Chi-squared = "Badness of Fit"

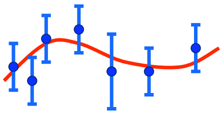
$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{X_i - \mu_i(\alpha)}{\sigma_i} \right)^2 \sim \chi^2_{N-M}$$

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

σ_i = 1 - σ error bar

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

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



N = number of data points

M = number of fitted parameters

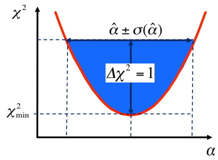
$N - M$ = degrees of freedom

The Dancing χ^2 Landscape

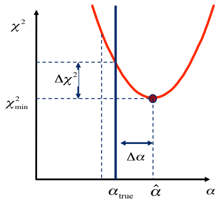
Fit M parameters to N data points.

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

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



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



$$\boxed{\begin{aligned} \hat{\alpha} &\sim G\left(\alpha_{\text{true}}, \sigma^2(\hat{\alpha})\right) \\ \chi^2(\alpha_{true}) &\sim \chi^2_N \\ \chi^2_{\min} \equiv \chi^2(\hat{\alpha}) &\sim \chi^2_{N-M} \\ \Delta\chi^2 \equiv \chi^2(\alpha_{\text{true}}) - \chi^2_{\min} &\sim \chi^2_M \end{aligned}}$$

χ^2 distribution
N degrees of freedom

$$f(x) = \frac{1}{\Gamma(N/2)\, 2^{N/2}}\, x^{(N/2-1)}\, e^{-x/2}$$

$\Gamma(1) = 1 \qquad \Gamma(1/2) = \sqrt{\pi}$

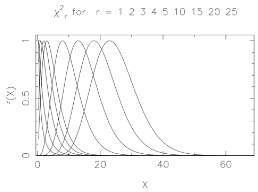
$\Gamma(n) = (n-1)! \qquad \Gamma(\alpha + 1) = \alpha\, \Gamma(\alpha)$

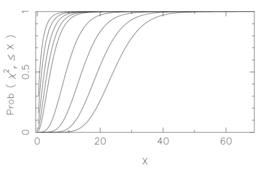
$e.g.\quad \Gamma(3/2) = (1/2)\, \Gamma(1/2) = \sqrt{\pi}/2$

$$\chi^2_1 : \quad f(x) = \left(\frac{e^{-x}}{2\, \pi\, x} \right)^{1/2}$$

$$\chi^2_2 : \quad f(x) = \frac{1}{2} e^{-x/2}$$

$$\boxed{\begin{aligned} \left\langle \chi^2_N \right\rangle &= N \\ \sigma^2\left(\chi^2_N\right) &= 2N \end{aligned}}$$





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.

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

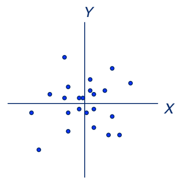
$X^2 \sim \chi^2_1 \qquad Y^2 \sim \chi^2_1$

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

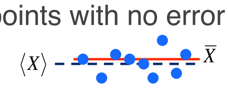
and so on for each

new degree of freedom:

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



Data points with no error bars ☹



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

Sample mean: $\quad \bar{X} \equiv \frac{1}{N} \sum_i X_i \quad$ unbiased: $\quad \langle \bar{X} \rangle = \langle X \rangle$

But σ_i unknown. How to estimate σ?

Variance: $\quad \sigma^2(X) \equiv \left\langle \left(X - \langle X \rangle \right)^2 \right\rangle$

Try: $\quad s^2 \equiv \frac{1}{N} \sum_i \left(X_i - \bar{X} \right)^2$

Is $\quad \langle s^2 \rangle = \sigma^2 \quad$?

No. $\quad \langle s^2 \rangle < \sigma^2 \quad$ Must correct for a bias.

Sample Variance S^2: Unbiased for σ^2

$S^2 \equiv A \sum_{i=1}^{N} (X_i - \bar{X})^2 \quad$ Pick A so that $\langle S^2 \rangle = A \sum_{i=1}^{N} \langle (X_i - \bar{X})^2 \rangle = \sigma^2$

$$\begin{aligned} \langle (X_i - \bar{X})^2 \rangle &= \left\langle \left[(X_i - \langle X \rangle) - (\bar{X} - \langle X \rangle) \right]^2 \right\rangle \\ &= \left\langle (X_i - \langle X \rangle)^2 - 2\,(X_i - \langle X \rangle)\,(\bar{X} - \langle X \rangle) + (\bar{X} - \langle X \rangle)^2 \right\rangle \\ &= \sigma^2(X_i) - 2\,\mathrm{Cov}(X_i, \bar{X}) + \sigma^2(\bar{X}) \\ &= \sigma^2 \quad - 2\frac{\sigma^2}{N} \quad + \frac{\sigma^2}{N} \\ &= \left(1 - \frac{1}{N}\right)\sigma^2 = \left(\frac{N-1}{N}\right)\sigma^2 \end{aligned}$$

Note: $\quad \mathrm{Cov}(X_i, \bar{X}) = \frac{\sigma^2}{N}$

$\therefore \quad \langle S^2 \rangle = A\,(N-1)\,\sigma^2 \quad$ Pick $\quad A = \frac{1}{N-1}$

$$\boxed{S^2 \equiv \frac{1}{N-1} \sum_{i=1}^{N} \left(X_i - \bar{X} \right)^2}$$

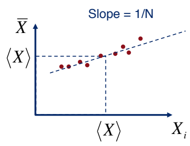
Evaluation of Cov(X_i , $\overline{X}$)

$$\mathrm{Cov}(X_i,\overline{X}) \equiv \left\langle (X_i - \langle X_i \rangle)\,(\overline{X} - \langle \overline{X} \rangle) \right\rangle$$

$$\text{Note: } \langle X_i \rangle = \langle \overline{X} \rangle = \langle X \rangle$$

$$\text{Shift coords to put } \langle X \rangle = 0 \ :$$

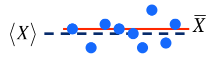
$$\mathrm{Cov}(X_i,\overline{X}) = \left\langle (X_i - 0)\,(\overline{X} - 0) \right\rangle = \left\langle X_i \frac{1}{N} \sum_k X_k \right\rangle = \frac{1}{N} \sum_k \langle X_i\, X_k \rangle = \frac{1}{N} \sum_k \sigma^2 \delta_{ik} = \frac{\sigma^2}{N}$$



$$\mathrm{Cov}(X_i,X_j) = \sigma^2\, \delta_{ij}$$

Sample Variance S^2 : Unbiased for σ^2

$$\boxed{S^2 \equiv \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \overline{X})^2}$$



Why $\dfrac{1}{N\text{-}1}$, not $\dfrac{1}{N}$?

Because $\overline{X}$ "chases" the dancing data points, removing 1 "degree-of-freedom" from the dance.

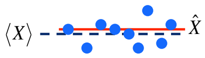
$$\boxed{\begin{aligned} S^2 &\sim \frac{\sigma^2}{N-1} \chi^2_{N-1} \\ \langle S^2 \rangle &= \frac{\sigma^2}{N-1} \langle \chi^2_{N-1} \rangle \\ &= \frac{\sigma^2}{N-1} (N-1) = \sigma^2 \end{aligned}}$$

$$\boxed{\begin{aligned} \mathrm{Var}[S^2] &= \left(\frac{\sigma^2}{N-1}\right)^2 \mathrm{Var}[\chi^2_{N-1}] \\ &= \left(\frac{\sigma^2}{N-1}\right)^2 2(N-1) = \frac{2\sigma^4}{N-1} \\ \frac{\sigma(S^2)}{\langle S^2 \rangle} &= \left(\frac{2}{N-1}\right)^{1/2} = \text{fractional accuracy} \end{aligned}}$$

Degrees of Freedom (DoF)

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

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

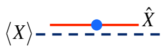


If $\langle X \rangle$ unknown, use $\hat{X}$ instead:

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

If N=1 data point: $\hat{X}$=X_1

$$\left(\frac{X_1 - \langle X \rangle}{\sigma_1} \right)^2 \sim \chi^2_1. \quad 1 \text{ degree of freedom}$$



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

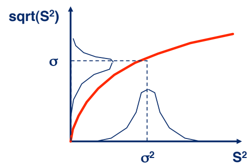
Fit M parameters to N data points:

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

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

sqrt(S^2) is biased

- The sample variance S^2 is unbiased for σ^2.
- Is the sqrt(S^2) unbiased for σ ?
- No. The square root introduces a bias:



$$\left\langle \sqrt{S^2} \right\rangle < \sigma, \text{ even though } \sqrt{\sigma^2} = \sigma.$$

Example: Correct the Bias in sqrt(S^2)

Define g(x) = sqrt(x) $\quad \overline{S} \equiv g(S^2)$

and its derivatives:

$$g(x) = x^{1/2}, \quad g'(x) = \frac{1}{2} x^{-1/2}, \quad g''(x) = -\frac{1}{4} x^{-3/2}$$

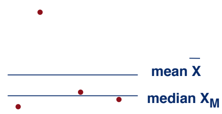
Evaluate the bias:

$$\begin{aligned} \langle \overline{S} \rangle &= g(\langle S^2 \rangle) + \frac{g''(\langle S^2 \rangle)}{2} \mathrm{Var}(S^2) + \ldots \\ &= g(\sigma^2) + \frac{g''(\sigma^2)}{2} \cdot \frac{2\sigma^4}{N-1} + \ldots \\ &= \sigma - \frac{1}{8\sigma^3} \cdot \frac{2\sigma^4}{N-1} + \ldots = \sigma \left(1 - \frac{1}{4\,(N-1)} + \ldots \right) = \sigma \left(\frac{4N-5}{4N-4} \right) + \ldots \end{aligned}$$

$$\boxed{\text{Bias-corrected:} \quad \overline{S} \equiv \left(\frac{4N-4}{4N-5} \right) (S^2)^{1/2}}$$

"Robust" estimation methods

- Robust => less sensitive to "bad" data.
- Example: using **median** rather than **mean**:



- Sample Mean $\overline{X}$ minimizes the Sample Variance:

$$S^2 \equiv \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \mu_i)^2$$

$$\frac{\partial}{\partial \mu} \left[\sum_{i=1}^{N} (X_i - \mu)^2 \right] = 0$$

$$\text{for } \mu = \overline{X}$$

- Median X_M minimizes the "Mean Absolute Deviation" :

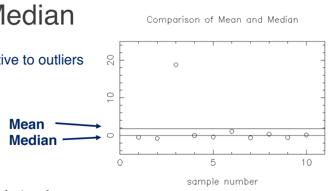
$$\mathrm{MAD} \equiv \frac{1}{N} \sum_{i=1}^{N} |X_i - \mu_i|$$

$$\frac{\partial}{\partial \mu} \left[\sum_{i=1}^{N} |X_i - \mu| \right] = 0$$

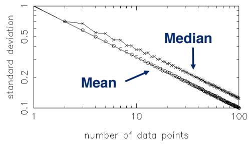
$$\text{for } \mu = X_M \equiv \mathrm{Median}(X_i)$$

Mean vs Median

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

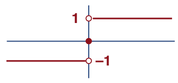


- The median is **unbiased**, but **not** a minimum-variance estimator.
- Note how the standard deviations of the median and of the mean vary with sample size.



"Proof" that the Median minimises the MAD

$$H(x) \equiv \begin{cases} +1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases} \qquad \frac{\mathrm{d}H}{\mathrm{d}x} = 2\delta(x)$$



$$MAD \equiv \frac{1}{N}\sum_{i=1}^{N}\left|\mu - X_i\right| = \frac{1}{N}\sum_{i=1}^{N}(\mu - X_i)\,H(\mu - X_i)$$

$$\frac{d\,MAD}{d\mu} = \frac{1}{N}\sum_{i=1}^{N}\left[H(\mu - X_i) + \cancel{(\mu - X_i)\,H'(\mu - X_i)}\right]$$

$= 0$

$$= \frac{1}{N}\sum_{i=1}^{N} H(\mu - X_i) \qquad \text{since } H'(x) = 0 \text{ whenever } x \neq 0$$

$$= 0 \quad \text{if} \quad \mu = \text{median}(X_i)$$

Finding the Median without Sorting

$$\frac{X_i - X_M}{\left|X_i - X_M\right|}$$

- A useful application:

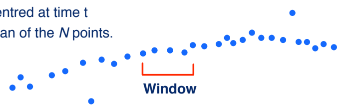
Since $\displaystyle\sum_{i=1}^{N}\frac{X_i - X_M}{\left|X_i - X_M\right|} = 0$, first make a guess at X_M.

Then estimate a new $X_M = \dfrac{\displaystyle\sum_{i=1}^{N}\frac{X_i}{\left|X_i - X_M\right|}}{\displaystyle\sum_{i=1}^{N}\frac{1}{\left|X_i - X_M\right|}}$,

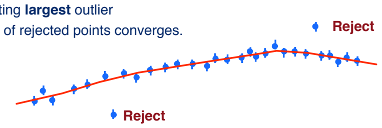
and iterate to convergence.

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 ± K σ or more
- Repeat fit omitting **largest** outlier
- Iterate until set of rejected points converges.



Various "Badness-of-Fit" Statistics

Sample Variance — mean

$$S^2 \equiv \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \mu_i)^2 \quad \rightarrow \quad \bar{X}$$

Chi-squared — optimal average

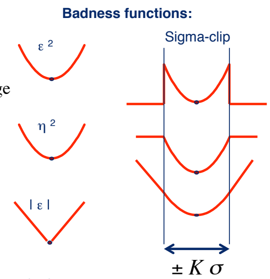
$$\chi^2 \equiv \sum_{i=1}^{N}\left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \quad \rightarrow \quad \hat{X}$$

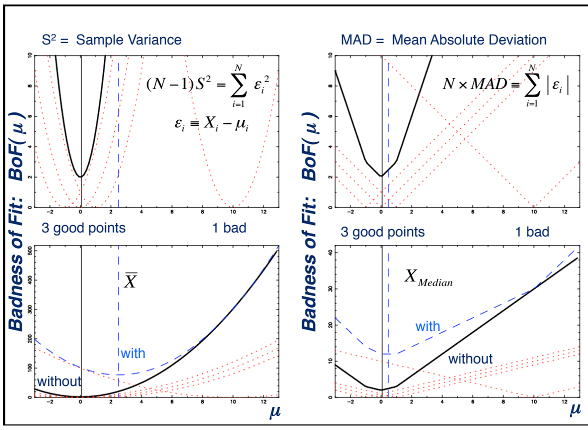
Mean Absolute Deviation — median

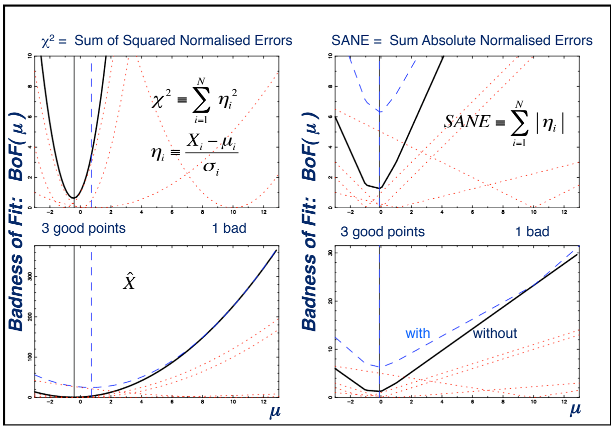
$$MAD \equiv \frac{1}{N}\sum_{i=1}^{N}\left|X_i - \mu_i\right| \quad \rightarrow \quad X_M$$

Sum Absolute Normalised Errors:

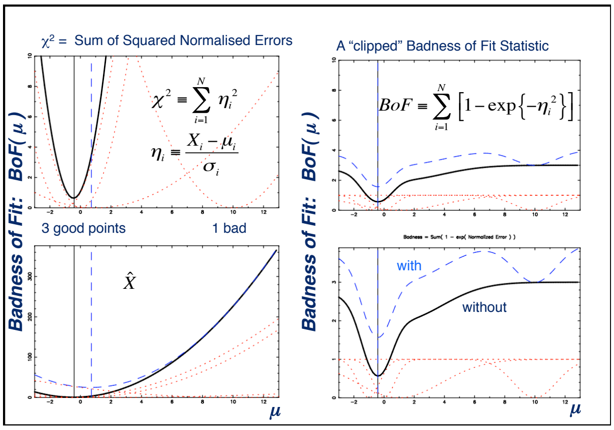
$$SANE \equiv \sum_{i=1}^{N}\left|\frac{X_i - \mu_i}{\sigma_i}\right|$$







χ² = Sum of Squared Normalised Errors
SANE = Sum Absolute Normalised Errors
$\chi^2 \equiv \sum_{i=1}^{N} \eta_i^2$
$\eta_i \equiv \frac{X_i - \mu_i}{\sigma_i}$
$SANE \equiv \sum_{i=1}^{N} |\eta_i|$
Badness of Fit: BoF(μ)
3 good points
1 bad
$\hat{X}$
with
without
μ



χ² = Sum of Squared Normalised Errors
A "clipped" Badness of Fit Statistic
$\chi^2 \equiv \sum_{i=1}^{N} \eta_i^2$
$\eta_i \equiv \frac{X_i - \mu_i}{\sigma_i}$
$BoF \equiv \sum_{i=1}^{N} \left[1 - \exp\left\{-\eta_i^2\right\}\right]$
Badness of Fit: BoF(μ)
3 good points
1 bad
$\hat{X}$
with
without
μ