Using $\chi^2_{\min}$ to reject models

Fit M parameters to N data points:

$$\chi^2_{\min} = \sum_{i=1}^{N} \left[\frac{X_i - \mu_i(\alpha_1, ..., \alpha_M)}{\sigma_i} \right]^2 \sim \chi^2_{N-M}$$

$$\left\langle \chi^2_{N-M} \right\rangle = N - M \qquad \sigma^2(\chi^2_{N-M}) = 2(N-M)$$

Why $N - M$ degrees of freedom?
Fitting $M = N$ parameters should fit N points exactly.

If model is good, then the best-fit $\chi^2_{\min}$ should be:

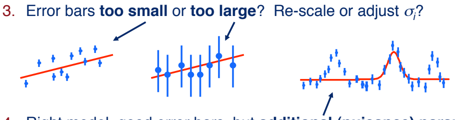
$$\chi^2_{\min} \approx N - M \pm \sqrt{2(N-M)}$$

$$\frac{\chi^2_{\min}}{N-M} \approx 1 \pm \sqrt{\frac{2}{N-M}}$$

What if $\chi^2_{\min}$ is too high (or low)?

Several possibilities:

1. Statistical fluke? Use χ^2_{N-M} distribution to estimate probability
2. Wrong model? Use χ^2_{N-M} distribution to reject model
3. Error bars **too small** or **too large**? Re-scale or adjust σ_i?

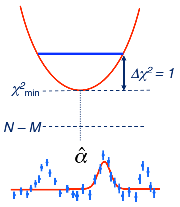


4. Right model, good error bars, but **additional (nuisance) parameters** omitted or not optimised?

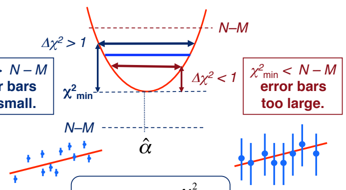
Failure to optimise nuisance parameters increases $\chi^2_{\min}$, but may leave the χ^2 curvature the same, **if the nuisance parameters are orthogonal to the parameters of interest.**

Can then still use $\chi^2_{\min}+1$ to set 1-σ confidence intervals on parameters **orthogonal to the nuisance parameters**.

Diagnosis of $\chi^2_{\min}$ too large or small



$\chi^2_{\min} > N - M$ due to failure to optimise **orthogonal** nuisance parameters ?
If so, then curvature unchanged.
Use $\Delta\chi^2 = 1$.



$\chi^2_{\min} > N - M$ **error bars too small.**

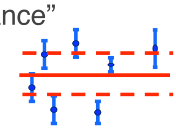
$\chi^2_{\min} < N - M$ **error bars too large.**

Use: $\Delta\chi^2 = \frac{\chi^2_{\min}}{N-M}$

Equivalent to **re-scaling the error bars**

$$\sigma \Rightarrow \sigma \sqrt{\frac{\chi^2_{\min}}{N-M}}$$

Estimate the "Extra Variance"



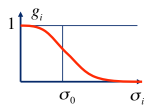
Assume two independent noise sources:

$$\mathrm{Var}[X_i] = \sigma_0^2 + \sigma_i^2 = \frac{\sigma_0^2}{g_i} \qquad g_i \equiv \frac{\sigma_0^2}{\sigma_0^2 + \sigma_i^2} = \frac{1}{1 + (\sigma_i/\sigma_0)^2}$$

$$-2\ln L = \sum_{i=1}^{N} \frac{(X_i-\mu)^2}{\sigma_0^2+\sigma_i^2} + \sum_{i=1}^{N} \ln(\sigma_0^2+\sigma_i^2) = \sum \left(\frac{X_i-\mu}{\sigma_0}\right)^2 g_i + \sum \ln(\sigma_0^2/g_i)$$

$$0 = \frac{\partial(-2\ln L)}{\partial \mu} = -2\sum \frac{(X_i-\mu)}{\sigma_0^2+\sigma_i^2} = -2\sum \frac{(X_i-\mu) g_i}{\sigma_0^2}$$

$$0 = \frac{\partial(-2\ln L)}{\partial \sigma_0^2} = -\sum \frac{(X_i-\mu)^2 g_i^2}{\sigma_0^4} + \sum \frac{g_i}{\sigma_0^2}$$



$$\hat{\mu} = \frac{\sum \frac{X_i}{\sigma_0^2+\sigma_i^2}}{\sum \frac{1}{\sigma_0^2+\sigma_i^2}} = \frac{\sum X_i g_i}{\sum g_i} \qquad \mathrm{Var}[\hat{\mu}] = \frac{\sigma_0^2}{\sum g_i} \qquad \hat{\sigma}_0^2 = \frac{\sum (X_i-\mu)^2 g_i^2}{\sum g_i}$$

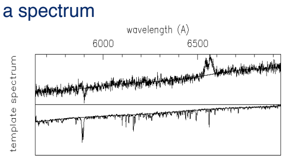
Need to iterate.

Background Functions

Smooth functions with adjustible flexibility.

- Polynomials
- Splines
- Running Optimal Average
  - with sigma-clipping
- Running Median

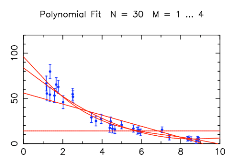
Example: Continuum fit to a spectrum



Polynomials

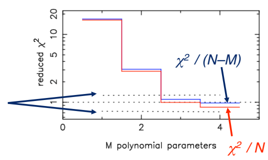
Fit $N = 30$ points with $M = 1, 2, 3, 4$ polynomial coefficients.

Higher M = more flexible model.
Use lowest M that gives good fit.



Reject $M = 1, 2.$
Accept $M = 3, 4.$
Based on Reduced χ^2

$$\frac{\chi^2}{N-M} \approx 1 \pm \sqrt{\frac{2}{N-M}}$$



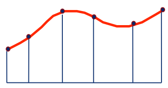
Splines – e.g. piecewise cubic

***N* nodes**: x_i, y_i $i = 1, \dots N$.
x_i fixed, y_i adjustable.

***4 (N – 1)* parameters** (*4* cubic coefficients for each of the *N - 1* segments)

***3 (N – 2)* matching conditions** (value, slope, curvature at each of the *N - 2* internal nodes)

***N + 2* degrees of freedom** (*N* values y_i values plus either slope or curvature at 2 end points).

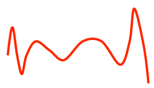


8-parameter cubic spline

- First, **distribute the nodes x_i**, e.g. equally spaced, or equal weight $\Sigma(1/\sigma^2)$ on each segment.
- Then, **fit the *N + 2* parameters**, e.g. optimise y_i by χ^2 minimization, set endpoint curvatures (or slopes) to zero.

Low-order polys good for simple background fits.

Splines better than high-order polys. Better control over the *x* distribution of the degrees of freedom.



8-parameter polynomial

Running Optimal Average

$$\hat{X}(t) = \frac{\sum X_i\, w_i(t)}{\sum w_i(t)} \qquad \sigma^2\left(\hat{X}(t)\right) = \frac{1}{\sum w_i(t)}$$

$$w_i(t) = \frac{G(t - t_i)}{\sigma_i^2}$$

Memory function *G(t)* expands the error bars as time-difference increases.

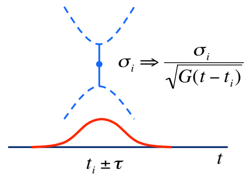
Parameter τ controls time interval over which the data point retains its $1/\sigma^2$ weight.

Memory functions:

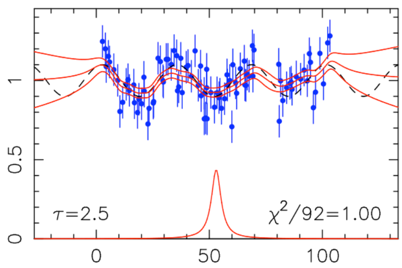
$$\text{boxcar}: \quad G(t) = \begin{cases} 1 & |t| < \tau \\ 0 & |t| > \tau \end{cases}$$

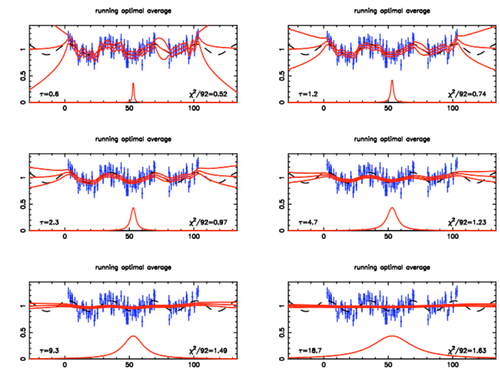
$$\text{Gaussian}: \quad = \exp\left\{ -\frac{1}{2}\left(\frac{t}{\tau}\right)^2 \right\}$$

$$\text{Lorentzian}: \quad = \frac{1}{1 + (t/\tau)^2}$$

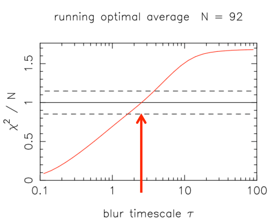


Running Optimal Average

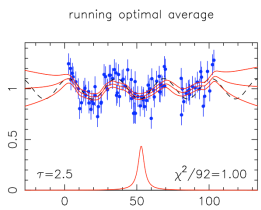




Running Optimal Average



Blur timescale τ chosen to make $\chi^2 / N = 1$.

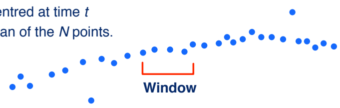


Interpolates across gaps.
Extrapolates past ends.
Averages appropriately.
Error bars provided.
(Almost) model-free.

Median Filter and Sigma-Clip

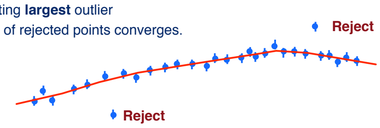
Median filter:

- window of *N* points centred at time *t*
- medfilt(*t*) is the median of the *N* points.



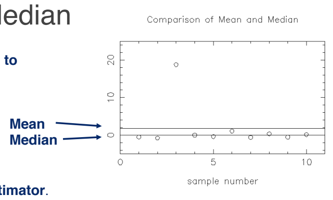
Sigma-clip:

- Fit all points by minimising χ^2
- Set threshold *K* and check for outliers at $\pm K\,\sigma$ or more
- Repeat fit omitting **largest** outlier
- Iterate until set of rejected points converges.

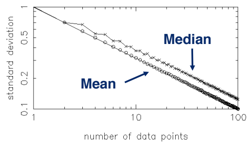


Mean vs Median

- The median is **less sensitive to outliers** than the mean.



- The median is **unbiased** but **not a minimum-variance estimator**.
- Note how **standard deviation of the median** varies with sample size *N* in comparison to **standard deviation of the mean.**



Variance of the Median is larger by a factor *π / 2 = 1.57* (for large *N*) than the Variance of the Mean.

Variance of Median vs Mean

N gaussian random numbers:

$$\langle X_i \rangle = 0 \quad \mathrm{Var}[X_i] = \sigma^2 \quad i = 1 \ldots N$$

$$f(x) = dF/dx = \exp\{-x^2/2\} \big/ (2\pi\sigma^2)^{1/2}$$

P = fraction of positive values:

$$p_i = \begin{cases} 1 & X_i > 0 \\ 0 & X_i < 0 \end{cases} \quad \langle p_i \rangle = \frac{1}{2} \quad \sigma^2(p_i) = \frac{1}{4}$$

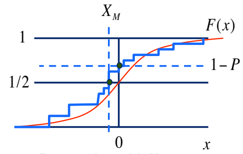
$$P \equiv \frac{1}{N}\sum_{i=1}^{N} p_i \qquad \langle P \rangle = \frac{1}{2} \quad \sigma^2(P) = \frac{1}{4N}$$

$$\text{Median: } X_M \approx \frac{P - \langle P \rangle}{dF/dx\big|_{x=0}} = \left(P - \tfrac{1}{2}\right)\left(2\pi\sigma^2\right)^{1/2}$$

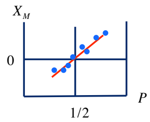
$$\frac{\partial X_M}{\partial P} = \frac{1}{dF/dx\big|_{x=0}} = \left(2\pi\sigma^2\right)^{1/2}$$

$$\sigma^2(X_M) = \sigma^2(P)\left|\frac{\partial X_M}{\partial P}\right|^2 = \frac{1}{4N}\left(2\pi\sigma^2\right) = \frac{\pi}{2}\frac{\sigma^2}{N}$$

$$\sigma^2(\bar{X}) = \frac{\sigma^2}{N}$$



***P* co-varies with X_M :**



Variance of the Median is larger by a factor *π / 2 = 1.57* (for large *N*) than the Variance of the Mean.