STACUP-2023

Hands-On
Session-3

Python Based Statistics and its Application in Astrophysics & Cosmology

Darshan Kumar Beniwal
Research Scholar,
Dept of Physics and Astrophysics,
University of Delhi, Delhi-110007

Date: October 18 2023

Least Square Fitting

Observational dataset:

x	y	σ_{y}
1.0	2.3	0.08
2.0	4.1	0.12
3.0	6.2	0.20
4.0	8.1	0.16
5.0	10.0	0.28

Assumptions: Observed data are normal distributed with center y and width σ_y

Theoretical Model: $y^{\mathrm{th}} = a + bx$

Least Square Fitting

- Theoretical Term: $y^{\mathrm{th}} = a + bx$
- Observational Term: $y_i^{\mathrm{obs}}, \quad \sigma_{y_i}$

The probability of obtaining the observed value (y_i^{obs}) is

$$\mathrm{Prob}_{a,b}\left(y_i\right) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\left(y_i^{\mathrm{obs}} - a - bx_i\right)^2 / 2\sigma_y^2}$$

$$\mathcal{L}(x_i; a, b) = \prod_{i=1}^{n} \mathrm{Prob}_{a,b}\left(y_i\right) \Rightarrow \text{Maximise it}$$

...... gives the parameters values for which observed data have the highest probability

The probability of obtaining the observed value (y_i^{obs}) is

$$\mathcal{L}(x_i; a, b) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_{y_i}^2}} \exp\left\{-\frac{1}{2}\left[\frac{\left(y_i^{\mathrm{obs}} - a - bx_i\right)^2}{\sigma_{y_i}^2}\right]\right\}$$

$$-2\ln\mathcal{L}(x_i; a, b) = \sum_{i=1}^{N} \frac{\left(y_i^{\mathrm{obs}} - a - bx_i\right)^2}{\sigma_{y_i}^2} \equiv \chi^2$$

To estimate parameters: either maximize the likelihood or minimise the Chi-square

..... log *makes math easier, doesn't change answer (monotonic)!*

Chi-square Test: describes the goodness-of-fit of the data to the model.

$$\chi^2 = \sum_i \left(\frac{\text{observed} - \text{expected}}{\text{error}} \right)^2$$

- Observational dataset: x_i, $y_{\rm obs}(x_i)$, σ_{y_i}
- Theoretical model: $y_{\rm th}(x_i; a, b) = f(x_i; a, b)$
- Define

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_{\rm obs}(x_i) - y_{\rm th}(x_i; a, b)}{\sigma_{y_i}} \right)^2$$

Maximum Likelihood Method ...cntd

- Chi-square:

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(y_i^{\rm obs} - a - bx_i\right)^2}{\sigma_{y_i}^2}$$

- Minimize the Chi-square: $\dfrac{\partial\chi^2}{\partial a}$ $\dfrac{\partial\chi^2}{\partial b}$

$$a = \frac{\sum wx^2 \sum wy - \sum wx \sum wxy}{\Sigma w \sum wx^2 - \left(\sum wx\right)^2} \qquad b = \frac{\sum w \sum wxy - \sum wx \sum wy}{\Sigma w \sum wx^2 - \left(\sum wx\right)^2}$$

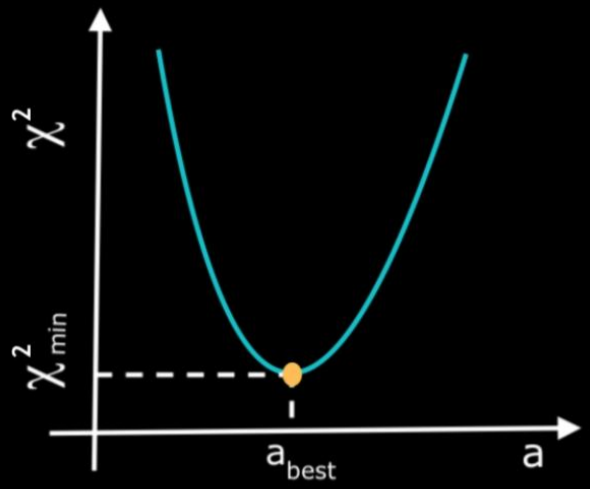
$$\sigma_a = \sqrt{\frac{\sum wx^2}{\Sigma w \sum wx^2 - \left(\sum wx\right)^2}} \qquad \sigma_b = \sqrt{\frac{\Sigma w}{\Sigma w \sum wx^2 - \left(\sum wx\right)^2}}$$

$$w = 1/\sigma_{y_i}^2$$

Chi-square:

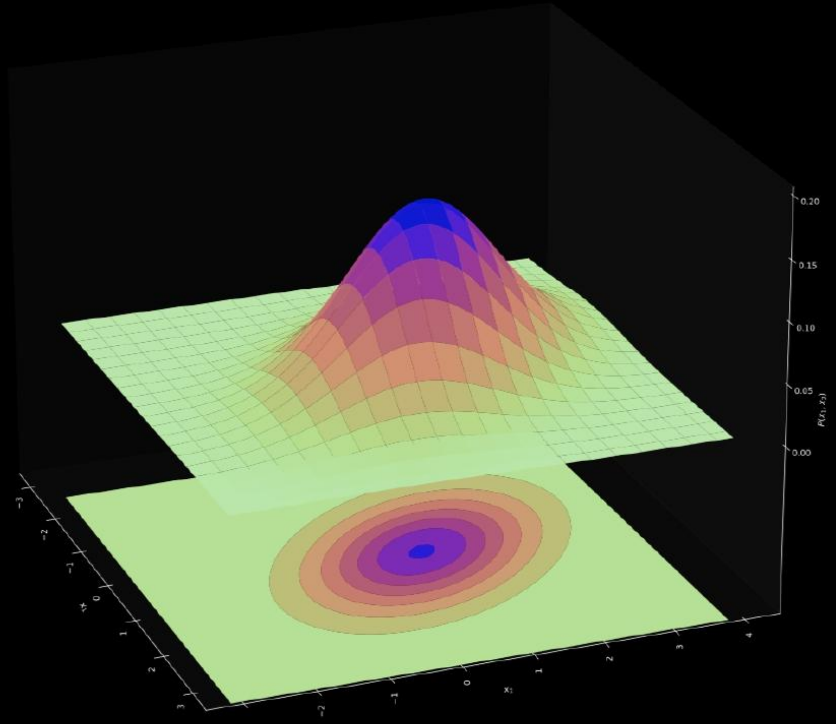
$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_{\mathrm{obs}}(x_i) - y_{\mathrm{th}}(x_i; a, b)}{\sigma_{y_i}} \right)^2 \Rightarrow \chi^2_{\min} \rightarrow \text{Best fit value of parameter}$$

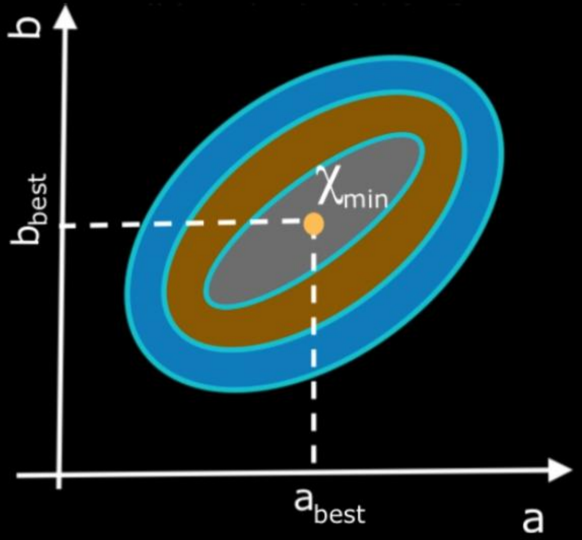
Case:1 One Parameter Model:



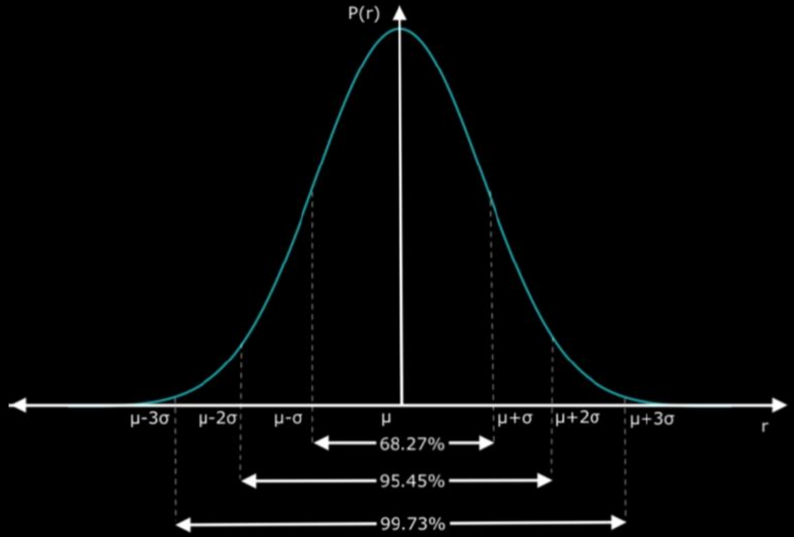
Chi-square Statistics ...cntd

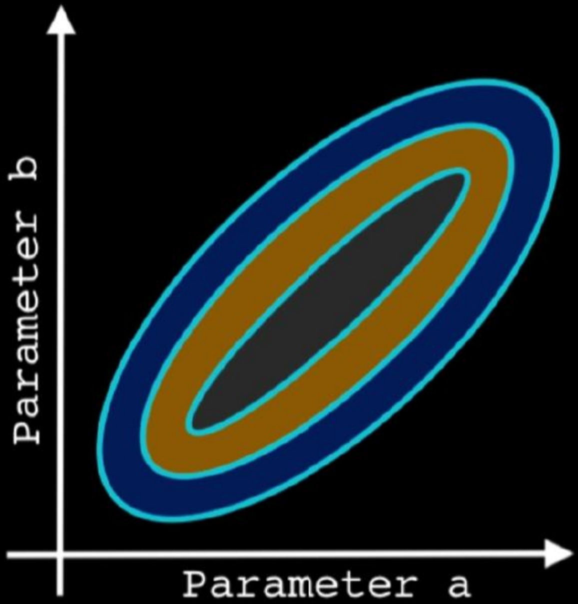
Case:2 Two Parameters Model:



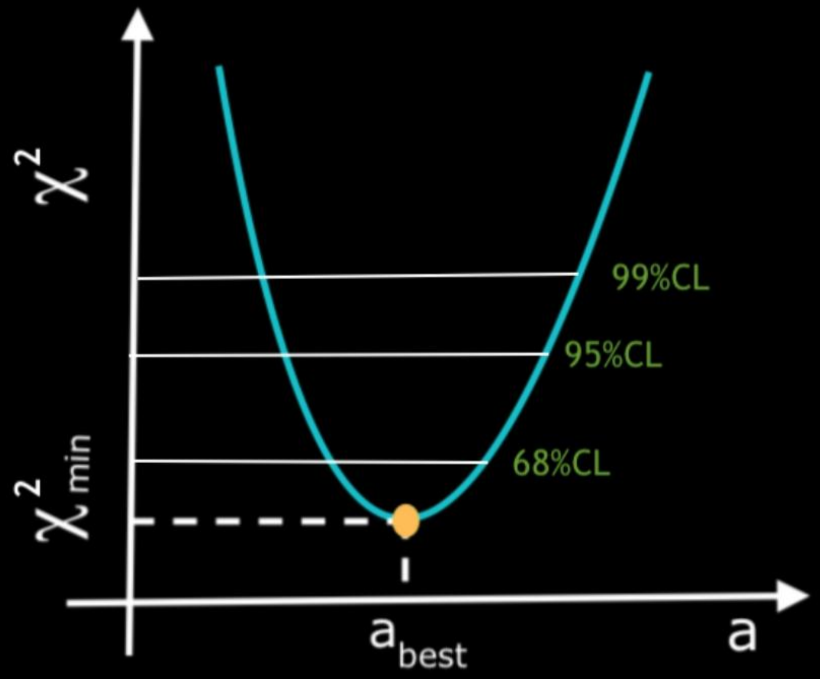


Confidence Intervals: Range of estimates for an unknown parameter.





Chi-squared Distribution with Sigma Values: $\chi^2_{n\sigma} = \chi^2_{\text{min}} + \Delta\chi^2_{n\sigma}$

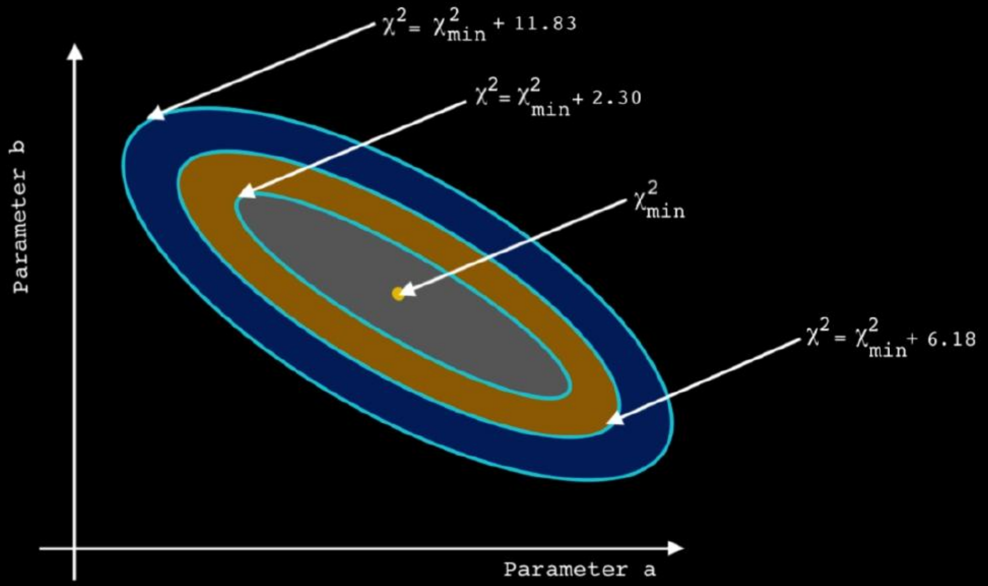


Dimensionality	1σ	2σ	3σ
1	1.00	4.00	9.00
2	2.30	6.18	11.83
3	3.53	8.02	14.16
4	4.72	9.72	16.25
5	5.89	11.31	18.21

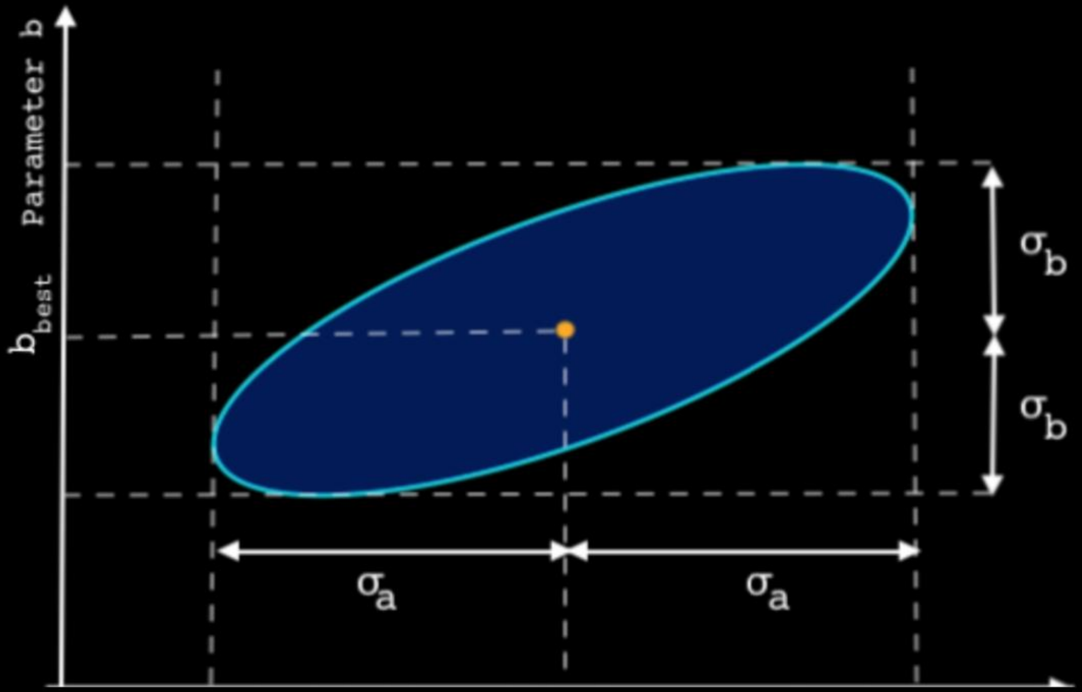
Chi-squared distribution $\Delta\chi^2_{n\sigma}$ upto 5 parameters (5D).

Chi-squared Distribution with Sigma Values: $\chi^2_{n\sigma} = \chi^2_{\text{min}} + \Delta\chi^2_{n\sigma}$

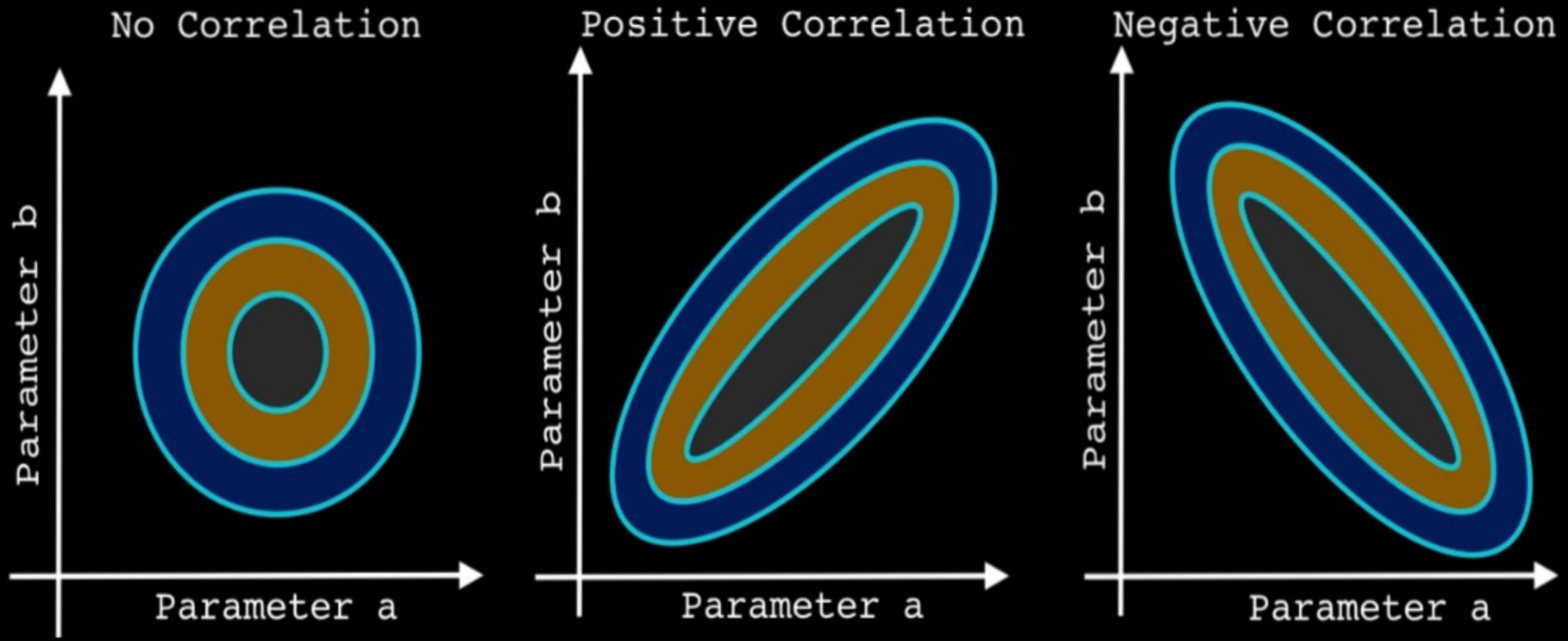
Two-parameters model:



Parameter error estimation:



Correlation among parameters:



c). Example: Hand-on Session

Example-1: Hand-on Session

Mock dataset:

x	y	σ_{y}
1.0	2.3	0.08
2.0	4.1	0.12
3.0	6.2	0.20
4.0	8.1	0.16
5.0	10.0	0.28

Theoretical Model: $y^{\mathrm{th}} = a + bx$

Best Fit values: $a = - - - \pm - - -, \quad b = - - - \pm - - -$

Key Takeaways

Observational dataset: x_i, $y_{\rm obs}(x_i)$, σ_{y_i}

Theoretical model: $y_{\rm th}(x_i; a, b) = f(x_i; a, b)$

Define Chi-square: $\chi^2 = \sum_{i=1}^{N} \left(\frac{y_{\rm obs}(x_i) - y_{\rm th}(x_i; a, b)}{\sigma_{y_i}} \right)^2$

Minimize Chi-square: $\chi^2_{\rm min} \Rightarrow$ **Best fit value of parameters**

Draw Confidence Level: $\chi^2_{n\sigma} = \chi^2_{\text{min}} + \Delta\chi^2_{n\sigma}$

Error in parameters: $a = a_{\rm best} \pm \sigma_a$ **and** $b = b_{\rm best} \pm \sigma_b$