# Lecture 4 : Bayesian inference

#### Lecture 4 : Bayesian inference

- What is the Bayesian approach to statistics? How does it differ from the frequentist approach?
- Conditional probabilities, Bayes' theorem, prior probabilities
- Examples of applying Bayesian statistics
- Bayesian correlation testing and model selection
- Monte Carlo simulations

## What is conditional probability?

- The concept of conditional probability is central to understanding Bayesian statistics
- P(A|B) means "the probability of A on the condition that B has occurred"
- Adding conditions makes a huge difference to evaluating probabilities
- On a randomly-chosen day in CAS , P(free pizza)  $\sim 0.2$
- P(free pizza|Monday) ~ I, P(free pizza|Tuesday) ~ 0

## Lies, damn lies and statistics



#### Example 6

O.J.Simpson's defence attorney : "Only 0.1% of the men who abuse their wives end up murdering them. The fact that Simpson abused his wife is irrelevant to the case"

Why was this poor statistics? This ignores the fact that Simpson's wife was actually murdered. What is relevant is P(abusive husband is guilty | wife is murdered), not P(husband murders wife | husband abuses wife). Using reasonable data : P(guilty) ~ 0.8.

## What is a "frequentist approach" to statistics?

- Frequentist statistics assign probabilities to a measurement, i.e. they determine P(data|model)
- For example : What is the probability that chi-squared has a particular value, given the model?
- We are defining probability by imagining a series of hypothetical experiments repeatedly sampling the population, which have not actually taken place.
- Philosophy of science : we attempt to "rule out" or falsify models if P(data|model) is too small.

- Bayesian statistics assign probabilities to a model, i.e. they give us tools for calculating P(model|data)
- We will see that this cannot be done without assigning a prior probability to each model [see later]
- We update the model probabilities in the light of each new dataset (rather than imagining many hypothetical experiments)
- Philosophy of science : we do not "rule out" models, just determine their relative probabilities

• An important role is played by **Bayes' theorem**, which can be derived from elementary probability

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

 Small print : for example, this formula can be derived by just writing down the joint probability of both A and B in two ways :

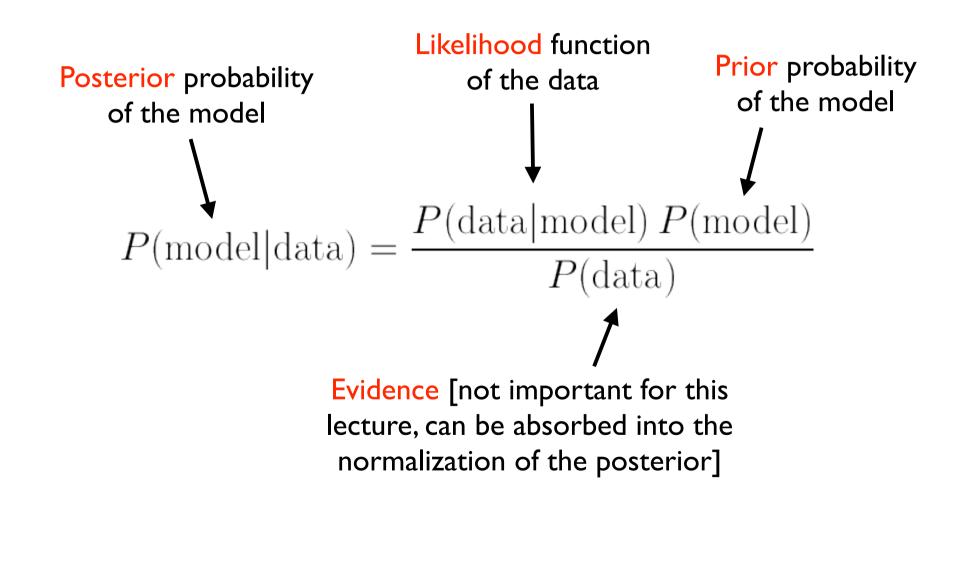
$$P(A \cap B) = P(A|B) P(B) = P(B|A) P(A)$$

• Remember  $P(A|B) \neq P(B|A)$  in general !!

- Example I : the probability of a certain medical test being positive is 90%, if a patient has disease D. 1% of the population have the disease, and the test records a false positive 5% of the time. If you receive a positive test, what is your probability of having D?
- P(+|D)=0.9, P(D)=0.01, P(+|no D)=0.05, we want P(D|+)
- $P(D|+) = \frac{P(+|D) P(D)}{P(+)} = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|\text{no}D)P(\text{no}D)}$
- Substituting in the numbers : P(D|+) = 0.15

- A frequentist might argue "either the person has the disease or not it is meaningless to apply probability in this way"
- A Bayesian might argue "there is a prior probability of 1% that the person has the disease. This probability should be updated in the light of the new data using Bayes' theorem"

• Bayes' theorem re-written for science :



- The importance of the prior probability is both the strong and weak point of Bayesian statistics
- A Bayesian might argue "the prior probability is a logical necessity when assessing the probability of a model. It should be stated, and if it is unknown you can just use an uninformative (wide) prior"
- A frequentist might argue "setting the prior is subjective - two experimenters could use the same data to come to two different conclusions just by taking different priors"

## The uniform prior

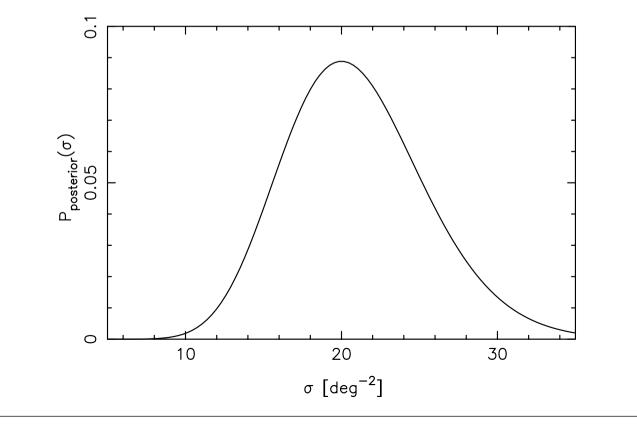
- In the absence of other information, a uniform (or constant) prior is often assumed. This is effectively equivalent to the fitting range.
- In Lecture 3 we implicitly assumed a uniform prior when determining the probability distribution of a model parameter "a" in chi-squared fitting :

$$P(a|\mathrm{data}) \propto P(\mathrm{data}|a) P(a) \propto \exp\left(-\chi^2/2\right)$$
  
We assume the prior is constant

## **Applications of Bayesian statistics**

• Example 2 : A I-degree survey finds 20 quasars. What is the posterior probability distribution for the quasar number density s?

•  $P(s|D) \sim P(D|s) P(s) \sim P_{Poisson}(n=20|s)$  [uniform prior]



## **Applications of Bayesian statistics**

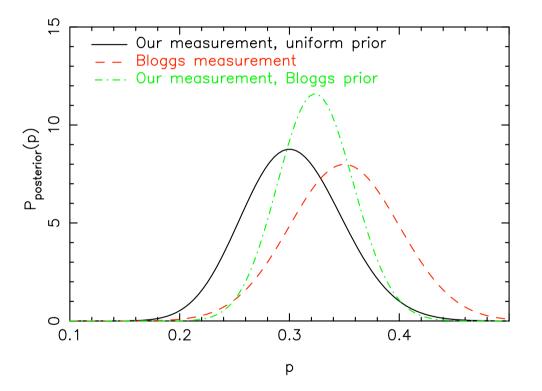
 Example 3 : I observe 100 galaxies, 30 of which are AGN. What is the posterior probability distribution of the AGN fraction p assuming (a) a uniform prior,
(b) Bloggs et al. have already measured that p has a Gaussian distribution with mean 0.35 and r.m.s. 0.05?

• 
$$P(p|D) \sim P(D|p) P(p) = P_{binomial}(n=30|N=100,p) P(p)$$

• Consider two cases for P(p) : (a) uniform (b) Gaussian

## Applications of Bayesian statistics

 Example 3 : What is the posterior probability distribution of the AGN fraction p ...



Bayesian statistics naturally allows for combination with previous measurements, via the prior

## Bayesian correlation testing

 In Lecture 2 we measured the correlation coefficient of two variables. To test the significance of the result we asked what is the probability of measuring this value of r if there is no correlation? Mathematically :

$$P(r|\rho=0)$$

 Using Bayesian statistics we can ask the opposite question : what is the posterior probability distribution for the correlation coefficient given the measured value of r? Mathematically :

$$P(\rho|r)$$

## Bayesian correlation testing

• Assume (x,y) data are drawn from bivariate Gaussian :

$$P(x,y) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

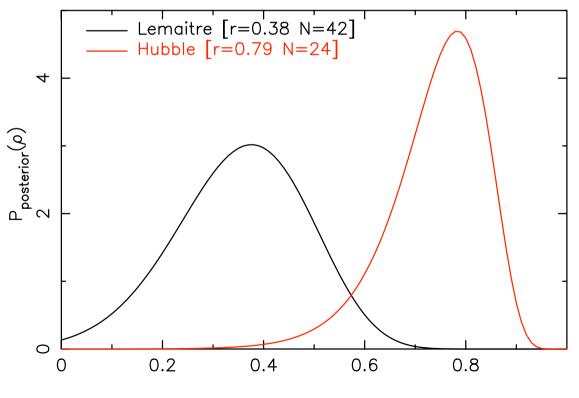
• We use Bayes' theorem to compute  $P(\rho|r)$  for this model, marginalizing over the other parameters :

$$P(\rho|r) \propto \frac{(1-\rho^2)^{\frac{N-1}{2}}}{(1-\rho r)^{N-\frac{3}{2}}} \left(1 + \frac{1}{N-\frac{1}{2}} \frac{1+r\rho}{8} + \dots\right)$$

• Then plug in values of N and r

## Bayesian correlation testing

 Example 4 : Use Bayesian correlation testing to determine the posterior probability distribution of the correlation coefficient of Lemaitre and Hubble's distance vs. velocity data, assuming a uniform prior.



ρ

## Bayes factor and model selection

Bayes theorem allows us to perform model selection.
Given models M<sub>1</sub> (parameter p<sub>1</sub>) and M<sub>2</sub> (parameter p<sub>2</sub>) and a dataset D we can determine Bayes factor :

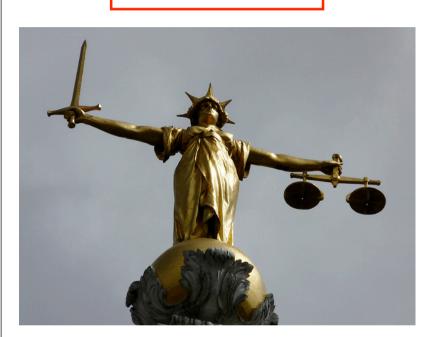
$$K = \frac{P(M_1|D)}{P(M_2|D)} = \frac{\int dp_1 P(D|p_1) P(p_1)}{\int dp_2 P(D|p_2) P(p_2)}$$

• The size of K quantifies how strongly we can prefer one model to another, e.g. the Jeffreys scale :

К	strength of evidence
I-3	"barely worth mentioning"
3-10	"substantial"
10-30	"strong"
>30	"very strong"

## Lies, damn lies and statistics

## Example 7



1999 : The 2 children of the U.K. Clark family both died within a few weeks of birth. Their mother was charged with double murder on the basis : cot death is a 1/8500 event. So the probability of both children dying naturally is "1 in 73 million"

Why was this poor statistics? (1) The events are not independent (genetic factors). (2) Just because an event is rare doesn't make it impossible. (3) Bayes' theorem is needed to assess the relative probabilities.

## Monte Carlo simulations

• A Monte Carlo simulation is a computer model of an experiment in which many random realizations of the results are created and analyzed like the real data





## Monte Carlo simulations

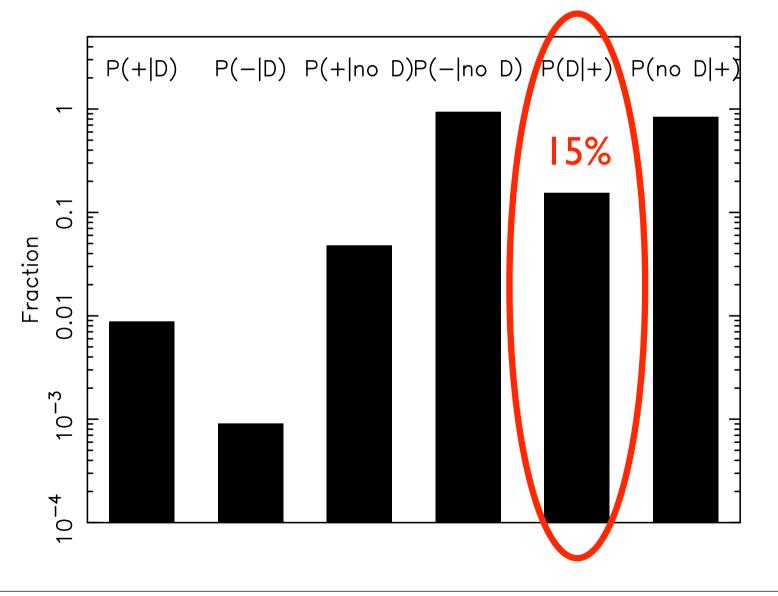
- A Monte Carlo simulation is a computer model of an experiment in which many random realizations of the results are created and analyzed like the real data
- This very powerful technique allows measurement of both statistical errors and systematic errors particularly in cases without an analytic solution
- Statistical errors can be obtained from the distribution of fitted parameters over the realizations
- Systematic errors can be explored by comparing the mean fitted parameters to their known input values

- Example 5 : simulate example 1 by Monte Carlo
- [The probability of a certain medical test being positive is 90%, if a patient has disease D. 1% of the population have the disease, and the test records a false positive 5% of the time. If you receive a positive test, what is your probability of having D?]
- We will compare the mathematical solution and the Monte Carlo solution
- [Bayes' theorem answer = 0.15]

- [The probability of a certain medical test being positive is 90%, if a patient has disease D. 1% of the population have the disease, and the test records a false positive 5% of the time. If you receive a positive test, what is your probability of having D?]
- Solution by maths : suppose 10,000 people are tested
- 100 have the disease, of which 90 return positive tests
- 9,900 do not have the disease, of which 495 return positive tests
- Probability of having disease, given positive test, is 90/ (90+495) = 0.15

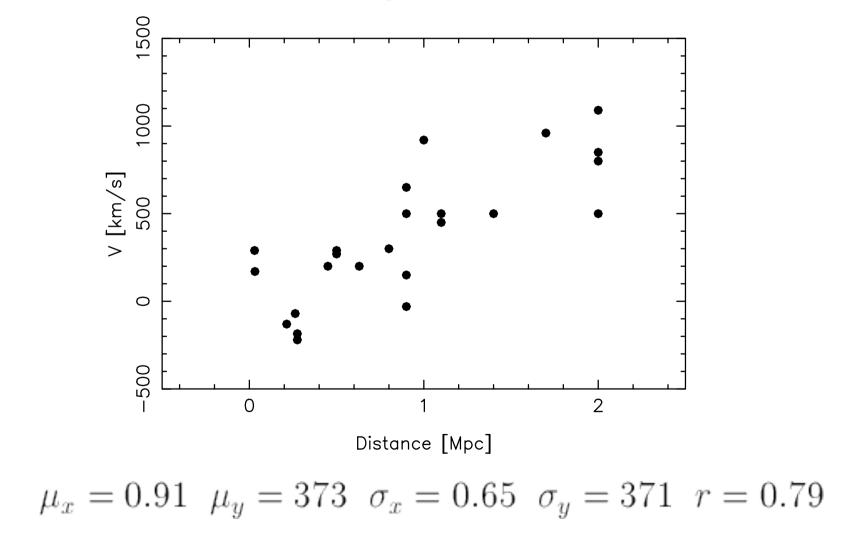
- [The probability of a certain medical test being positive is 90%, if a patient has disease D. 1% of the population have the disease, and the test records a false positive 5% of the time. If you receive a positive test, what is your probability of having D?]
- Solution by Monte Carlo : loop over N patients
- For each patient, assign the disease with probability 0.01
- For each patient with the disease, assign positive test result with probability 0.9
- For each patient without the disease, assign positive test result with probability 0.05



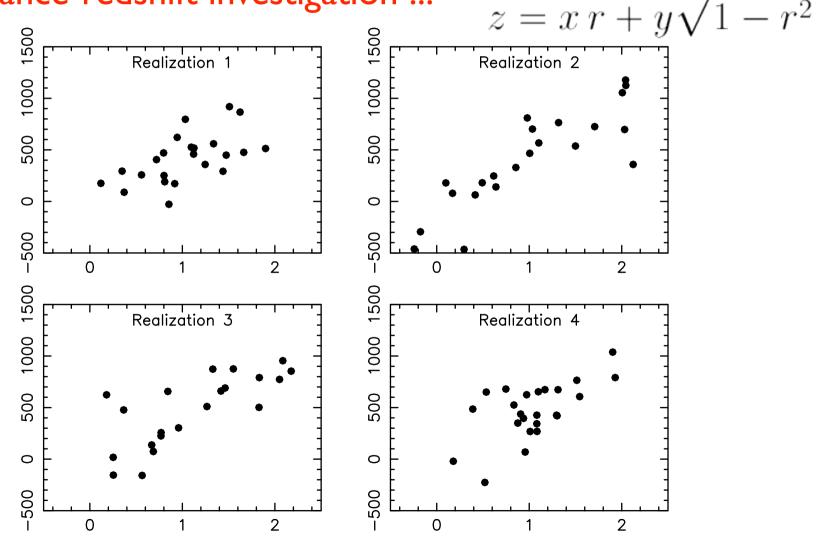


 Example 6 : Run a Monte Carlo simulation of Hubble's distance-redshift investigation, assuming that D and V are drawn from a bivariate Gaussian distribution. What is the resulting error in the Hubble parameter?

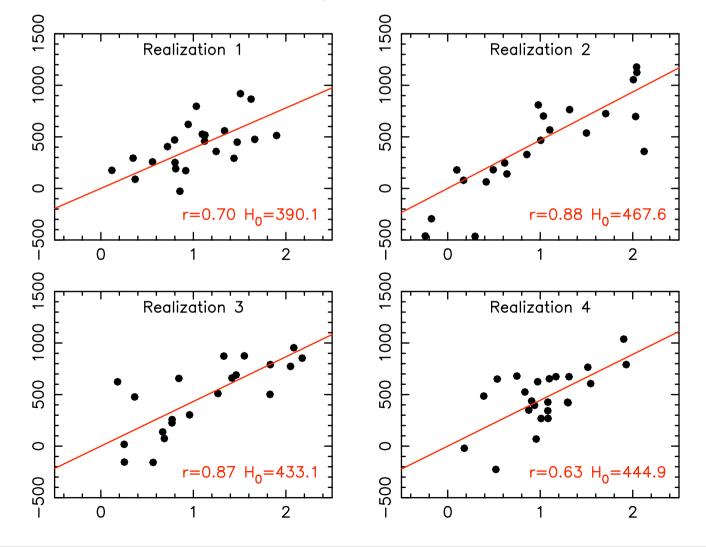
 Example 6 : Run a Monte Carlo simulation of Hubble's distance-redshift investigation ...



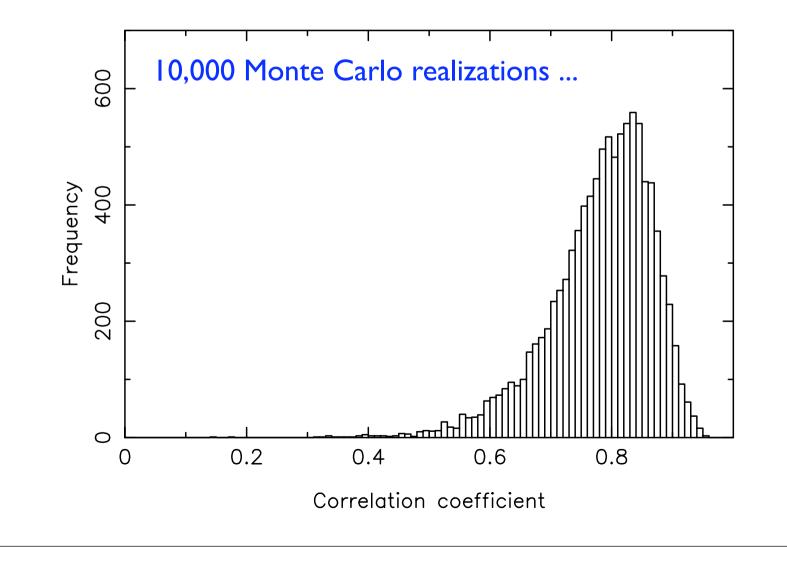
 Example 6 : Run a Monte Carlo simulation of Hubble's distance-redshift investigation ...



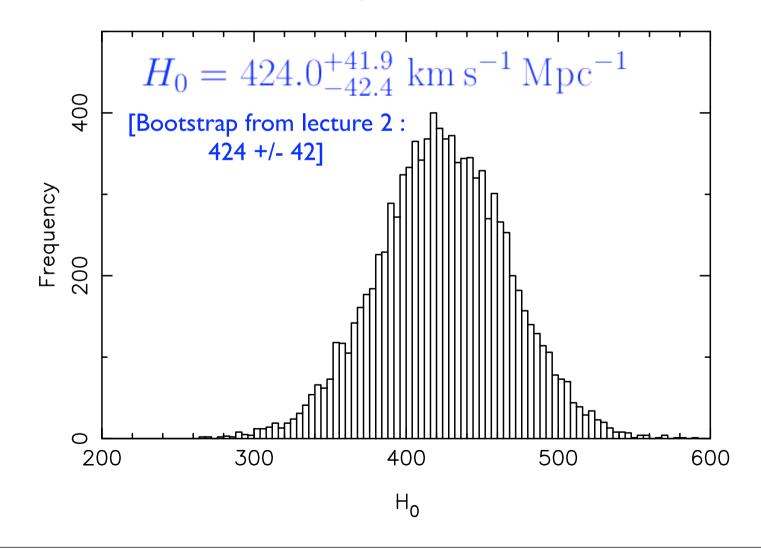
 Example 6 : Run a Monte Carlo simulation of Hubble's distance-redshift investigation ...



• Example 6 : Run a Monte Carlo simulation of Hubble's distance-redshift investigation ...



 Example 6 : Run a Monte Carlo simulation of Hubble's distance-redshift investigation ...



## Concluding remarks : which statistic to use?

- There are often multiple statistical approaches for any given problem. How do we decide which to use?
- Clarify : what is the question I am trying to answer?
- Are the conditions required by the test valid? [small numbers, Gaussian errors, data points independent, ...]
- What is the standard approach in your field? [... you want the community to understand you ...]
- Decide on the statistical test in advance : do not try multiple tests if seeking a result [confirmation bias...]
- Bootstrap or Monte Carlo offer powerful crosscheck

# Thank you for coming !

