Bayesian analysis

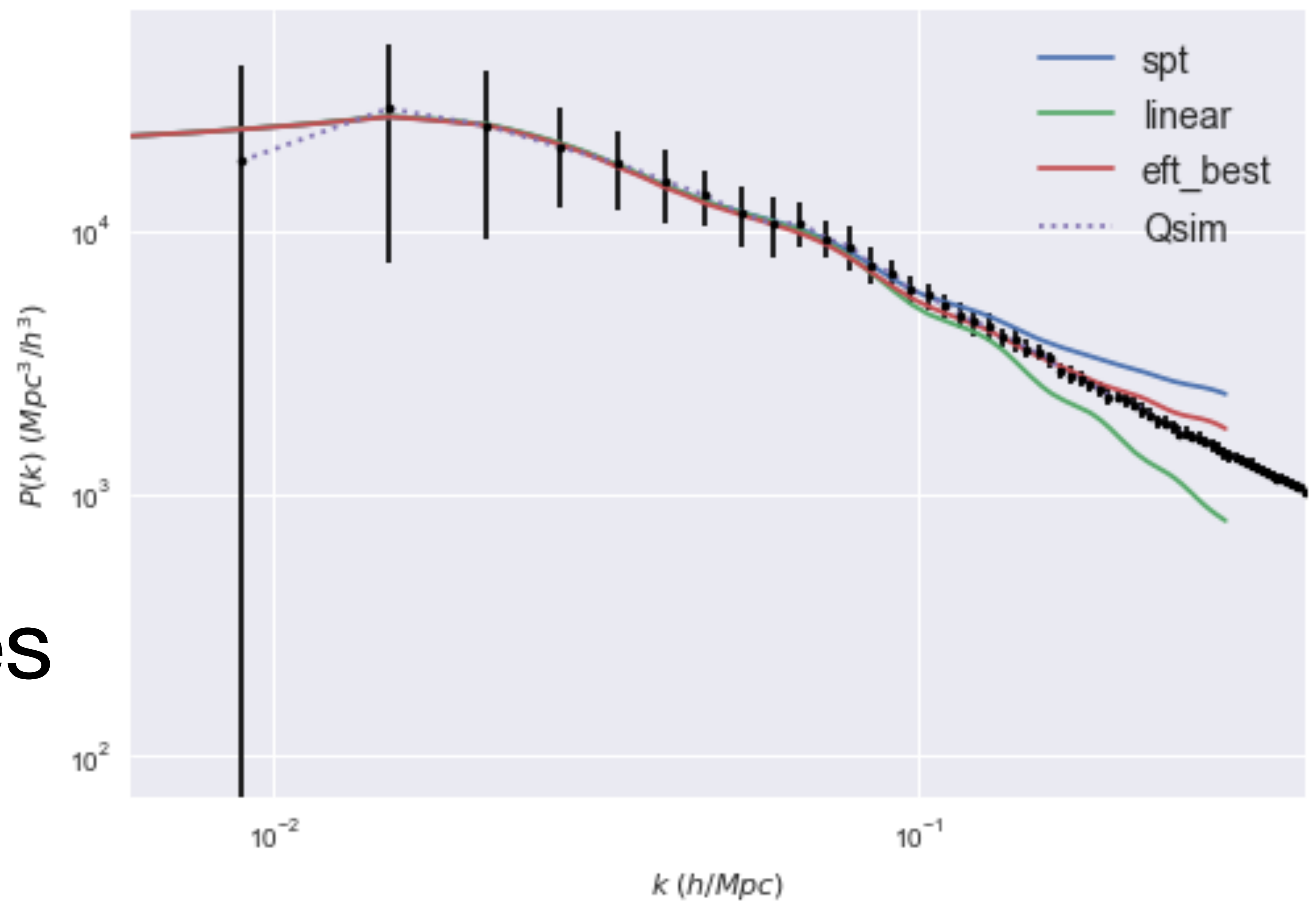
Parameter estimation and model selection

Lucia F. De la Bella

1. Why is this important?

Theory versus data

- How to compare/falsify my theory with reality?
- Statistics is the mathematical tool
    - To compare my theory with data
    - To separate good theories from bad theories
    - Note: uncertainties are crucial!



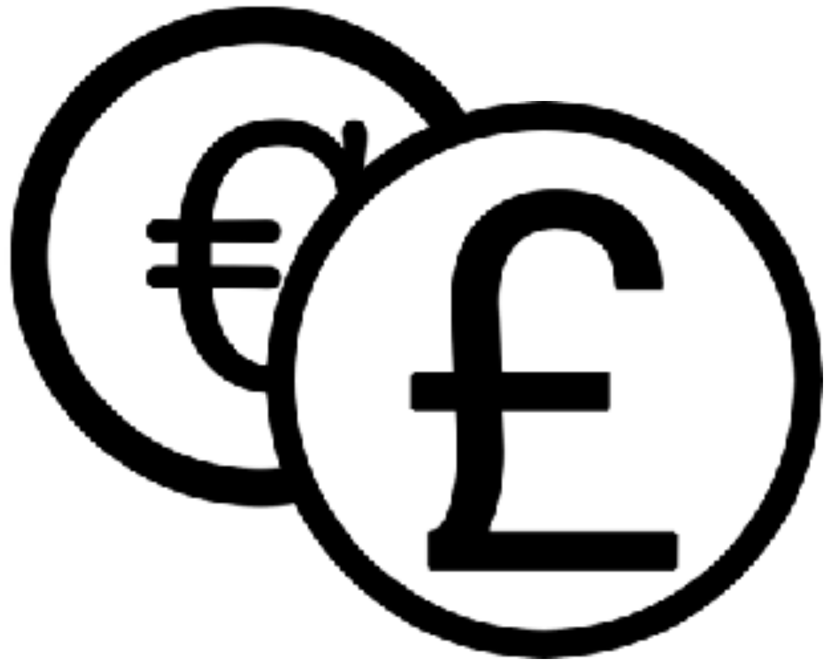
- Important: cosmology is an interesting case where there is only one experiment! We cannot reproduce lots of universes in our lab!!!

2. Probability

2.1. Bayesian definition

- *"An event's probability is a measure of an individual's* ***degree of belief*** *in assessing the uncertainty"*

- There are other definitions, e.g. frequentists. JACS session 03/12/2020.



- Given the data, what is the probability of the coin being biased?

Model: the coin is biased towards head

$b_h \in [0,1]$

$b_h = 0$ the coin always shows tails

$b_h = 1$ the coin always shows head

Data: HHH

How confident are you that the coin is fair?

$$P(model \mid data) = P(M \mid d) = P(b_h \mid d)$$

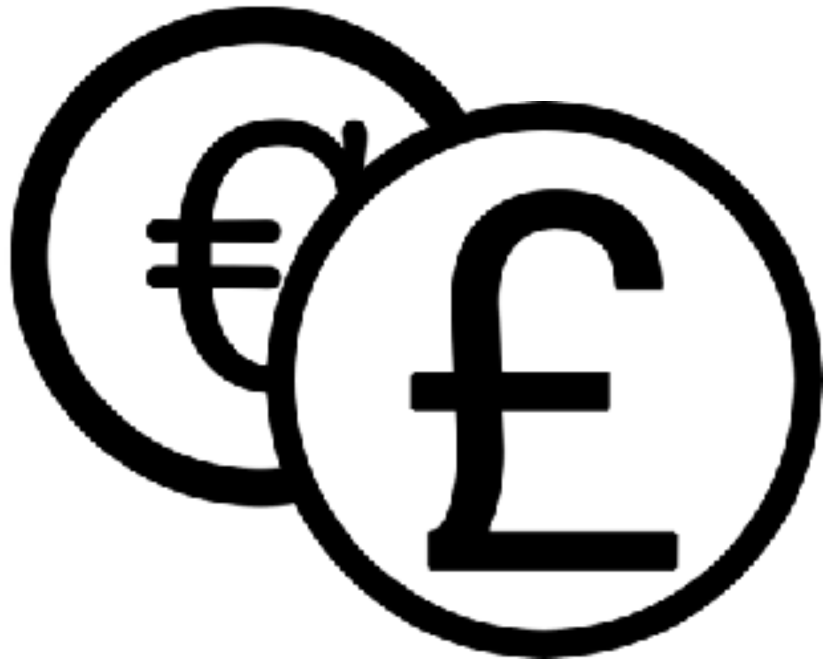
2. Probability

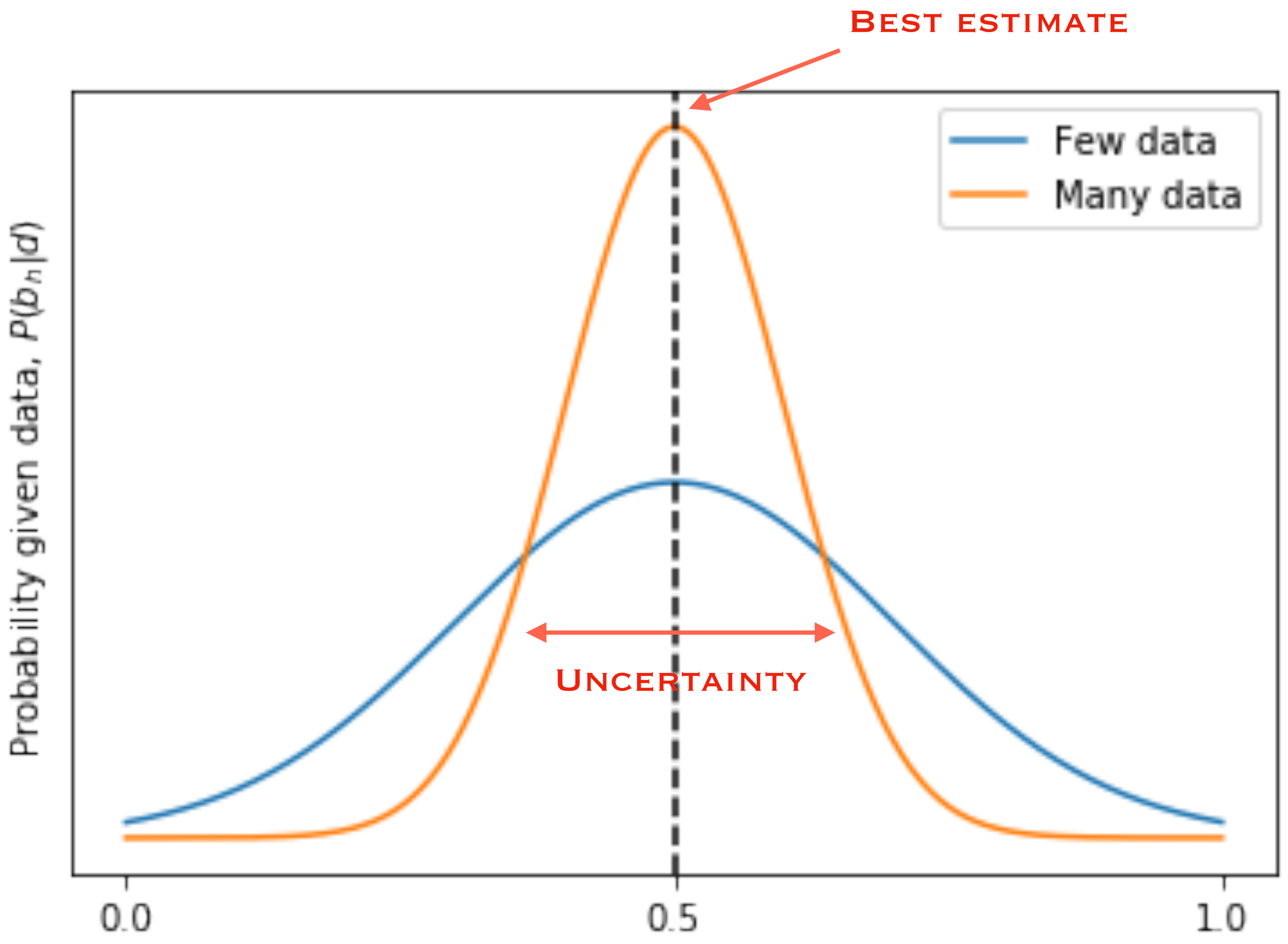
2.2. How to assess probability?

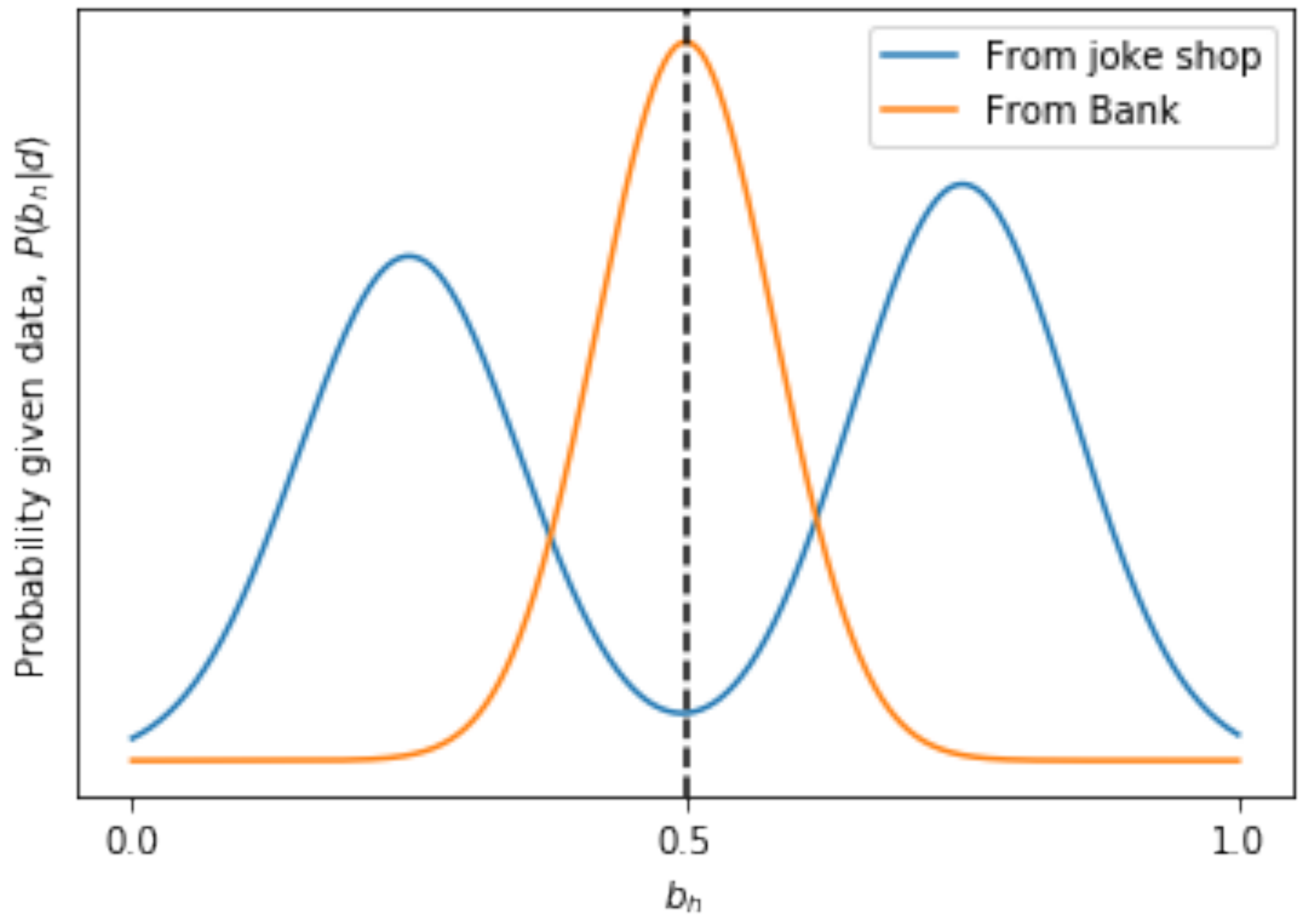
$$P(b_h | d)$$

- Given the data, what is the probability of the coin being biased?

Model: $b_h \in [0,1]$







Data: HHH
How confident are you that the coin is fair?

More data can change our degree of belief

But so other pieces of information: origin of coin

2. Probability

2.3. Bayes' Theorem

- Bayes' Theorem tell us how to change our degree of belief in a model based on new data.

$$P(M|d) = \frac{P(d|M)P(M)}{P(d)}$$

- $P(M|d)$ Posterior probability of the model
- $P(d|M)$ Data likelihood
- $P(M)$ Prior probability of the model
- $P(d)$ evidence

- In cosmology we use Bayes' theorem to infer the parameters $\vec{\theta} = \{\theta_1, \ldots\}$of our model.

$$P(\vec{\theta}|d, M) \propto P(d|\vec{\theta}, M)P(\vec{\theta}|M)$$

- $P(\vec{\theta}|d, M)$ Posterior probability distribution for values of parameters within our model
- $P(d|\vec{\theta}, M)$ Data likelihood
- $P(\vec{\theta}|M)$ Prior probability distribution for parameter values
- For now, forget about the evidence $P(d)$

2. Probability

2.4. Choosing a likelihood

$$P(\vec{\theta}\,|\,d, M) \propto P(d\,|\,\vec{\theta}, M)P(\vec{\theta}\,|\,M)$$

- Likelihood, $P(d\,|\,\vec{\theta}, M)$, is the probability to observe data points d_i with these values
- It tells you how the errors are distributed
- Gaussian distribution when you do not know any better

$$P(d\,|\,\vec{\theta}, M) = e^{-\chi^2/2}$$

$$\chi^2 = \sum_{ij} (f(\vec{\theta})_i - d_i) C_{ij}^{-1} (f(\vec{\theta})_j - d_j)$$

- $f(\vec{\theta})_i$ predicted values of data points in a model with particular parameters
- d_i vector of observed data points
- C_{ij}^{-1} Inverse covariance matrix, correlations between data points ($C_{ii} = \sigma^2$ variance)

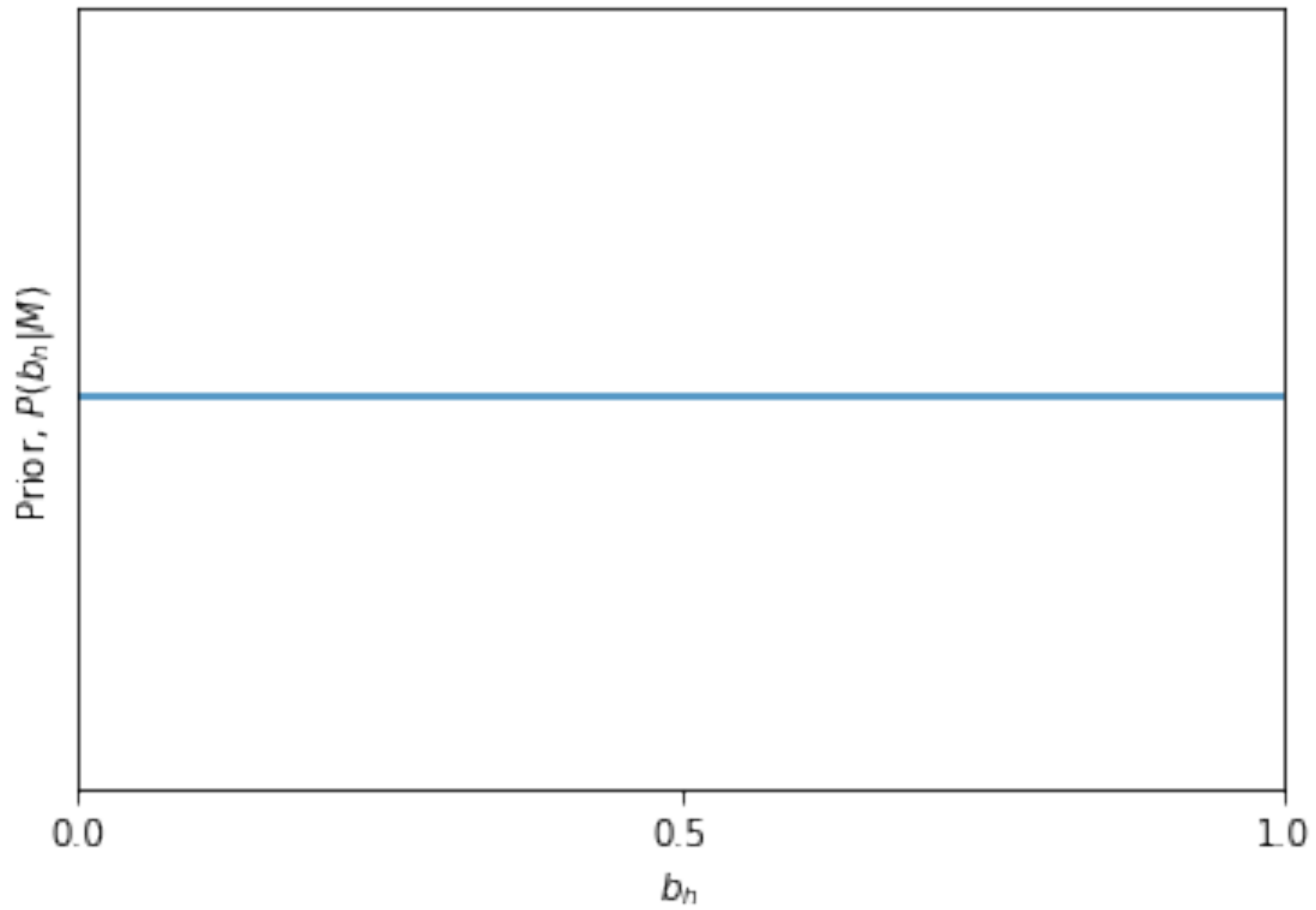
- Give a score to each potential set of parameters $\vec{\theta}$: the higher the score (likelihood), the closer to real data

2. Probability

2.5. The prior

$$P(\vec{\theta}\,|\,d, M) \propto P(d\,|\,\vec{\theta}, M)P(\vec{\theta}\,|\,M)$$

- Prior, $P(\vec{\theta}\,|\,M)$, is the initial information about your model
- It is similar to setting your initial conditions for a set of differential equations.
- It is subjective and free for you to choose.
- As data improves, the posterior converges to the same distribution regardless the prior.
- Uniform/flat priors, when you know nothing a priori



3. Contourplots

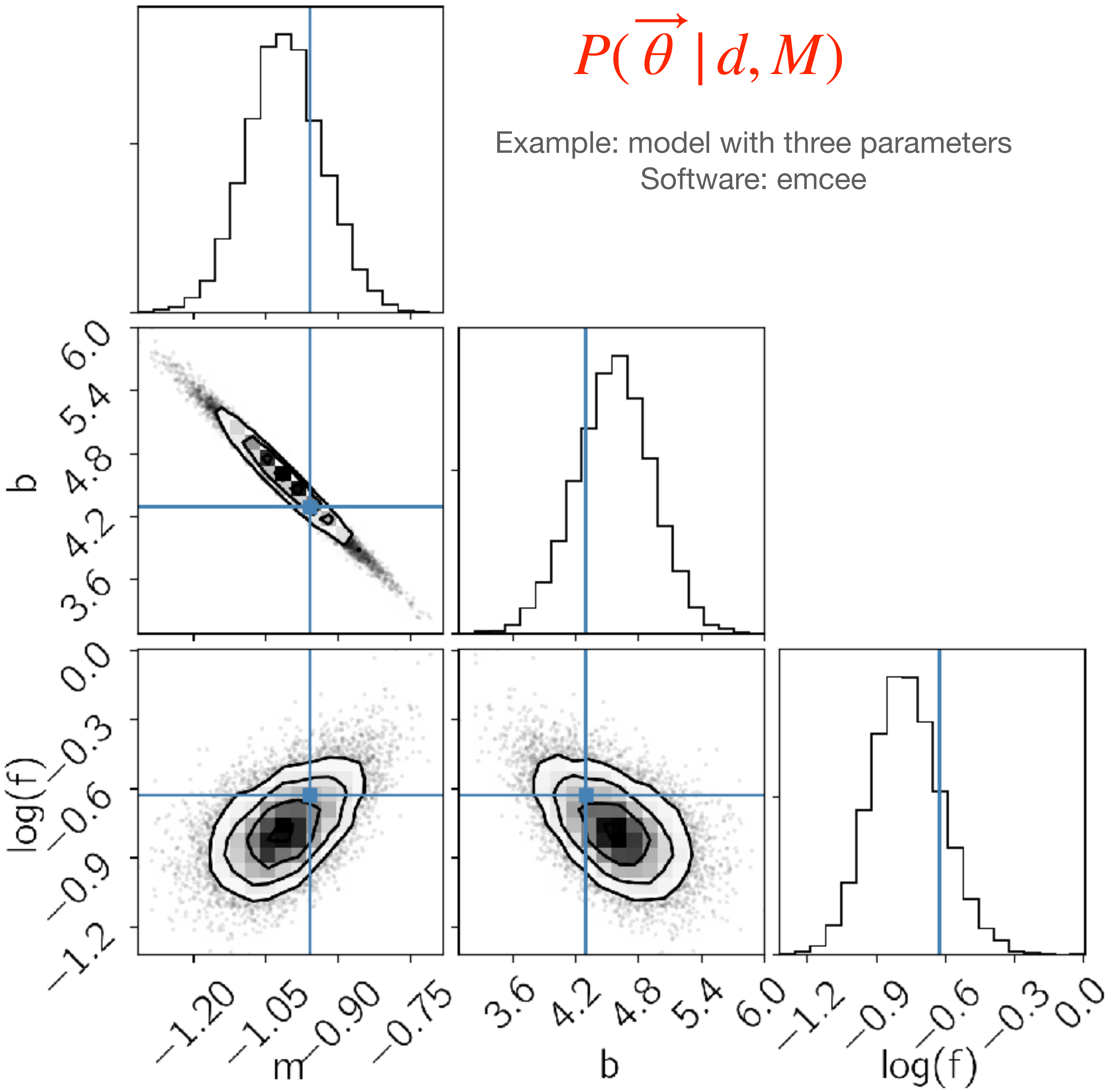
3.1. Parameters are correlated

- Corner plots show the one and two dimensional projections of the posterior probability distribution of the parameters.
- It is useful because it shows all the covariances/correlations between parameters.
- Levels of confidence: 66%, 95%, 99%.
- Which are the best parameters?
- To get uncertainties in one single parameter, we *marginalise* over the others

$$P(\theta_1 | M) = \int d\theta_2 P(\theta_1, \theta_2 | M)$$

$P(\vec{\theta} | d, M)$

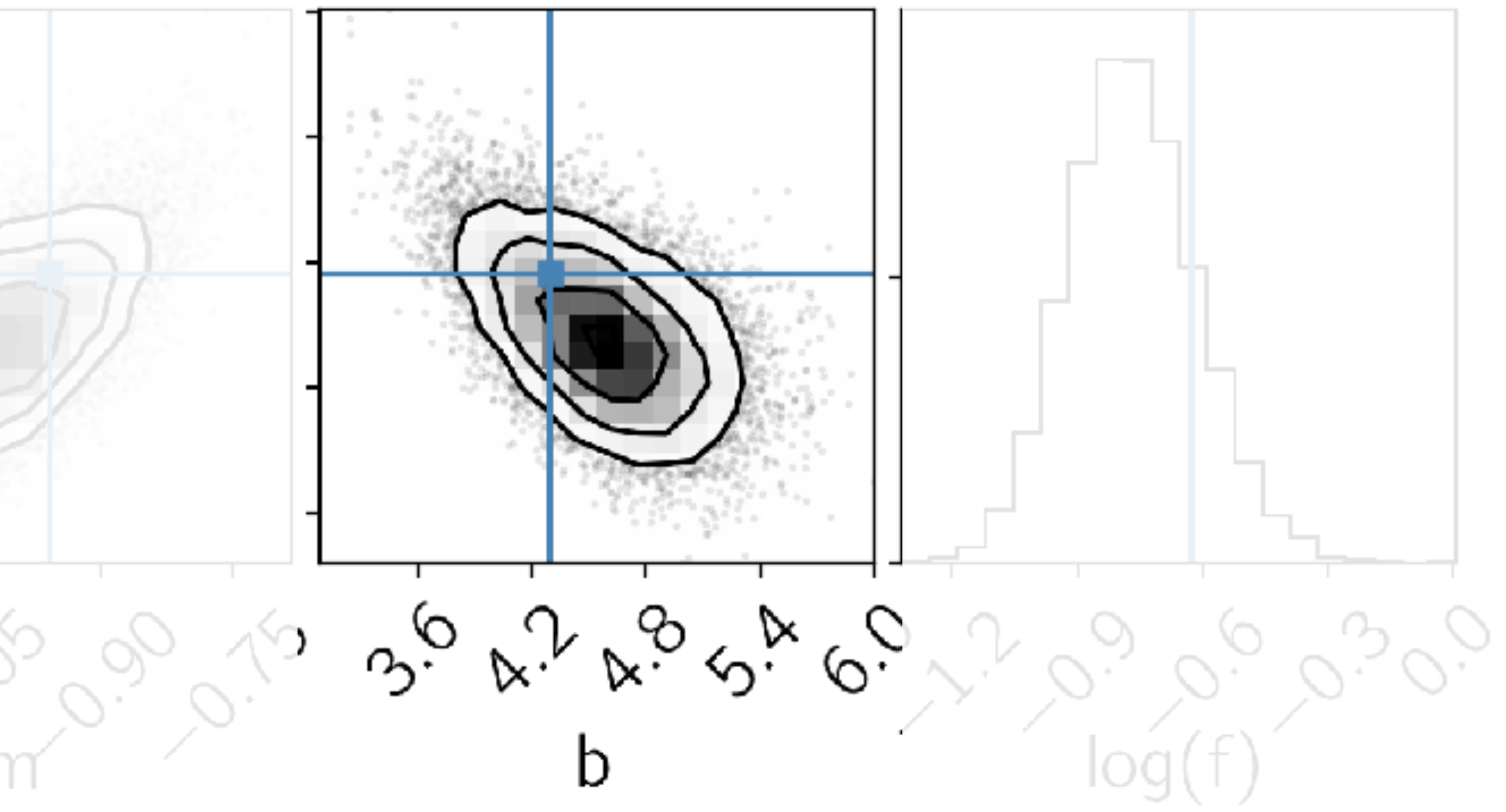
Example: model with three parameters
Software: emcee



3. Contourplots

3.2. Markov Chain Monte Carlo methods

- How do we get these contour plots of posterior probability?
- Use likelihood to randomly walk around the values of parameters $\vec{\theta}$.
- Many sampling algorithms (e.g. MCMC) design to do this efficiently.
- They use random walks algorithms.
- After a number of steps, we make a histogram.
- These are the probability contours!



4. Model comparison

4.1. The evidence

$$P(\vec{\theta}\,|\,d, M) = \frac{P(d\,|\,\vec{\theta}, M)P(\vec{\theta}\,|\,M)}{P(d)}$$

- As well as estimating parameters within one model
- We can compare different models.
- The evidence or marginal posterior for two different models

$$P(d, M_0) = \int d\vec{\theta}\,P(d\,|\,\vec{\theta}, M_0)P(\vec{\theta}\,|\,M_0)$$

$$P(d, M_1) = \int d\vec{\theta}\,P(d\,|\,\vec{\theta}, M_1)P(\vec{\theta}\,|\,M_1)$$

Which model is best?
Press-Schechter or Sheth-Tormen
Data: MICE sims

4. Model comparison

4.2. The Bayes factor

- The log of the ratio $$B_{01} = \frac{P(d, M_1)}{P(d, M_0)}$$

- Jeffrey's scale to assess the strength of evidence for model M_1.

|lnB|	relative odds	favoured model's probability	Interpretation
< 1.0	< 3:1	< 0.750	not worth mentioning
< 2.5	< 12:1	0.923	weak
< 5.0	< 150:1	0.993	moderate
> 5.0	> 150:1	> 0.993	strong

4. Task

Learn to use emcee

- Go to emcee documentation page https://emcee.readthedocs.io/en/stable/
- Install emcee following the installation guidelines
- Read the tutorials/Quickstart and reproduce their example in “Fitting a model to data”
- Together: Explain this example in our next meeting 27/11/2020 at 2.30 pm.
- 5-min talks:
    - emcee
    - MCMC methods

Good luck!!! We are almost there!