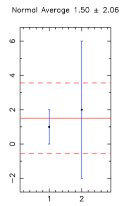
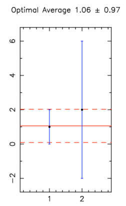
Review: Normal vs Optimal Average

$$\bar{X} = \frac{1}{N}\sum_i X_i \qquad \sigma^2(\bar{X}) = \frac{1}{N^2}\sum_i \sigma_i^2$$





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

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 the star, strength of a spectral line.
Assume (for now) zero background.

How to measure A ?

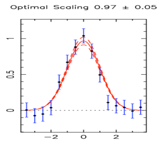
Simple method: **Integrate the Data**:

$$\langle X_i \rangle = A\,P_i \qquad \mathrm{Cov}[X_i, X_j] = \sigma_i^2\,\delta_{ij}$$

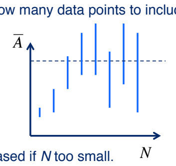
$$\bar{A} = \sum_{i=1}^{N} X_i$$

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

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.

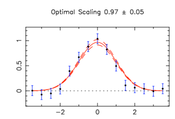
Data : $X_i \pm \sigma_i$
Model : $\langle X_i \rangle = \mu_i = A\,P_i$
Pattern : P_i

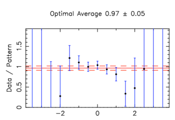
$A_i = X_i/P_i$ unbiased: $\langle A_i \rangle = A$ $\quad \mathrm{Cov}[A_i, A_j] = \left(\frac{\sigma_i}{P_i}\right)^2 \delta_{ij}$

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

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

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



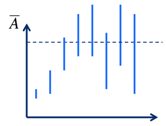


Sum the Data vs Optimal Scaling

Sum up the data.

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

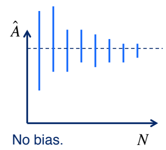
$$\langle \bar{A} \rangle = A\sum_{i=1}^{N} P_i \qquad \sigma^2[\bar{A}] = \sum_{i=1}^{N}\sigma_i^2$$



Biased if N too small.
Noisy if N too large.

Optimal Scaling of known Pattern.

$$\hat{A} = \frac{\sum_i X_i P_i/\sigma_i^2}{\sum_i P_i^2/\sigma_i^2} \qquad \mathrm{Var}[\hat{A}] = \frac{1}{\sum_i P_i^2/\sigma_i^2}$$



No bias.
Result improves with N.

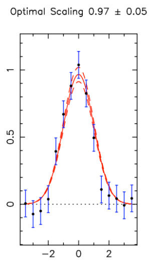
Optimal Scaling

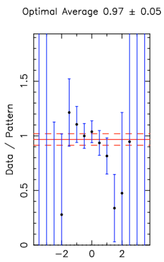
The Golden Rule of Optimal Data Analysis:

Data : $X_i \pm \sigma_i$
Model : $\langle X_i \rangle = \mu_i = A\,P_i$
Optimal Scaling :

$$\hat{A} = \frac{\sum_i X_i P_i/\sigma_i^2}{\sum_i P_i^2/\sigma_i^2} \qquad \mathrm{Var}[\hat{A}] = \frac{1}{\sum_i P_i^2/\sigma_i^2}$$

Memorise this result.
Know how to derive it.





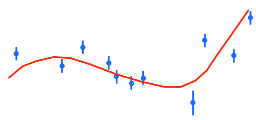
Optimal Average is a special case of Optimal Scaling, with pattern $P_i = 1$.

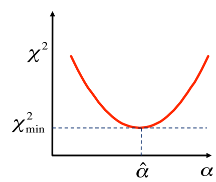
Fitting Models by minimising χ^2

Data: $X_i \pm \sigma_i \quad i = 1 \ldots N$
Model: $\langle X_i \rangle = \mu_i(\alpha)$
Parameters: $\alpha_k \quad k = 1 \ldots M$
Error: $\varepsilon_i = X_i - \mu_i(\alpha)$
Normalised Error: $\eta_i = \frac{\varepsilon_i}{\sigma_i} = \frac{X_i - \mu_i(\alpha)}{\sigma_i}$
"Badness-of-Fit" statistic:

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

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





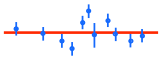
(BoF a.k.a. "Goodness-of-Fit" statistic).

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

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

Badness-of-Fit statistic:

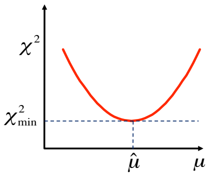
$$\chi^2 = \sum_i \left(\frac{X_i - \mu}{\sigma_i} \right)^2$$



Minimise χ^2 :

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

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



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

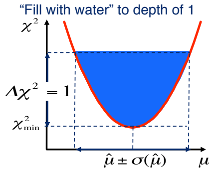
$$\frac{\partial \chi^2}{\partial \mu} = -2 \sum_i \frac{X_i - \mu}{\sigma_i^2} \qquad \frac{\partial^2 \chi^2}{\partial \mu^2} = +2 \sum_i \frac{1}{\sigma_i^2}$$

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

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

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

"Fill with water" to depth of 1



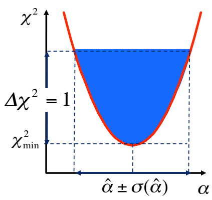
$$\therefore \quad \Delta\chi^2 = \chi^2 - \chi^2_{min} \approx \left(\frac{\mu - \hat{\mu}}{\sigma(\hat{\mu})} \right)^2 = 1 \quad \text{for} \quad \mu = \hat{\mu} \pm \sigma(\hat{\mu})$$

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

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

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

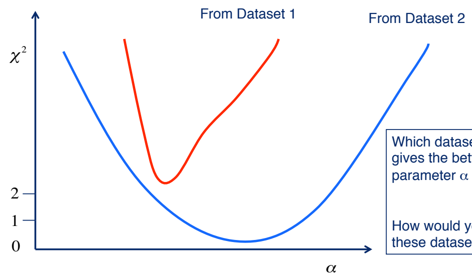
$$\therefore \quad \sigma^2(\hat{\alpha}) = \frac{2}{\left(\frac{\partial^2 \chi^2}{\partial \alpha^2} \right) \Big|_{\alpha = \hat{\alpha}}}$$



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 ?

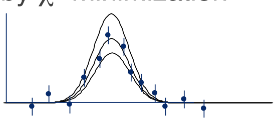
How would you combine these datasets ?

Scaling a Pattern by χ^2 minimization

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

Badness - of - fit :

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



Minimise :

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

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

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

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

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

Same result as Optimal Scaling.

☺

Summary

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

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

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

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

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

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