A brief introduction to Bayesian Statistics through Astronomical Applications (Lecture 1)

Cecilia Mateu J.
Centro de Investigaciones de Astronomía, Mérida, Venezuela

Universidad de Barcelona
24 de octubre de 2016

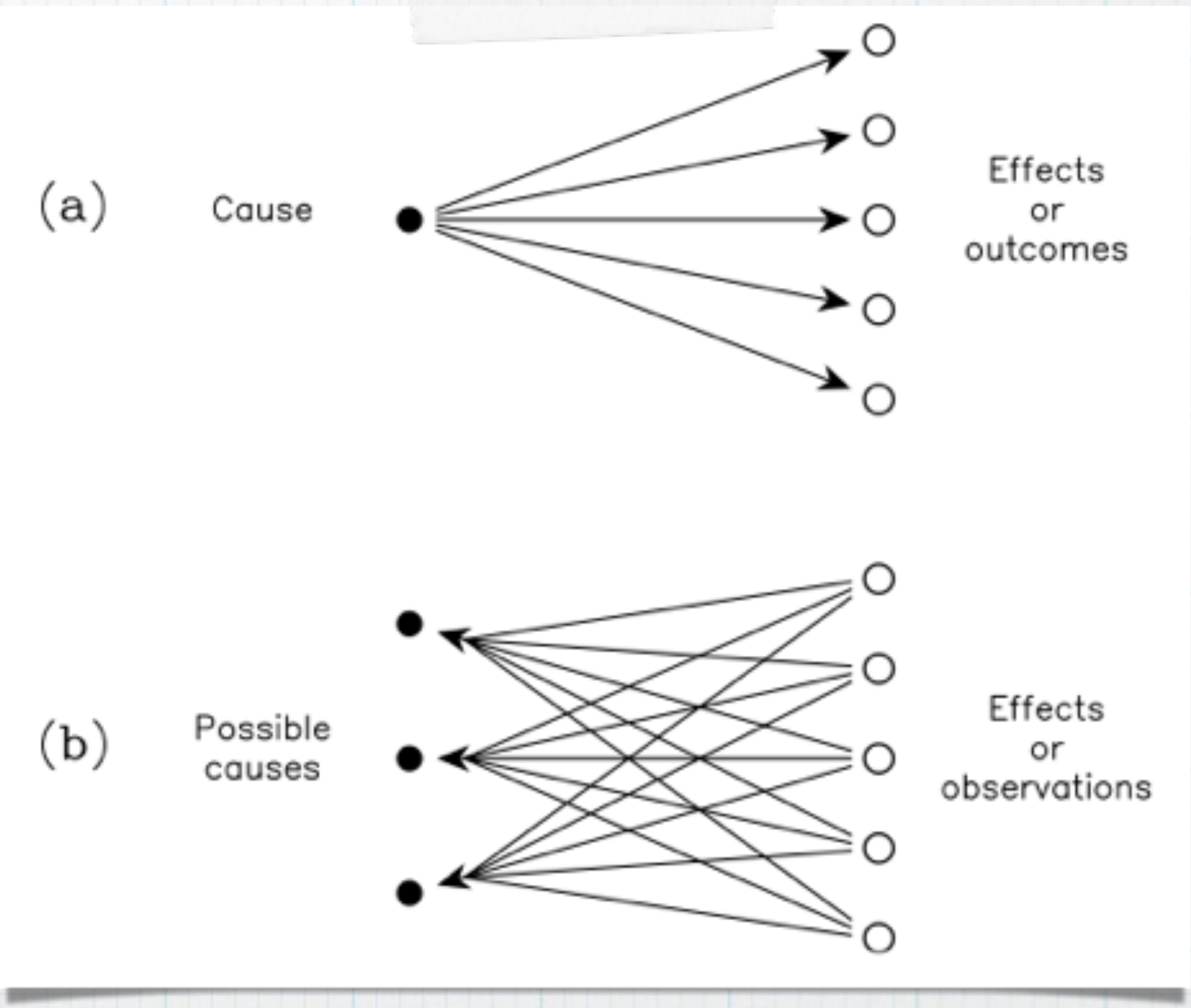
Inference and Deduction

* The forward problem:
  Given a cause predicting the possible effects

  → Deduction

* The inverse problem:
  Given a set of effects or observations, inferring the probable causes

  → Inference



(Sivia & Skilling 1996)

Hypothesis Testing

* Popper's Falsification Scheme
  * Hypotheses can only be falsified.
  * That's not enough, we'd like to have a way of ranking hypotheses according to their degree of success in reproducing observables
  * '..since it is impossible to demonstrate with certainty that a theory is true, it becomes impossible to decide among the infinite number of hypotheses which have not been falsified' (D'Agostini, 1998)

Some Motivations for Bayesian Inference

* Bayesian Statistics provides a clear framework for Inference - Hypothesis testing
* Probability is related to the state of uncertainty in a physical variable/model/theory, not only on the outcome of repeated experiments
* Our prior knowledge, assumptions, prejudices or lack thereof, must be stated explicitly in our model
* Propagation of uncertainties follows naturally

The Definition of Probability

"Probability is what everybody knows before going to school and continues to use afterwards, in spite of what one has been taught"

–G. D'Agostini (1998)

The Definition (and interpretation) of Probability

For an event or proposition A, probability is defined as:

* The Frequentist definition:

  P(A) is the relative frequency of occurrence of A in a series of Bernoulli trials, as the number of trials tends to infinity

* The Bayesian definition:

  P(A|I) is the plausibility (or our degree of belief) that A will occur, given I

  I denotes our assumptions (all available info) which in Bayesian statistics **must** be explicit. No such thing as absolute probabilities, all probabilities are conditional.

The Definition (and interpretation) of Probability

* The Frequentist definition:

  P(A) is the relative frequency of occurrence of A in a series of Bernoulli trials, as the number of trials tends to infinity

* The Bayesian definition:

  P(A|I) is the plausibility (or our degree of belief) that A will occur, given our assumptions I (available info)

Probability of getting heads in a coin toss?

$P(H)=1/2$

only IF heads and tails are equally probable!

$P(H|\text{fair coin})=1/2$

The Definition of Probability

* The 'frequentist' or classical definition can be recovered as an evaluation rule under appropriate conditions (usually when the indifference principle applies)

Probability of getting heads in a coin toss?

If heads and tails are equally probable

$P(H|\text{fair coin})=1/2$

Conventions

* A,B means A and B
* $A|B$ means A given B is True
* $P(A|B)$:= probability of A conditional on B = probability of A given B
* $P(A,B|C)$:= joint probability = probability of A and B, given C

Probability Rules

* Our definition of probability + Boolean logic implies that a probability must obey the following rules (see Jaynes 2003):
    * $0<P<1$
    * Sum Rule:

$$P(A|I) + P(\text{not-}A|I) = 1$$

    * Product Rule:

$$P(A,B|I)=P(A|B,I)\ P(B|I)$$

Bayesian Inference

* We have a set of data D, and a set of hypotheses H (possible causes)
* We would like to infer the probability for each hypothesis given that we have observed the data, and given all available information I at the time of the experiment

$$P(H,D|I) = P(D|H,I)\ P(H|I)$$

$$P(H,D|I) = P(H|D,I)\ P(D|I)$$

* we can thus write

$$P(H|D,I) = P(D|H,I)P(H|I)/P(D|I)$$

Bayes Theorem

* Bayes' Theorem:

$$P(H|D,I) = P(D|H,I)\, P(H|I) \,/\, P(D|I)$$

$P(H|D,I)$: Posterior probability

$P(D|H,I)$: Likelihood

$P(H|I)$: Prior probability

$P(D|I)$ = Normalization constant

Bayes' Theorem

* $P(H|I)$: Prior probability
    * The probability of the hypothesis, given our previous knowledge I, i.e. before any data is acquired: what is the plausibility of H?
* $P(D|H,I)$: Likelihood
    * The probability of having observed the data, given the hypothesis H
* $P(H|D,I)$: Posterior probability
    * The probability of the Hypothesis, given the data
* $P(D|I)$ = Normalization constant (called also Bayes factor)

Bayes Theorem

* Bayes' Theorem:

$$P(H|D,I) = P(D|H,I)\ P(H|I)\ /\ P(D|I)$$

translation:

$$\text{Posterior} = \frac{\text{Likelihood} * \text{Prior}}{\text{Constant}}$$

The Coin Example

The Coin Example

* Lets say we're at a casino and see a coin tossed N times, with the following outcome
    * H, T, H, H, H, T, H, t
* We'd like to know if the coin is biased

The Coin Example

* Let h be the coin bias, i.e. the probability of getting heads in a single coin toss

* The probability of having observed N_h heads in N tosses is

$$hhh \ldots\ldots h \qquad (N_h \text{ times})$$

* and the probability of getting $(N-N_h)$ tails is

$$(1-h)(1-h)(1-h) \ldots\ldots (1-h) \quad (N-N_h \text{ times})$$

The Coin Example

* So, we can write our likelihood function as

$$P(N, N_h | h, I) = h^{N_h}(1-h)^{N-N_h}$$

The Binomial Distribution

and the posterior is given by Bayes' Theorem as

$$P(h | N, N_h, I) = C h^{N_h}(1-h)^{N-N_h} P(h | I)$$

* where C is the normalization constant

The Coin Example

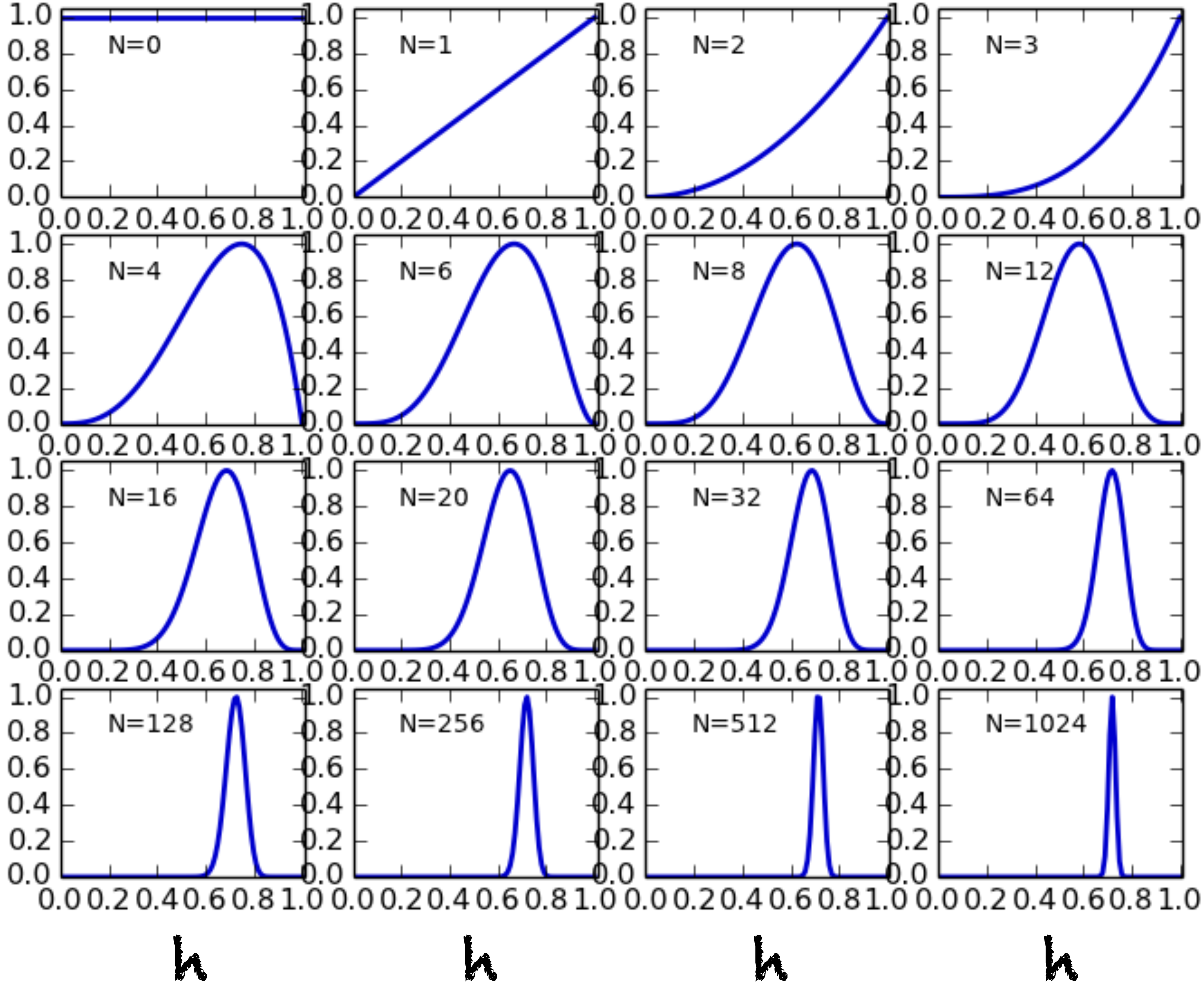
* Lets recap
  * N and N_h are our data (known)
  * Our goal is to get $P(h|N,N_h,I)$ remember this is a function of h

$\downarrow$

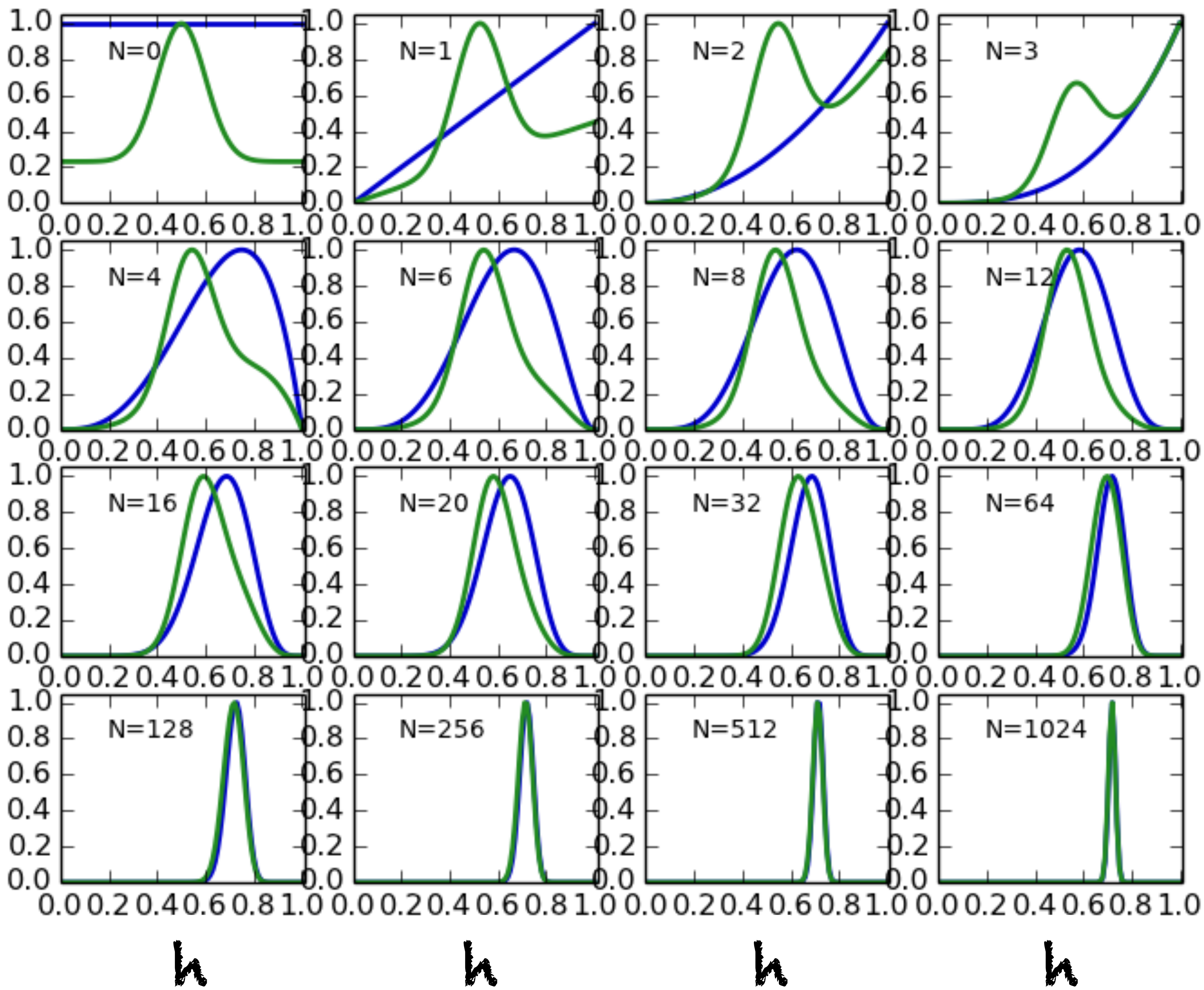
The full posterior IS the answer to our problem

$$P(h|N,N_h,I) = Ch^{N_h}(1-h)^{N-N_h}\,P(h|I)$$

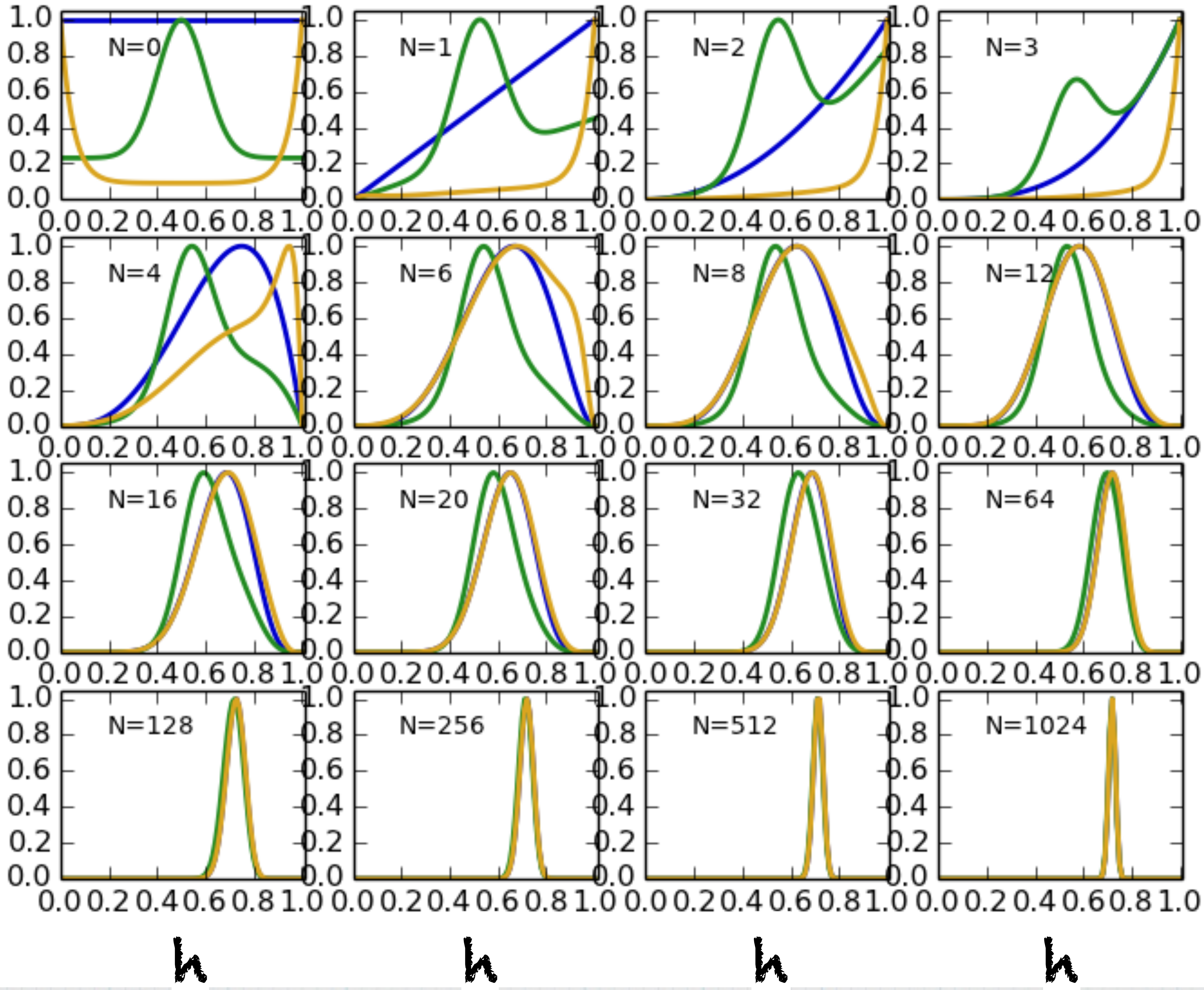
The Coin Example



The Coin Example



The Coin Example



The Coin Example

The full posterior IS the answer to our problem

$$P(h|N,N_h,I) = Ch^{N_h}(1-h)^{N-N_h}\, P(h|I)$$

* anything else we may want can be calculated from it, e.g.
  * the most probable value of h
  * credible regions (Bayesian term for confidence intervals)
  * The probability that h>0.5
    * int P(h|N,n,I) dp
  * ... more on this tomorrow ...

More examples

The full posterior IS the answer to our problem

$$P(h|N,N_H,I) \propto h^{N_H}(1-h)^{N-N_H}P(h|I)$$

* The probability that the coin is biased:
  * Lets say if $0.45<h<0.55$ we can safely take the coin as fair

$$P_{fair} = \int_{0.45}^{0.55} P(h|N,N_H,I)dh$$

* so, the probability that it is biased is $P_{biased} = 1 - P_{fair}$

$$P_{biased} = \int_{0}^{0.45} P(h|N,N_H,I)dh + \int_{0.55}^{1.} P(h|N,N_H,I)dh$$

The Importance (or not) of Priors

* The prior probability reflects our knowledge or ignorance on the problem
* In practice, for many applications the posterior is dominated by the likelihood
* If radically different priors are thought to be acceptable and the 'answer' depends strongly on the choice of the prior, it just means the data is not constraining enough! (see Jaynes 2003, D'Agostini 1998)

Github Repository

* Classes and programs are available in the following GitHub repository

https://github.com/cmateu/intro_to_bayes_UB

Updating Information

Updating Information

* Lets consider the case of having two independent data points D_1 and D_2. Bayes' Theorem states

$$P(H|D,I) \propto \prod_{i=1,2} P(D_i|H,I)\, P(H|I)$$

* Expanding the product in the likelihood term:

$$P(H|D,I) \propto P(D_2|H,I)\, \underbrace{\boxed{P(D_1|H,I)\, P(H|I)}}_{P(H|D_1,I)}$$

Updating Information

* Lets consider the case of having two independent data points D_1 and D_2. Bayes' Theorem states

$$P(H|D_1,D_2,I) \propto \prod_{i=1,2} P(D_i|H,I)\ P(H|I)$$

* Expanding the product in the likelihood term:

$$P(H|D_1,D_2,I) \propto P(D_2|H,I)\ \boxed{P(D_1|H,I)\ P(H|I)}$$

$$P(H|D_1,D_2,I) \propto P(D_2|H,I)\,P(H|D_1,I)$$

* Here $P(H|D_1)$ the posterior on H given D_1 is acting as an updated prior!

Least Squares

Gaussian Uncertainties: Least Squares derived

* In a problem, whatever it may be, where our model differs from our observations because of gaussian uncertainties, the posterior is simply:

$$P(model|data, I) = \prod_{i=1}^{N} e^{-\frac{(x_i - x_{model})^2}{2\sigma_i^2}} P(model|I)$$

$$= e^{-\frac{1}{2}\sum_{i=1}^{N} \frac{(x_i - x_{model})^2}{\cdots}}$$

* For a uniform prior we have

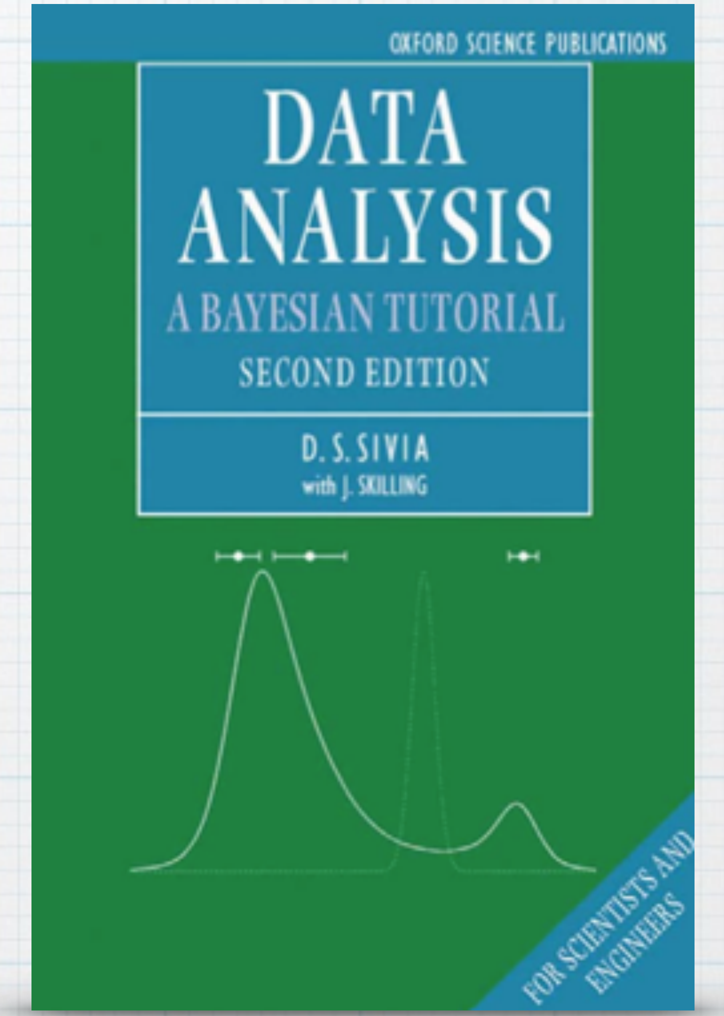
$$P(model|data, I) = e^{-\frac{1}{2}\chi^2}$$

The least squares or χ2 minimisation method derived !!!! (assumptions are explicit!)

Very short bibliography

* Highly recommended introductory bibliography:
    * Sivia & Skilling book
    * Giulio D'Agostini's notes available at Tom Loredo's BIPS web page:

http://www.astro.cornell.edu/staff/loredo/bayes/



Very short bibliography

* 'Statistics, Data Mining, and Machine Learning in Astronomy' by Ivezic, Connolly, VanderPlas & Gray
* Python code freely available at:
* http://www.astroml.org

Disponible en recursos en línea biblioteca UNAM