Computational Astrophysics I: Introduction and basic concepts

Helge Todt

Astrophysics
Institute of Physics and Astronomy
University of Potsdam

SoSe 2024, 9.7.2024

Insertion: data analysis

Recursive mean I

arithmetic mean

$$\langle x \rangle = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1}$$

Problem: calculation of the mean for measured data

- Eq. (1) must be evaluated again for every new data point
- for $n \gg 1$ and at the same time $x_i \ll 1$ numerical inaccuracy for strict use of Eq. (1) because of $\rightarrow$ saturation in x_i

$\rightarrow$ hence: definition of the recursive mean:

$$\langle x \rangle_i = \frac{i-1}{i} \langle x \rangle_{i-1} + \frac{1}{i} x_i \tag{2}$$

proof:

$$i-1 : \langle x \rangle_{i-1} = \frac{x_1 + \ldots + x_{i-1}}{i-1} \tag{3}$$

$$i : \langle x \rangle_i = \frac{x_1 + \ldots + x_{i-1} + x_i}{i} \tag{4}$$

$$= \frac{(i-1)\frac{x_1+\ldots x_{i-1}}{i-1} + x_i}{i} \tag{5}$$

$$= \frac{i-1}{i}\langle x \rangle_{i-1} + \frac{x_i}{i} \tag{6}$$

q.e.d.

analogously: **recursive variance**

$$\sigma_i^2 = \frac{i-1}{i}\sigma_{i-1}^2 + \frac{1}{i-1}(x_i - \langle x\rangle_i)^2 \tag{7}$$

proof similar as for recursive mean

correction of a single value:

$$\begin{aligned} \langle x\rangle_{\text{new}} &= \langle x\rangle_{\text{old}} + \frac{x_{\text{new}} - x_{\text{old}}}{n} \\ \sigma^2_{\text{new}} &= \sigma^2_{\text{old}} + \frac{x^2_{\text{new}} - x^2_{\text{old}}}{n} - \frac{x_{\text{new}} - x_{\text{old}}}{n}\left(\langle x\rangle_{\text{old}} + \frac{x_{\text{new}} - x_{\text{old}}}{n}\right) \end{aligned} \tag{8}$$

proof for the correction of the mean:

$$\begin{aligned} \langle x\rangle_{\text{new}} &= \frac{1}{n}\left(\sum_{i=1}^{n} x_i - x_{\text{old}} + x_{\text{new}}\right) = \\ &= \langle x\rangle_{\text{old}} + \frac{x_{\text{new}} - x_{\text{old}}}{n} \qquad \text{q.e.d.} \end{aligned} \tag{9}$$

We already know

Straight line fit without errors

$$y = b \cdot x + a \tag{10}$$

$$\text{with slope} \quad b = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \langle x\rangle)(y_i - \langle y\rangle)}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \langle x\rangle)^2} \tag{11}$$

$$\text{and} \quad a = \langle y\rangle - b \cdot \langle x\rangle \tag{12}$$

quality of fit $y = a + bx$ measured by χ^2:

$$\chi^2(a,b) = \sum_{i=1}^{n}\left(\frac{y_i - a - bx_i}{\sigma_i}\right)^2 \tag{13}$$

with error σ_i in measuring of y_i (x_i exact)

Best fit for χ^2 minimum, hence (see also *Numerical Recipes*)

$$0 \stackrel{!}{=} \frac{\partial \chi^2}{\partial a} = -2 \sum_{i=1}^{n} \frac{y_i - a - b x_i}{\sigma_i^2} \tag{14}$$

$$0 \stackrel{!}{=} \frac{\partial \chi^2}{\partial b} = -2 \sum_{i=1}^{n} \frac{x_i (y_i - a - b x_i)}{\sigma_i^2} \tag{15}$$

can be rewritten as system of equations:

$$a \sum_{i=1}^{n} \frac{1}{\sigma_i^2} + b \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} = \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2} \tag{16}$$

$$a \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} + b \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} = \sum_{i=1}^{n} \frac{x_i y_i}{\sigma_i^2} \tag{17}$$

$$\tag{18}$$

solution for the system:

$$a = \frac{\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} \sum_{i=1}^{n} \frac{x_i y_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} \right)^2} \tag{19}$$

$$b = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{x_i y_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} \right)^2} \tag{20}$$

errors in a and b from error propagation for a quantity f:

$$\sigma_f^2 = \sum_{i=1}^{n} \sigma_i^2 \left(\frac{\partial f}{\partial y_i} \right)^2 \tag{21}$$

where

$$\frac{\partial a}{\partial y_i} = \frac{\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} - x_i \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sigma_i^2 \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} \right)^2 \right)} \tag{22}$$

$$\frac{\partial b}{\partial y_i} = \frac{x_i \sum_{i=1}^{n} \frac{1}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sigma_i^2 \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} \right)^2 \right)} \tag{23}$$

adding up according to Eq. (21)

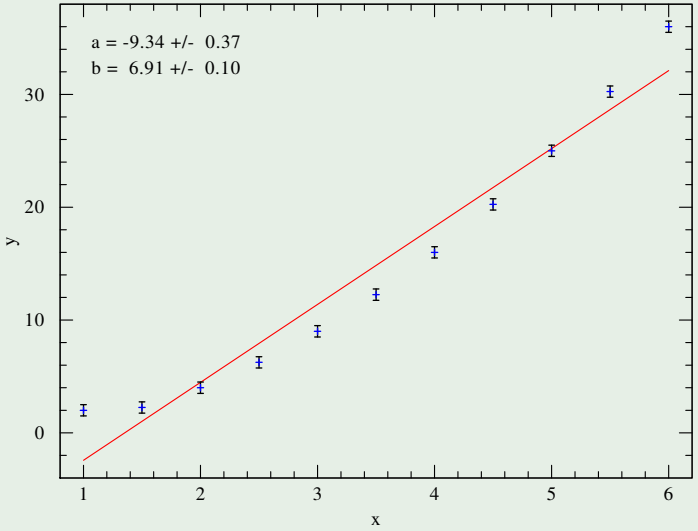
$$\sigma_a^2 = \frac{\sum_{i=1}^n \frac{x_i^2}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2}\right)^2} \tag{24}$$

$$\sigma_b^2 = \frac{\sum_{i=1}^n \frac{1}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2}\right)^2} \tag{25}$$

Caution!

This (purely formal) error may drastically underestimate the real error in a, b!

Example: bad fit but small error



→ small formal error, as error in the measurements is small

but:
→ model does not fit to data

Our original case: errors σ_i not available.

Then: Set $\sigma_i = 1$ in equations for a, b and multiply factor $\sqrt{\frac{\chi^2}{n-2}}$ to the formal errors

$$\sigma_a^2 = \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\sqrt{\frac{\chi^2}{n-2}} \tag{26}$$

$$\sigma_b^2 = \frac{n}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\sqrt{\frac{\chi^2}{n-2}} \tag{27}$$

$$\text{where} \quad \chi^2 = \sum_{i=1}^n \left(y_i - a - bx_i\right)^2 \tag{28}$$

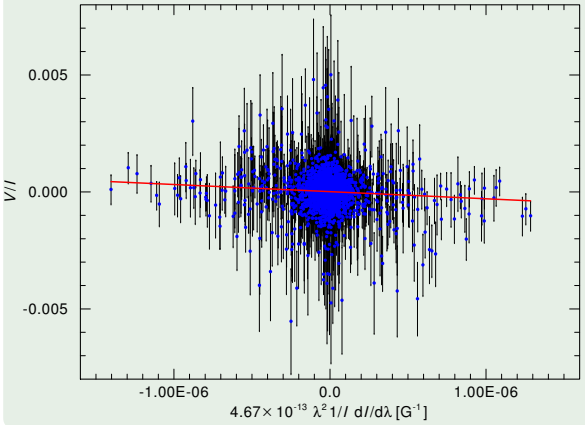
Bootstrapping I

Estimating the errors in fit variables (e.g., the slope b)
Methods:

1. formal error from errors in measuring in $y_i \rightarrow$ without consideration of the fit quality χ^2
2. error from χ^2 without consideration of the errors in measuring y_i

$\rightarrow$ usually results in an underestimation of σ_b

Example: measuring the magnetic field from polarization



Stokes I : intensity

Stokes $V = I_R - I_L$ (so: right-hand circularly polarized – left-hand circularly polarized)

$\rightarrow V/I$: fraction of polarized light

$\rightarrow \lambda^2/I \; dI/d\lambda$: Zeeman shift

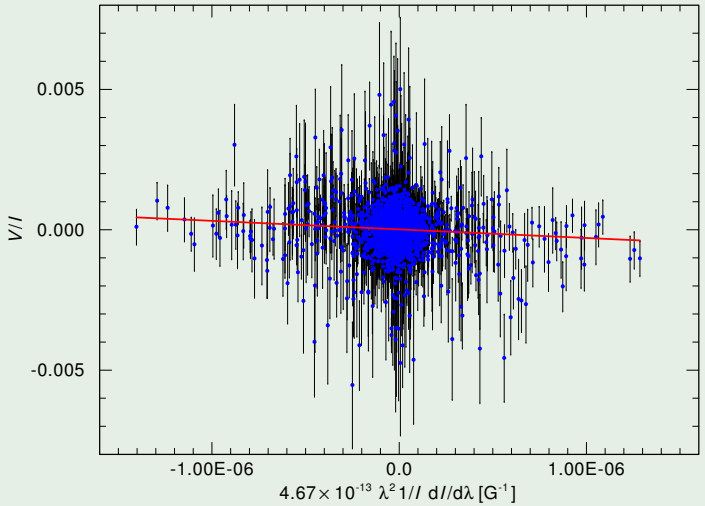
Idea: for broad spectral lines (Balmer lines in WD, WR emission lines) Zeeman splitting not directly detectable because of Doppler shifts. Therefore: measure "line displacement" per pixel per line together with V/I.

No B-field $\rightarrow$ no correlation. Otherwise, slope gives longitudinal $\langle B_z \rangle$

$$\frac{V}{I} = -\frac{g_{\rm eff} e \lambda^2}{4\pi m_e c^2} \frac{1}{I} \frac{dI}{d\lambda} \langle B_z \rangle \tag{29}$$

with average effective Landé factor

Example: measuring the magnetic field from polarization



→ slope dominated by only few data points?

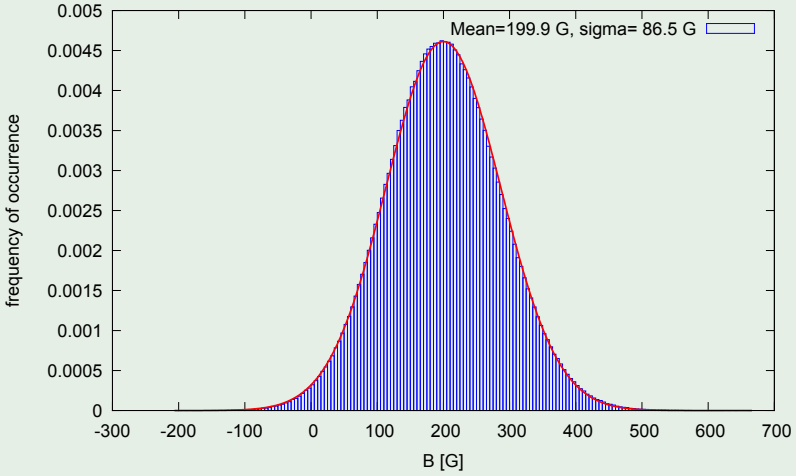
Problem: the distribution of b is usually not known
Idea: construct a distribution with help of *Bootstrapping*

Construction of a Bootstrapping distribution

random sample j by random drawing of n data (x_i, y_i) from the complete set of n data with *repetition* and determination of b_j. Repeating m-times this procedure, where $m \gg n$.
In each random sample are only $\sim 1/e \approx 37\,\%$ of the original data because of *repetitions*.
$\rightarrow$ result: sample of m measured quantity b_j.

Then, determination of the expectation value, variance, etc. for the obtained bootstrapping sample, e.g.,

Example: magnetic field B_z from polarization



$\rightarrow$ interpretation: $\mu > 2\sigma \rightarrow$ marginal detection (5% proability that actually $B_z = 0$)