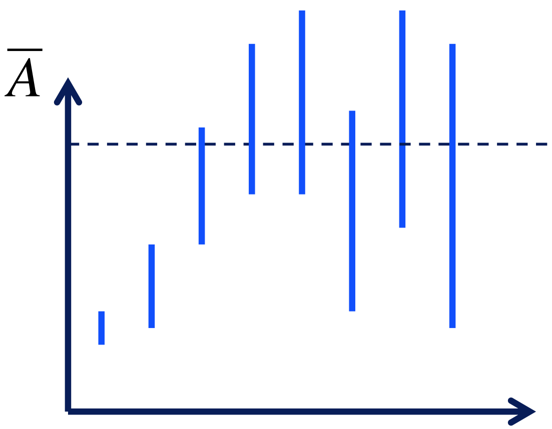


Sum the Data vs Optimal Scaling

Sum up the data.

$$\overline{A} \equiv \sum_{i=1}^{N} X_i$$

$$\left\langle \overline{A} \right\rangle = A \sum_{i=1}^{N} P_i \qquad \sigma^2\left[\overline{A}\right] = \sum_{i=1}^{N} \sigma_i^{\,2}$$



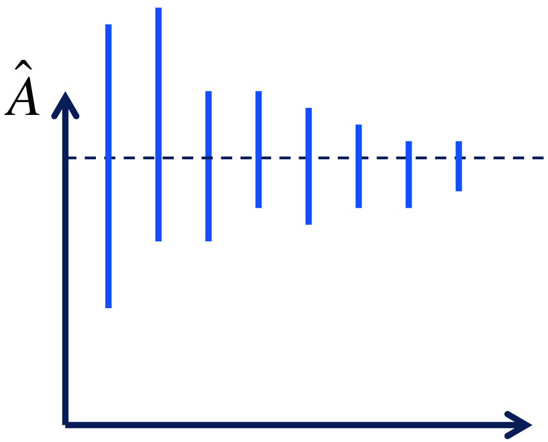
Biased if N too small. N

Noisy if N too large.

Optimal Scaling of known Pattern.

$$\hat{A} = \frac{\sum_i X_i \, P_i \big/ \sigma_i^{\,2}}{\sum_i P_i^2 \big/ \sigma_i^{\,2}}$$

$$\mathrm{Var}\left[\hat{A}\right] = \frac{1}{\sum_i P_i^2 \big/ \sigma_i^{\,2}}$$



No bias. N

Result improves with N.

Optimal Scaling

"Golden Rule" of Optimal Data Analysis:

$$\text{Data}: \quad X_i \pm \sigma_i$$

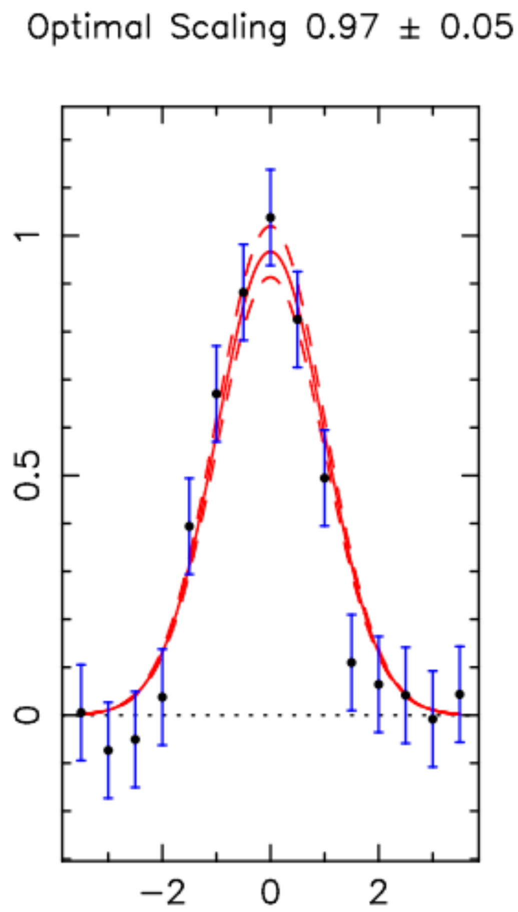
$$\text{Model}: \quad \langle X_i \rangle \equiv \mu_i = A\, P_i$$

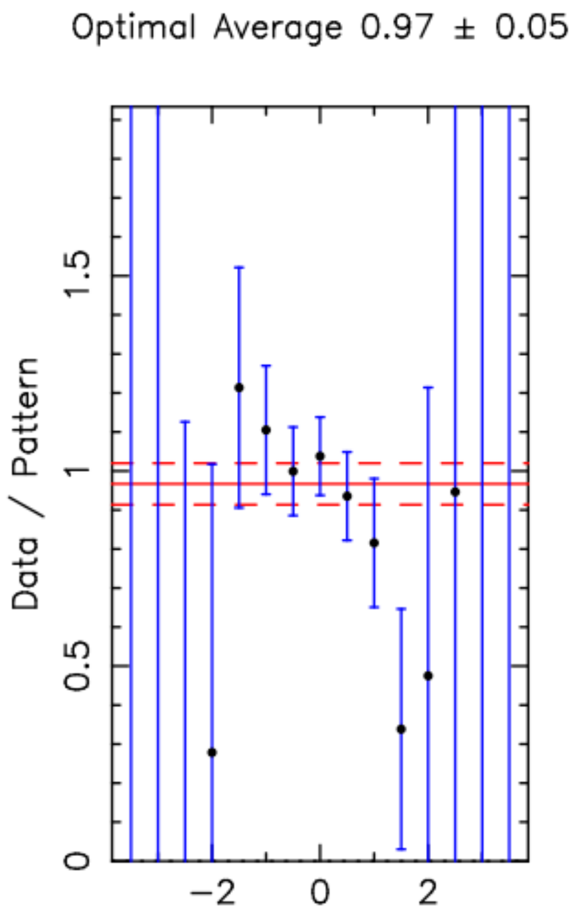
Optimal Scaling:

$$\hat{A} = \frac{\sum_i X_i\, P_i / \sigma_i^2}{\sum_i P_i^2 / \sigma_i^2}$$

$$\text{Var}\left[\hat{A}\right] = \frac{1}{\sum_i P_i^2 / \sigma_i^2}$$

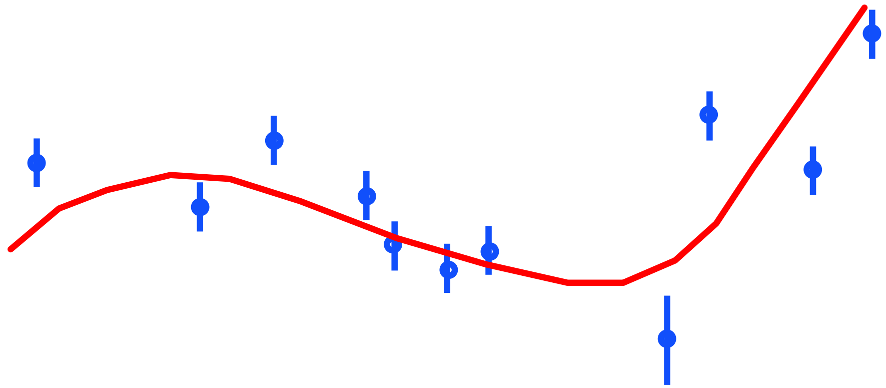
Memorise this result.
Know how to derive it.





Optimal Average is a special case of Optimal Scaling, with pattern $P_i = 1$.

Fitting Models by minimising χ^2



Data: $X_i \pm \sigma_i \quad i = 1 \ldots N$

Model: $\langle X_i \rangle \equiv \mu_i(\alpha)$

Parameters: $\alpha_k \quad k = 1 \ldots M$

Error: $\varepsilon_i \equiv X_i - \mu_i(\alpha)$

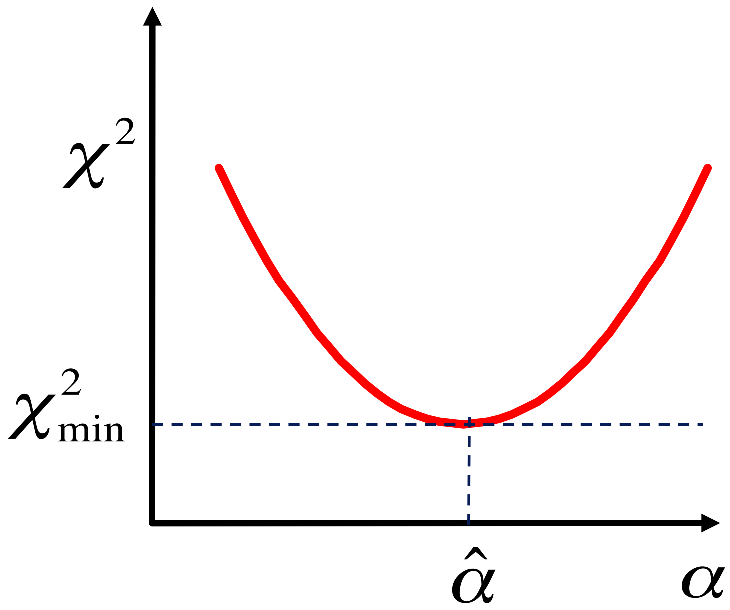
Normalised Error: $\chi_i \equiv \dfrac{\varepsilon_i}{\sigma_i} = \dfrac{X_i - \mu_i(\alpha)}{\sigma_i}$

"**Badness-of-Fit**" statistic:

$$\chi^2(X, \sigma, \alpha) \equiv \sum_{i=1}^{N} \chi_i^2 = \sum_{i=1}^{N} \left(\frac{X_i - \mu_i(\alpha)}{\sigma_i} \right)^2$$

Best-fit parameters $\hat{\alpha}$ minimise χ^2.

(BoF a.k.a. "Goodness-of-Fit" statistic).

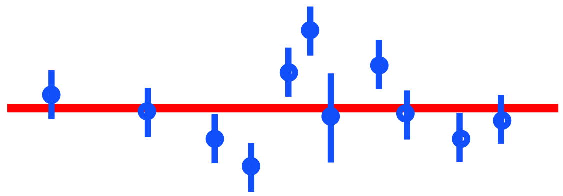


Example: Estimate $< X >$ by χ^2 Fitting

Model: $\langle X_i \rangle = \mu \quad \mathrm{Cov}\left[X_i, X_j\right] = \sigma_i^2\, \delta_{ij}$

Badness-of-Fit statistic:

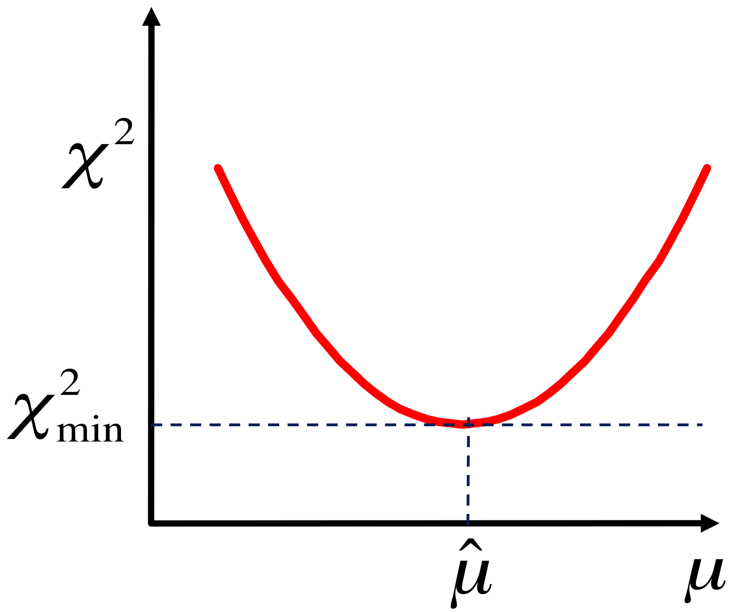
$$\chi^2 = \sum_i \left(\frac{X_i - \mu}{\sigma_i} \right)^2$$



Minimise χ^2 :

$$\frac{\partial \chi^2}{\partial \mu} = -2 \sum_i \frac{X_i - \mu}{\sigma_i^2} = 0 \quad \text{at} \quad \mu = \hat{\mu}$$

$$\sum_i \frac{X_i}{\sigma_i^2} = \sum_i \frac{\hat{\mu}}{\sigma_i^2} \quad \Rightarrow \quad \hat{\mu} = \frac{\sum_i X_i / \sigma_i^2}{\sum_i 1/\sigma_i^2} = \hat{X}.$$



The Optimal Average minimises χ^2 !

Parameter Error Bar: $1-\sigma$ at $\Delta\chi^2 = 1$

From χ^2 fit : $\hat{\mu} = \hat{X} =$ Optimal Average

Must have $\sigma^2(\hat{\mu}) = \sigma^2(\hat{X}) = \dfrac{1}{\sum_i 1/\sigma_i^2}$

$$\frac{\partial \chi^2}{\partial \mu} = -2 \sum_i \frac{X_i - \mu}{\sigma_i^2}$$

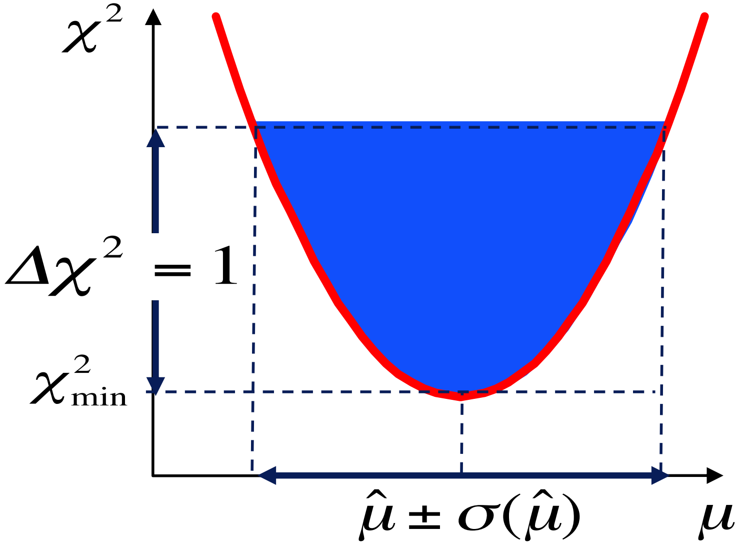
$$\frac{\partial^2 \chi^2}{\partial \mu^2} = +2 \sum_i \frac{1}{\sigma_i^2}$$

$$\chi^2 = \chi^2_{\min} + \frac{1}{2} \left. \frac{\partial^2 \chi^2}{\partial \mu^2} \right|_{\mu = \hat{\mu}} (\mu - \hat{\mu})^2 + \ldots$$

$$= \chi^2_{\min} + \left(\sum_i \frac{1}{\sigma_i^2} \right) (\mu - \hat{\mu})^2 + \ldots$$

$$= \chi^2_{\min} + \left(\frac{\mu - \hat{\mu}}{\sigma(\hat{\mu})} \right)^2 + \ldots$$

“Fill with water” to depth of 1



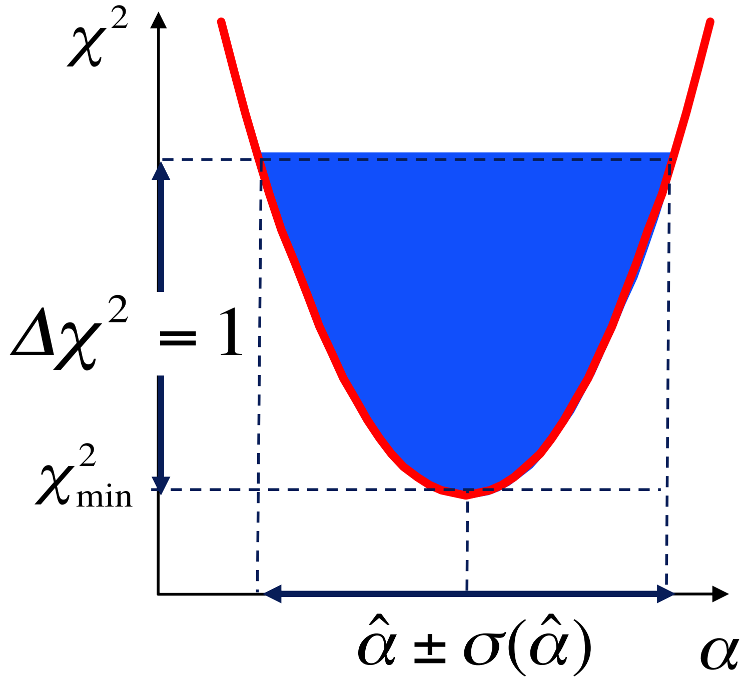
$$\therefore \quad \Delta\chi^2 \equiv \chi^2 - \chi^2_{\min} \approx \left(\frac{\mu - \hat{\mu}}{\sigma(\hat{\mu})} \right)^2 = 1 \quad \text{for} \quad \mu = \hat{\mu} \pm \sigma(\hat{\mu})$$

Parameter Error Bar: 1-σ from χ^2 Curvature

$$\Delta\chi^2 \equiv \chi^2 - \chi^2_{\min} \approx \frac{1}{2}\left(\frac{\partial^2 \chi^2}{\partial \alpha^2}\right)\Bigg|_{\alpha=\hat{\alpha}} (\alpha - \hat{\alpha})^2$$

$$= \left(\frac{\alpha - \hat{\alpha}}{\sigma(\hat{\alpha})}\right)^2 = 1 \quad \text{for} \quad \alpha = \hat{\alpha} \pm \sigma(\hat{\alpha})$$

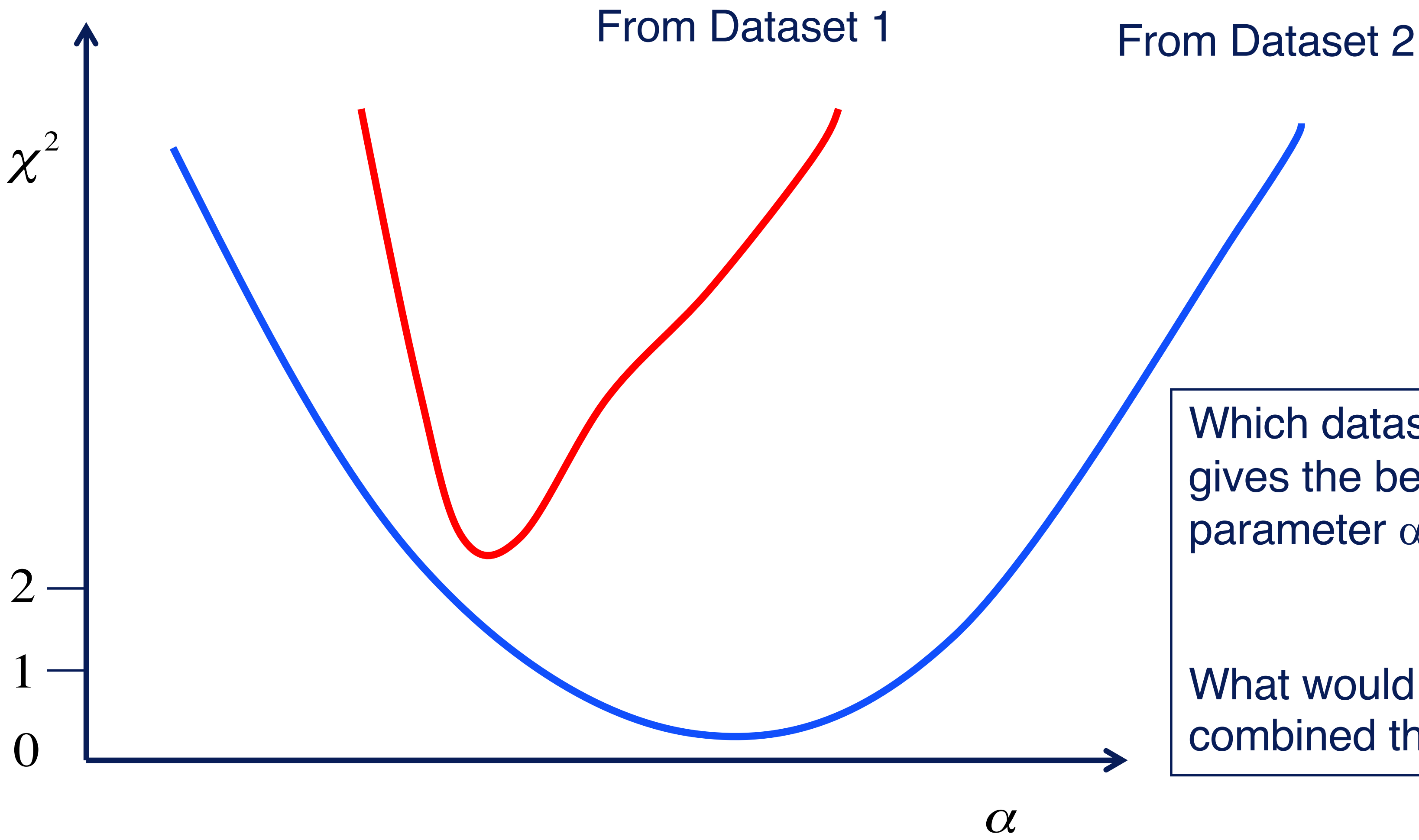
$$\therefore \quad \sigma^2(\hat{\alpha}) = \frac{2}{\left(\frac{\partial^2 \chi^2}{\partial \alpha^2}\right)\Big|_{\alpha=\hat{\alpha}}}$$



Exact for linear models, BoF(α) quadratic in α.

Approximate for non-linear models, BoF(α) not quadratic in α.

Test Understanding



Which dataset (1 or 2) gives the better estimate of parameter α ?

What would $\chi^2(\alpha)$ be if you combined these 2 datasets ?

Scaling a Pattern by χ^2 minimization

Model: $\mu_i \equiv \langle X_i \rangle = A\,P_i$

Badness-of-fit:

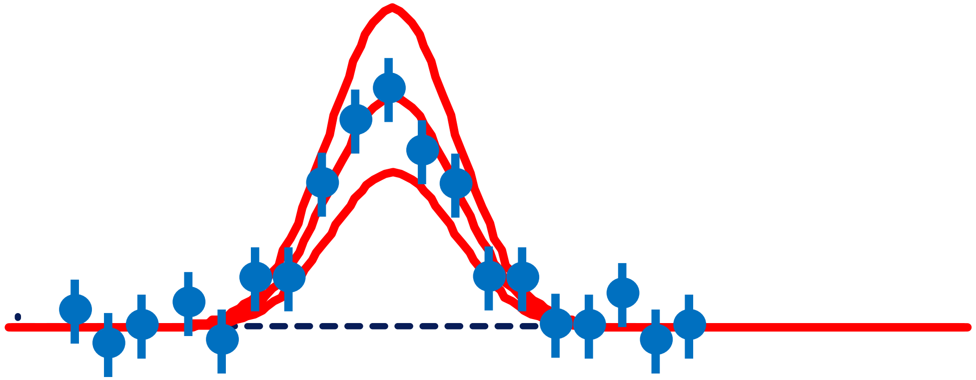
$$\chi^2 = \sum_i \left(\frac{X_i - A\,P_i}{\sigma_i} \right)^2$$

Minimise χ^2:

$$0 = \frac{\partial \chi^2}{\partial A} = -2 \sum_i \frac{(X_i - A\,P_i)\,P_i}{\sigma_i^2}$$

$$\Rightarrow \quad \sum_i \frac{X_i\,P_i}{\sigma_i^2} = \sum_i \frac{\hat{A}\,P_i^2}{\sigma_i^2}$$

$$\hat{A} = \frac{\sum_i X_i\,P_i / \sigma_i^2}{\sum_i P_i^2 / \sigma_i^2}.$$



$$\frac{\partial^2 \chi^2}{\partial A^2} = +2 \sum_i \frac{P_i^2}{\sigma_i^2}$$

$$\sigma^2\left(\hat{A}\right) = \frac{2}{\dfrac{\partial^2 \chi^2}{\partial A^2}} = \frac{1}{\sum_i P_i^2 / \sigma_i^2}$$

Same result as Optimal Scaling.

Summary: Optimal Average/Scaling is equivalent to Minimise χ^2

- Two 1-parameter models:
  - Estimating $< X >$: $\mu_i \equiv \langle X_i \rangle = \mu$
  - Scaling a pattern: $\mu_i \equiv \langle X_i \rangle = A\, P_i$
- Two equivalent methods:
  - **Algebra of Random Variables: Optimal Average and Optimal Scaling**

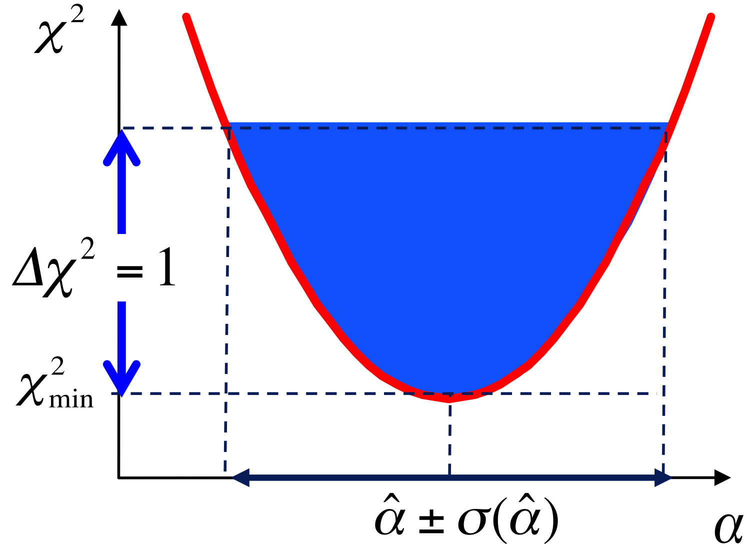
$$\hat{X} = \frac{\sum_i X_i / \sigma_i^2}{\sum_i 1/\sigma_i^2} \qquad \sigma^2\left(\hat{X}\right) = \frac{1}{\sum_i 1/\sigma_i^2}$$

$$\hat{A} = \frac{\sum_i X_i\, P_i / \sigma_i^2}{\sum_i P_i^2 / \sigma_i^2} \qquad \sigma^2\left(\hat{A}\right) = \frac{1}{\sum_i P_i^2 / \sigma_i^2}$$

  - **Minimising χ^2** gives same result:

$$\Delta\chi^2 \equiv \chi^2 - \chi^2_{\min} = \left(\frac{\alpha - \hat{\alpha}}{\sigma\left(\hat{\alpha}\right)}\right)^2 + \dots$$

$$\sigma^2\left(\hat{\alpha}\right) = \frac{2}{\left.\frac{\partial^2 \chi^2}{\partial \alpha^2}\right|_{\alpha=\hat{\alpha}}}$$



χ^2_{min} = "Badness of Fit" statistic

χ^2_{min} is a statistic.
It has a probability distribution:

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{X_i - \mu_i(\alpha)}{\sigma_i} \right)^2 \sim \chi^2_{N-M}$$

X_i = data values $i = 1 \ldots N$

$\sigma_i = 1$ - σ error bar

$\mu_i(\alpha)$ = model predicted data value

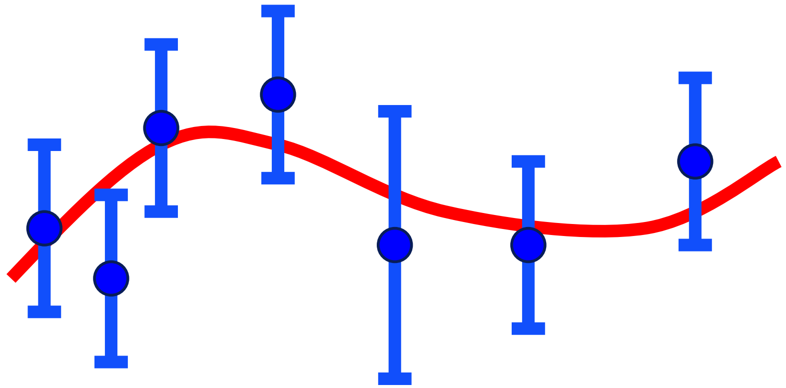
α_k = parameters of the model $k = 1 \ldots M$

N = number of data points

M = number of fitted parameters

$N - M$ = degrees of freedom

To fit *N* data points, adjust *M* parameters to minimise χ^2.



Review : Constructing χ^2_N from N Gaussians

- Sum of squares of N independent Gaussian random variables

$\chi^2_N \equiv$ Chi-squared with N degrees of freedom

X and Y are independent Gaussian random variables.

$X \sim G(0,1) \qquad Y \sim G(0,1)$

$X^2 \sim \chi^2_1 \qquad Y^2 \sim \chi^2_1$

$X^2 + Y^2 \sim \chi^2_2$

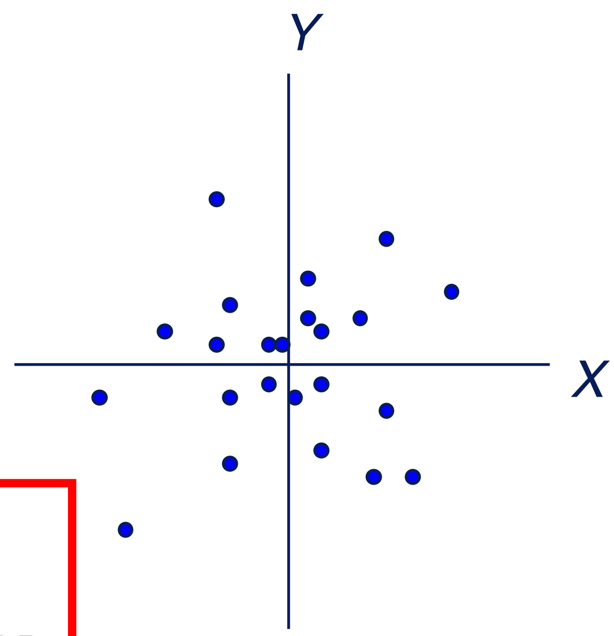
and so on for each

new degree of freedom:

$$\chi^2_N + \chi^2_M \sim \chi^2_{N+M}$$

$$\left\langle \chi^2_N \right\rangle = N$$

$$\sigma^2\left(\chi^2_N\right) = 2N$$

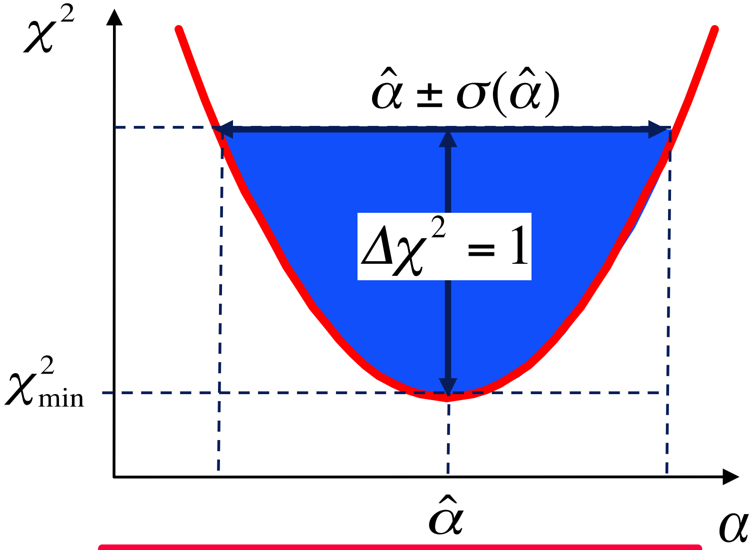


Dancing Data => Dancing χ^2 Landscape

Fit M parameters to N data points.

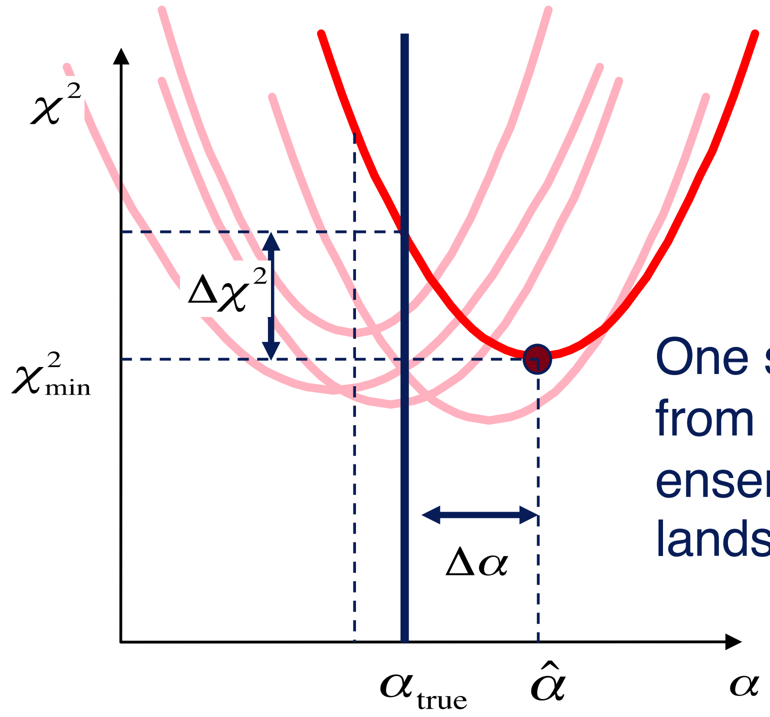
$$\chi^2(X,\sigma,\alpha) \equiv \sum_{i=1}^{N}\left(\frac{X_i-\mu_i(\alpha)}{\sigma_i}\right)^2$$

Best - fit parameters $\hat{\alpha}$ minimise χ^2.



$$\sigma^2(\hat{\alpha}) = \frac{2}{\left.\frac{\partial^2\chi^2}{\partial\alpha^2}\right|_{\alpha=\hat{\alpha}}}$$

Caveat: Assumes orthogonal parameters.
Generalise to correlated parameters later.



One sample from an ensemble of χ^2 landscapes.

$$\hat{\alpha} \sim G\left(\alpha_{\text{true}}, \sigma^2(\hat{\alpha})\right)$$

$$\chi^2(\alpha_{true}) \sim \chi^2_N$$

$$\chi^2_{\min} \equiv \chi^2(\hat{\alpha}) \sim \chi^2_{N-M}$$

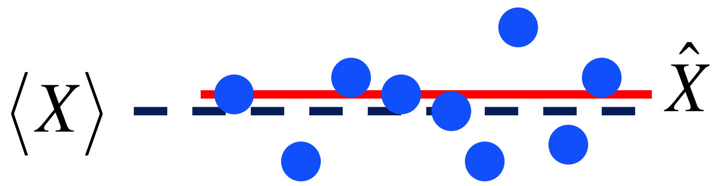
$$\Delta\chi^2 \equiv \chi^2(\alpha_{\text{true}}) - \chi^2_{\min} \sim \chi^2_M$$

Degrees of Freedom (DoF)

N data : $\langle X_i \rangle = \langle X \rangle \quad \mathrm{Cov}(X_i, X_j) = \sigma_i^2 \, \delta_{ij}$

$$\sum_{i=1}^{N} \left(\frac{X_i - \langle X \rangle}{\sigma_i} \right)^2 \sim \chi_N^2. \quad N \text{ degrees of freedom.}$$

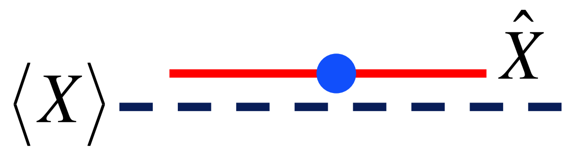
If $\langle X \rangle$ unknown, use $\hat{X}$ instead:



$$\sum_{i=1}^{N} \left(\frac{X_i - \hat{X}}{\sigma_i} \right)^2 \sim \chi_{N-1}^2. \quad N-1 \text{ degrees of freedom.}$$

For a single datum : $N = 1, \quad \hat{X} = X_1$

$$\left(\frac{X_1 - \langle X \rangle}{\sigma_1} \right)^2 \sim \chi_1^2. \quad 1 \text{ degree of freedom}$$



$$\left(\frac{X_1 - \hat{X}}{\sigma_1} \right)^2 = 0. \quad 0 \text{ degrees of freedom.}$$

Fit M parameters to N data :

$$\sum_{i=1}^{N} \left(\frac{X_i - \mu_i(\alpha)}{\sigma_i} \right)^2 \sim \chi_{N-M}^2. \quad N-M \text{ degrees of freedom.}$$

Each fitted parameter removes 1 degree of freedom from the residuals.

Fini -- ADA 05