Exploring Astrophysical Data: Statistical Analysis and Practical Applications in Python

Workshop on Python Programming in Astronomy, Astrophysics & Cosmology

Darshan Kumar Beniwal

July 20th 2023

Organized by

Department of Physics and Astrophysics,
University of Delhi, Delhi, India-110007

IUCAA Centre for Astronomy Research & Development (ICARD)
GLA University, Mathura, July 20th-22th, 2023

Why do we study statistics in Astrophysics & Cosmology?

Astrophysics & Cosmology in the Era of Big Data

- **Motivations:**
  - Availability of vast and precise data
  - Different kinds of Statistical techniques
- **Goal:**
  - To understand the universe with high accuracy,
  - To extract the maximum amount of information from the observational data.
- **Methodology used:**
  - Frequentist Statistics: Likelihood or Chi-square.
  - Bayesian Statistics: Markov Chain Monte Carlo.

Outline

1. Introduction to Statistics

2. Least Square Fitting

    a). Maximum Likelihood Estimator (MLE)

    b). Minimum Chi-square Statistics

    c). Example: Hand-on Session

3. Bayesian Statistics

Introduction to Statistics

Statistics: *Study of collecting, analyzing, interpreting, and presenting numerical data.*

Basic form of Statistics:

- Descriptive Statistics: *Summarizing and describing the data*
  - mean, median, mode
  - standard deviation, variance
  - achieved with the help of charts, graphs, tables, etc.
- Inferential Statistics: *Generalizations or draw conclusions about a larger dataset.*
  - testing hypotheses
  - estimating parameters
  - achieved by probability.

Descriptive Statistics:

1	2	1	1	3	4	100

- Mean: *Average of observed values*: 16
- Median: *Value which divides the dataset into half* : 2 [Data should be in Ascending order]
- Mode: *Value which occurs with greatest frequency*: 1

outlier: 100

- Mean: 2
- Median: 1.5
- Mode: 1

Descriptive Statistics:

1	2	1	1	3	4

- Standard Deviation: *Dispersion of data points from the mean of a dataset.*

$$\sigma = \sqrt{\sum_{i=0}^{n} \frac{(x_i - \mu)^2}{n}} = 1.155$$

- Variance: *Average of the squared differences from the mean.*

$$\sigma^2 = \sum_{i=0}^{n} \frac{(x_i - \mu)^2}{n} = 1.334$$

Inferential Statistics: *Generalizations or draw conclusions about a larger dataset.*

1. Hypothesis Testing:

Year	2018	2019	2020	2021	2022
Rainfall (inches)	8	5	7	5	6

$\Rightarrow$ Test the hypothesis that the average rainfall in a given area is 8 inches?

- Null Hypothesis (H_0): *The average annual rainfall from 2018-2022 is the same as the overall average annual rainfall of 8 inches.*
- Alternative Hypothesis (H_A): *The average annual rainfall from 2018-2022 is the same not as the overall average annual rainfall of 8 inches*

Inferential Statistics:

1. Hypothesis Testing:

Year	2018	2019	2020	2021	2022
Rainfall (inches)	8	5	7	5	6

$\Rightarrow$ Test the hypothesis that the average rainfall in a given area is 8 inches?

Sample mean ($\bar{x}$) $= 6.2$, sample size (n) $= 5$

Sample Standard Deviation (σ) $= 1.30$, Population mean (μ) $= 8$

t-test:

$$t_{obs} \equiv \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = -3.10$$

Degree of freedom: $d \equiv N - 1 = 4$

Inferential Statistics:

1. Hypothesis Testing:

Year	2018	2019	2020	2021	2022
Rainfall (inches)	8	5	7	5	6

$\Rightarrow$ Test the hypothesis that the average rainfall in a given area is 8 inches?

$t_{obs} < t_{tab}$

Reject the null hypothesis.

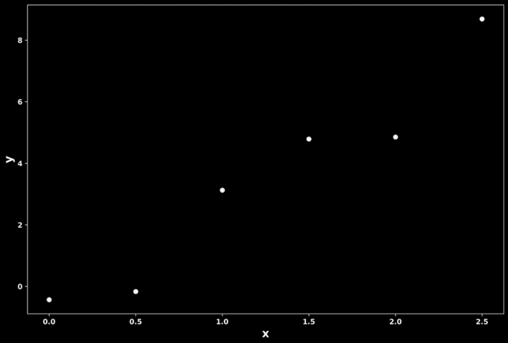
Difference is not purely due to the random error.

Inferential Statistics:

2. Parameter Estimation:

Observational Dataset

i	x	y
1	0.0	-0.4326
2	0.5	-0.1656
3	1.0	3.1253
4	1.5	4.7877
5	2.0	4.8535
6	2.5	8.6909



Least Square Fitting

Parameter Estimation

i	x	y
1	0.0	-0.4326
2	0.5	-0.1656
3	1.0	3.1253
4	1.5	4.7877
5	2.0	4.8535
6	2.5	8.6909

- y_{obs}: Observational dataset
- y_{th}: Theoretical Model: $y = a + bx$
- AIM: $a =?$ and $b =?$

Least Square Fit: *Minimizing the sum of the squares of the residuals.*

- Observable quantity: y^{obs}
- Theoretical quantity: $y^{\mathrm{th}}(x_i; a, b)$

$$\min_{a,b} : \sum_{i=1}^{n} \left[y_i^{\mathrm{obs}} - y^{\mathrm{tr}}(x_i; a, b) \right]^2$$

$$\text{Residuals} = \sum_{i=1}^{n} \left[y_i^{\mathrm{obs}} - (a + bx_i) \right]^2$$

Least Square Fit: *Minimizing the sum of the squares of the residuals.*

- Observable quantity: y^{obs}
- Theoretical quantity: $y^{\mathrm{th}}(x_i; a, b)$

$$\frac{\partial \text{Residuals}}{\partial a} = 0 \qquad \frac{\partial \text{Residuals}}{\partial b} = 0$$

$$\Rightarrow aN + b\sum x_i = \sum y_i^{\mathrm{obs}} \qquad \Rightarrow a\sum x_i + b\sum x_i^2 = \sum x_i y_i^{\mathrm{obs}}$$

Least Square Fit: *Minimizing the sum of the squares of the residuals.*

$$\boxed{\begin{aligned} a &= \frac{\sum x^2 \sum y - \sum x \sum xy}{N \sum x^2 - (\sum x)^2} \\ b &= \frac{N \sum xy - \sum x \sum y}{N \sum x^2 - (\sum x)^2} \end{aligned}}$$

$$\begin{bmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum (x_i y_i) \end{bmatrix}$$

$$X \cdot P = Y \quad \Rightarrow P = X^{-1} \cdot Y$$

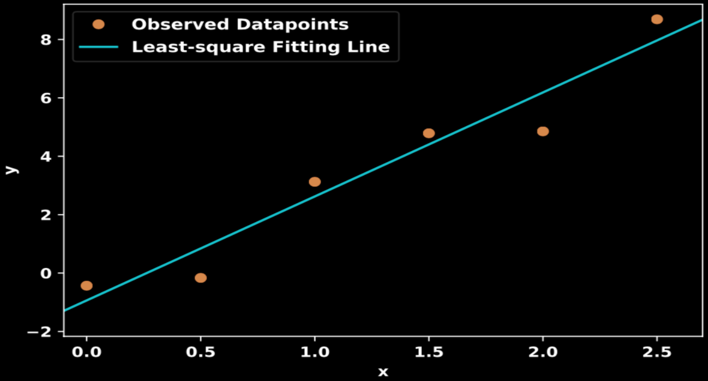
i	1	2	3	4	5	6
x	0	0.5	1.0	1.5	2.0	2.5
y	-0.4326	-0.1656	3.1253	4.7877	4.8535	8.6909

- $n = 6 \quad \sum x_i = 7.5, \quad \sum y_i = 20.8593, \quad \sum x_i^2 = 13.75, \quad \sum x_i y_i = 41.6584$

$$\begin{bmatrix} 6 & 7.5 \\ 7.5 & 13.75 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 20.8593 \\ 41.6584 \end{bmatrix} \Rightarrow \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 6 & 7.5 \\ 7.5 & 13.75 \end{bmatrix}^{-1} \times \begin{bmatrix} 20.8593 \\ 41.6584 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -0.975 \\ 3.561 \end{bmatrix}$$

$$y = -0.975 + 3.561x$$



Observed Datapoints
Least-square Fitting Line
8
6
4
2
0
−2
y
0.0
0.5
1.0
1.5
2.0
2.5
x

What if uncertainties or errors are associated with the observed data?

- Key-points of Error Bars:
  - How far the reported value is from the true (error-free) value?
  - Errors introduced by random fluctuations in the measurement: random errors.
  - Errors due to a faulty calibration of equipment or observer biasing: systematic errors.
- Total Error Estimation:
  - Sum and Difference: $q_o = x_o \pm y_o$ $\quad\quad \sigma_q = \sigma_x + \sigma_y$
  - Products and Quotients: $q_o = x_o . y_o$ & $q_o = x_o / y_o$ $\quad\quad \sigma_q = |q_o| \left[\frac{\sigma_x}{x_o} + \frac{\sigma_y}{y_o} \right]$

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

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

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

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

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

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

Least Square Fitting

a). Maximum Likelihood Estimator (MLE)

Maximum Likelihood Method ...cntd

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)!*

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: $\frac{\partial \chi^2}{\partial a}$ $\frac{\partial \chi^2}{\partial b}$

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

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

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

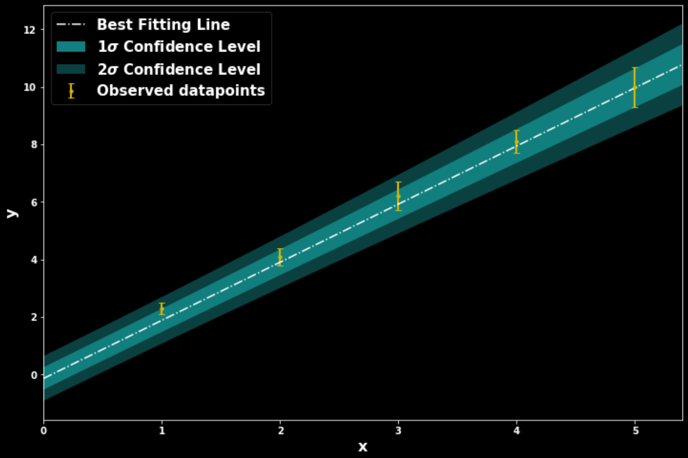
Observational dataset:

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

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

Best Fit values: $a = 0.22 \pm 0.14, \quad b = 2.03 \pm 0.07$

Maximum Likelihood Method ...cntd



Revise:

- Observational dataset: x_i^{obs}, y_i^{obs}, σ_{y_i}
- Theoretical model: $y = f(x; a_0, a_1, \cdot\cdot, a_j)$
- Likelihood: Maximise or Chi-square: Minimise
- Solve:

$$\begin{bmatrix} N & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^j \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{j+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \cdots & \sum x_i^{j+1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \sum x_i^j & \sum x_i^{j+1} & \sum x_i^{j+2} & \cdots & \sum x_i^{j+n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_j \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum (x_i y_i) \\ \sum (x_i^2 y_i) \\ \vdots \\ \sum \left(x_i^j y_i\right) \end{bmatrix} \Rightarrow \text{Inverse} = ...?$$

Least Square Fitting

b). Minimum Chi-square Statistics

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

$$\chi^2 = \sum_i \frac{(\text{ observed } - \text{ expected })^2}{\text{expected}} = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

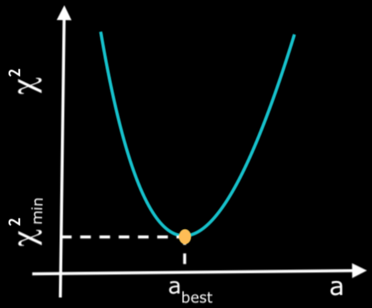
- **Observational dataset:** $x_i,\ y_{\rm obs}(x_i),\ \sigma_{y_i}$
- **Theoretical model:** $y_{\rm th}(x_i; a, b) = f(x_i; a, b)$
- **Define**

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

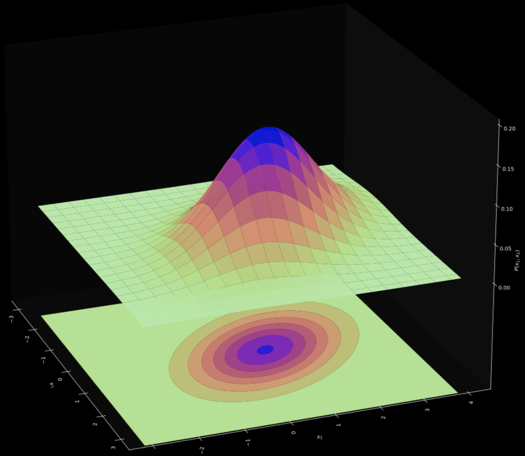
Chi-square:

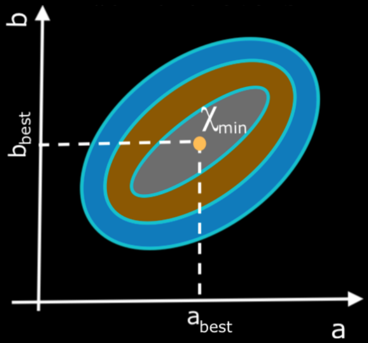
$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_{\rm obs}(x_i) - y_{\rm th}(x_i; a, b)}{\sigma_{y_i}} \right)^2 \Rightarrow \chi^2_{\rm min} \rightarrow \text{Best fit value of parameter}$$

Case:1 One Parameter Model:



Case:2 Two Parameters Model:





Question: *Why should we trust our model?*

→ *Fitting can be $\mathtt{overfit}$ or $\mathtt{underfit}$..*

→ *Model can be $\mathtt{linear}$, $\mathtt{quadratic}$ or so..*

- Chi-squared probability distribution:

$$P\left(\chi^2\right) \propto \left(\chi^2\right)^{\frac{\nu-2}{2}} \exp\left(-\chi^2/2\right); \qquad \nu : \text{d.o.f.=N-p}$$

- Mean: $\overline{\chi^2} = \nu$ Variance: $\mathrm{Var}\left(\chi^2\right) = 2\nu$

$$\Rightarrow \text{we expect} \quad \boxed{\chi^2 \sim \nu \pm \sqrt{2\nu}}$$

Reduced Chi-square:

$$\chi^2_\nu = \frac{\chi^2}{\nu}$$

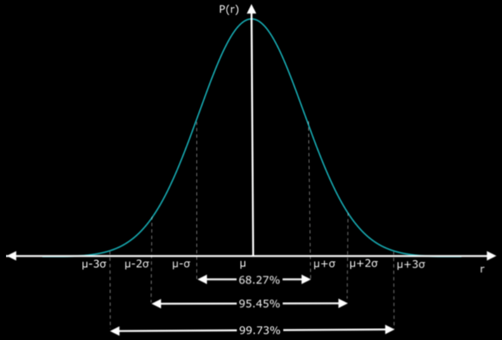
Goodness-of-fit:

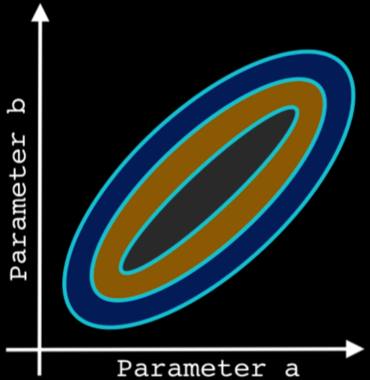
- $\chi^2_\nu \simeq 1 \rightarrow$ good match between data and model
- $\chi^2_\nu < 1 \rightarrow$ over-fitting of the data.
- $\chi^2_\nu > 1 \rightarrow$ poor model fit

Model Comparison: Linear model: M_L or Quadratic model: M_Q

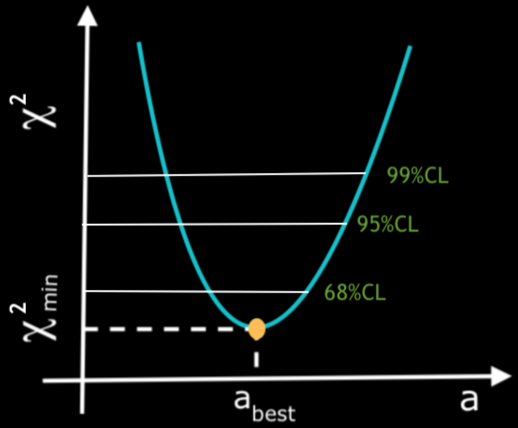
$\boxed{\textit{Best model is one whose value of } \chi^2_\nu \textit{ is closest to one!}}$

Confidence Intervals: Range of estimates for an unknown parameter.





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

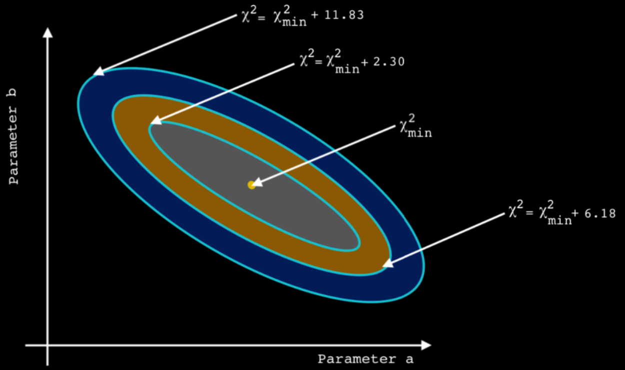


Dimensionality	1σ	2σ	3σ
1	1.00	4.00	9.00
2	2.30	6.18	11.83
3	3.53	8.02	14.16
4	4.72	9.72	16.25
5	5.89	11.31	18.21

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

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

Two-parameters model:



Parameter error estimation:

$$\sigma_{a_o} = \sqrt{\frac{\mathscr{B}}{\mathscr{C}^2 - \mathscr{A}\mathscr{B}}}, \quad \sigma_{b_o} = \sqrt{\frac{\mathscr{A}}{\mathscr{C}^2 - \mathscr{A}\mathscr{B}}}, \quad \sigma_{a_o b_o} = \sqrt{\frac{-\mathscr{C}}{\mathscr{C}^2 - \mathscr{A}\mathscr{B}}}$$

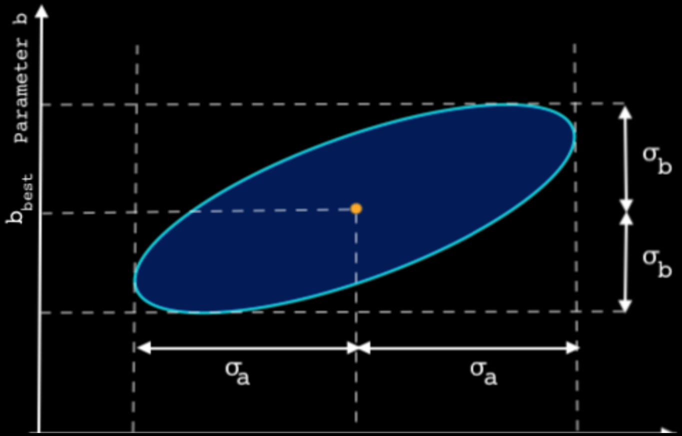
where $\mathscr{A} < 0, \quad \mathscr{B} < 0, \quad \mathscr{A}\mathscr{B} > \mathscr{C}^2$ and defined as

$$\mathscr{A} = \left.\frac{\partial^2 \mathcal{L}}{\partial a^2}\right|_{a_o, b_o}, \quad \mathscr{B} = \left.\frac{\partial^2 \mathcal{L}}{\partial b^2}\right|_{a_o, b_o}, \quad \mathscr{C} = \left.\frac{\partial^2 \mathcal{L}}{\partial a \partial b}\right|_{a_o, b_o}$$

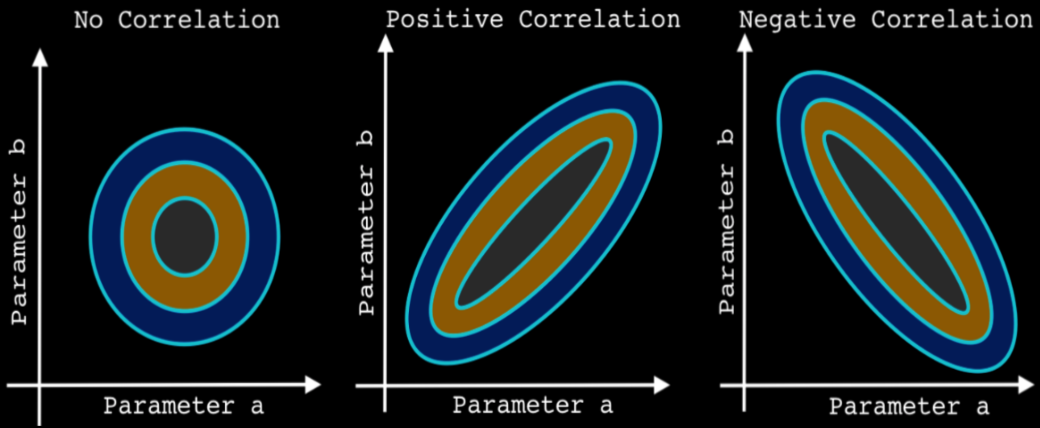
where a_o and b_o : best fit value of a and b parameters, and

$\mathcal{L}$: log-likelihood function.

Parameter error estimation:



Correlation among parameters:



Least Square Fitting

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: $a = - - - \pm - - -, \quad b = - - - \pm - - -$

Example-2: Hand-on Session

Astro-Observational dataset: Hubble parameter measurements of 30 datapoints

z	H(z)	$\sigma_H(z)$
0.07	69.0	19.6
..	..	..
..	..	..
..	..	..
..	..	..
1.965	186.5	50.4

Theoretical Model: $H^{\rm th}(z) = H_0\sqrt{\Omega_{m0}(1+z)^3 + 1 - \Omega_{m0}}$

Best Fit values: $H_0 = ---\pm---, \quad \Omega_{m0} = ---\pm---$

Key Points from this discussion

Revise:

- 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: $\chi^2 = \sum_{i=1}^{N} \left(\frac{y_{\rm obs}(x_i) - y_{\rm th}(x_i; a, b)}{\sigma_{y_i}} \right)^2$
- $\chi^2_{\rm min} \Rightarrow$ *Best fit value of parameters*
- Draw Confidence Level: $\chi^2_{n\sigma} = \chi^2_{\rm min} + \Delta\chi^2_{n\sigma}$
- Estimate error in parameters: $a = a_{\rm best} \pm \sigma_a$ and $b = b_{\rm best} \pm \sigma_b$

Limitation of Maximum Likelihood Estimation (MLE)

- Limited to simple models: This is generally well-suited for simple models with a small number of parameters.
- Computational time: This methods can be computationally intensive for large datasets or complex models.
- Large number of parameters: Problem arises with the large number of parameters.

Statistical Inference

Statistical Inference: A process to learn about what we do not observe (parameters) using what we know (data).

Two philosophical approaches to statistics

- Frequentist Approach:
  - Focus on the long-run frequency of events.
  - Uncertainty is viewed as inherent randomness in the data.
  - Underlying parameters remain constant during this process.
  - Here parameters are fixed.

Two philosophical approaches to statistics

- Frequentist Approach:
  - Focus on the long-run frequency of events.
  - Uncertainty is viewed as inherent randomness in the data.
  - Underlying parameters remain constant during this process.
  - Here parameters are fixed.
- Bayesian Approach:
  - Focus on the uncertainty of events and update beliefs based on new data.
  - Uncertainty is viewed as subjective and conditional on prior beliefs or knowledge.
  - Parameters are unknown and described probabilistically.
  - Data are fixed.

Bayesian Statistics

Some Elementary Concepts

Probability: Random process is known but we have to determine the outcome.
Measure of the chance of an event occurring is called *probability* $\rightarrow \mathbb{P}(A)$

Conditional Probability: Probability of an event occurring given that another event has already occurred $\rightarrow \mathbb{P}(A|B)$

Joint Probability: Probability of two or more events occurring together. $\rightarrow \mathbb{P}(A, B)$ or $\mathbb{P}(A \cap B)$

Some Elementary Concepts

Properties of Probability:

- The "sum" rule: $\mathbb{P}(A) + \mathbb{P}(\bar{A}) = 1$
- The "product" rule: $\mathbb{P}(A, B) \equiv \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$
- The "or" rule: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- The "marginalisation" rule:
  - $\mathbb{P}(A) = \sum_i \mathbb{P}(A, B_i) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)$ $\Rightarrow$ *for discrete prob. distributions.*
  - $\mathbb{P}(A) = \int_i \mathbb{P}(A, B_i) dB = \int_i \mathbb{P}(A|B_i)\mathbb{P}(B_i) dB$ $\Rightarrow$ *for continuous prob. distributions.*

Some Elementary Concepts

Random Variables: A mathematical function that maps the outcomes of a random event to a numerical value.

- Discrete random variable: *one that can only take on a finite or countably infinite number of values.* Example: the number of heads in three coin flips.
  - Binomial distribution.
  - Poisson distribution.
  - Discrete uniform distribution.

- Continuous random variable: *one that can take on any value in a specified range.* Example: the height of a person
  - Exponential distribution
  - Normal or Gaussian distribution
  - Uniform distribution

Some Elementary Concepts

Random Numbers: sequence of numbers that are generated in a way that appears unpredictable and without a discernible pattern.

- Pseudo-random numbers: generated using a deterministic algorithm. They appear random, but are not truly random.
- True random numbers: generated from physical phenomena such as atmospheric noise, thermal noise, or radioactive decay.
- Uniform random numbers: generated from a uniform distribution.
- Gaussian random numbers: generated from a Gaussian distribution.

Bayesian Statistics

Bayes Theorem: Conditional probability for two quantity.

The theorem is expressed mathematically as:

$$\mathbb{P}(A, B) \equiv \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

$$\Rightarrow \mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

OR

$$\boxed{\mathbb{P}(A|B, C) = \frac{\mathbb{P}(B|A, C)\mathbb{P}(A|C)}{\mathbb{P}(B|C)}}$$

...called Bayes Theorem

Bayesian Statistics

Bayes Theorem:

$$\mathbb{P}(A|B,C) = \frac{\mathbb{P}(B|A,C)\mathbb{P}(A|C)}{\mathbb{P}(B|C)}$$

We set:

- $A \leftarrow \Theta$: the parameters of a physical model,
- $B \leftarrow \mathcal{D}$: the experimental data,
- $C \leftarrow \mathcal{M}$: the physical model that includes all assumptions made in an analysis,

$$\therefore \quad \mathbb{P}(\Theta|\mathcal{D},\mathcal{M}) = \frac{\mathbb{P}(\mathcal{D}|\Theta,\mathcal{M})\mathbb{P}(\Theta|\mathcal{M})}{\mathbb{P}(\mathcal{D}|\mathcal{M})}$$

Bayesian Statistics

Bayes Theorem: $$\mathbb{P}(\Theta|\mathcal{D},\mathcal{M}) = \frac{\mathbb{P}(\mathcal{D}|\Theta,\mathcal{M})\mathbb{P}(\Theta|\mathcal{M})}{\mathbb{P}(\mathcal{D}|\mathcal{M})}$$

- Posterior probability distribution: $\mathbb{P}(\Theta|\mathcal{D},\mathcal{M})$
- Prior probability distribution: $\mathbb{P}(\Theta|\mathcal{M})$
- Likelihood function: $\mathbb{P}(\mathcal{D}|\Theta,\mathcal{M})$
- Model evidence: $\mathbb{P}(\mathcal{D}|\mathcal{M}) = \int \mathbb{P}(\mathcal{D}|\Theta,\mathcal{M})\mathbb{P}(\Theta|\mathcal{M})d\Theta$

Schematically,

$$\boxed{\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}}.$$

... updating our degree of belief when new information, in the form of data, becomes available

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: $\chi^2_{n\sigma} = \chi^2_{\mathsf{min}} + \Delta\chi^2_{n\sigma}$

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