Bibliography

- Bayesian Data Analysis, Carlin,Stern and Rubin, CHAPMAN & HAA/CRC
- Bayesian Reasoning in Data Analysis, Giulio D'Agnostini, World Scientific.
- ICIC Data Analysis Workshop 2016, Alan Heavens Lectures.
- MACSS 2016 LEcture notes.

Likelihood

- The probability , under the assumption of a model/theory, to observe the data as was actually obtained.

$$\mathcal{L} \qquad P(\text{Data}, \text{Model})$$

- For data that can be thought as samples of a sequence of normal random variables that have a mean and a variance, the likelihood is Gaussian

$$\mathcal{L} \propto \prod_i^n \frac{1}{2\pi\sigma_i^2} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma_i^2}\right)$$

mu will be the expected mean given our model

- A minimization of the Chi-square correspond to the maximization of the likelihood.

Gaussian Likelihood

$$-\ln(\mathcal{L}(\vec{x},\vec{y}|\vec{\theta})) \propto \frac{1}{2}\sum_i \left(\frac{(y_i - \lambda(x_i,\vec{\theta}))^2}{\sigma_i^2} \right)$$

In terms of data points and parameters.
Lambda is our model for y_i

How do we maximize the likelihood if there 2,3, or more parameters...?

How do we maximize the likelihood if we have a complex model?

What if the likelihood is not Gaussian...?

Bayesian approach

Probability (some definitions)

Typical answer

- "The ratio of the number of favorable cases to the number of all cases"
- " The ratio of the number of times the event occurs in a test series to the total number of trials in the series"

A subjective definition

- A formal definition would be: "The quality, state, or degree of something being supported by evidence strong enough make it likely though not certain to be true"
- A simple definition: "A measure of the degree of belief that an event will occur"

Probability Rules

$$0 < p(x) < 1$$

Probability of event x happens is coherent

$$p(x) + p(\sim x) = 1$$

Probability that event "x" happen, and probability of event x do not happen are complementary.

$$p(x, y) = p(x|y)p(y)$$

Product rule

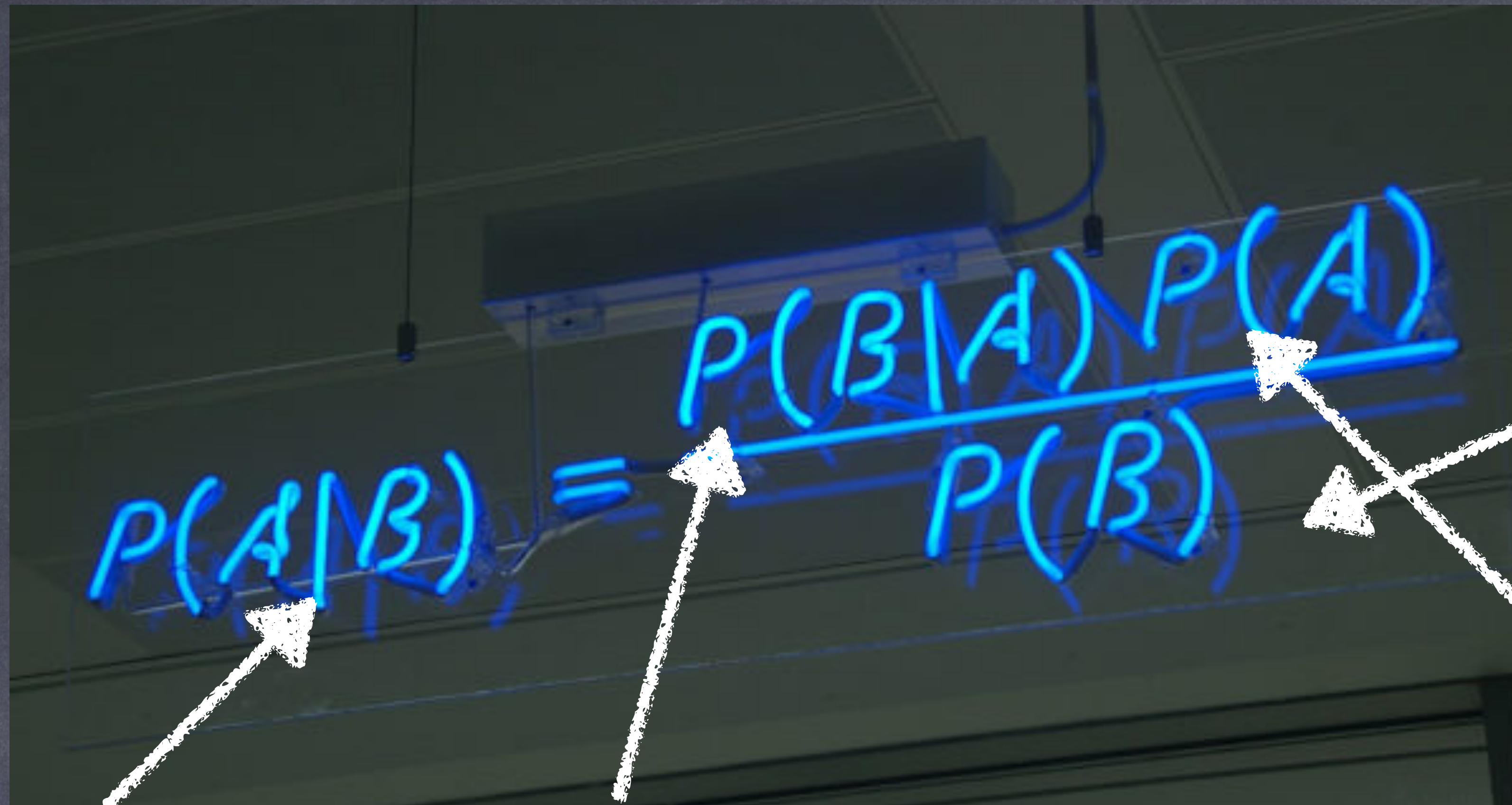
$$p(x) = \sum_i p(x, y_i)$$

Probability that event "x" happen, given that y happened: Marginalization

$$p(x) = \int p(x, y)dy$$

In the continuous limit we change the sum by an integral.

BAYES THEOREM arise from the these rules.



Evidence: The probability of de data, in all possibilities

Posterior

Degree of believe that some truth proposition event B implies that A is also truth.

Likelihood

Degree of believe to which some truth proposition A, event, implies that proposition B, event, is also true.

Prior

The degree of our a priori believe of proposition A, event, based on previous knowledge.

A question in cosmology would be: Given the observed CMB data with current experiments (A), what is the probability that the matter density of the Universe is between 0.9 and 1.1 (B).

$$P(0.9 < \Omega_m < 1.1 | Data) \,?$$

Bayes Theorem

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

Derive Bayes theorem from p(x,y)=p(y,x)

Example

Suppose you have the following results for an allergy test

1) It gives a positive result with probability 0.8, when patients have the allergy
2) It gives a false positive with probability 0.1.

If you make yourself the test, and result is positive, what is the probability that you actually have the allergy?

P(A): probability of having the allergy

P(T): probability of test being true

What is the question in terms of probability?

Example

Suppose you have the following results for an allergy test

1) It gives a positive result with probability 0.8, when patients have the allergy
2) It gives a false positive with probability 0.1
3) The probability of having thee allergy in the population is 0.01

If you make yourself the test, and result positive, what is the probability that you actually have the allergy?

P(A): probability of having the allergy

P(T): probability of test being true

What is the question in terms of probability?

We want p(A|T)

Example (Solution)

We want p(A|T)

We know p(T|A)=0.8, p(T|~A)=0.1   , P(A)=0.01

Bayes Theorem:

$$p(A|T) = \frac{p(T|A)p(A)}{p(T)}$$

$$p(A|T) = \frac{p(T|A)p(A)}{p(T,A) + p(T,\sim A)}$$

$$p(A|T) = \frac{p(T|A)p(A)}{p(T|A)p(A) + p(T,\sim A)p(\sim A)}$$

$$p(A|T) = \frac{0.8 * 0.01}{0.8 * 0.01 + 0.1 * 0.99} = 0.075$$

Ej.: Las catafixias de Chabelo (adapted from The Monty Hall problem)

There are three doors; behind one there is a prize. Suppose you choose door 1, and Chabelo opens door to show there is not a prize in there, and ask you whether you want to keep your original choice or want to change to door 3. What would you? Justify your response by computing the probabilites and use Bayes Theorem.

P(a): Prob prize is in door 1

P(b) : Prob prize is in door 2

P(c) : Prize is door 3

P(A|B) : Prob Chabelo opens door 1 given that you selected 2.

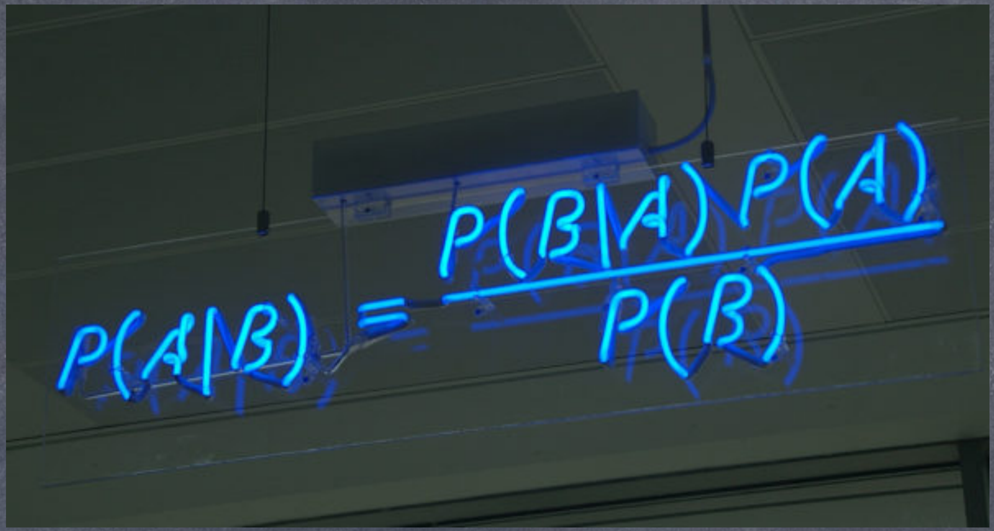
Events

B: You choose door 2.

A : Chabelo opens door 1, and prize is not in there.

Question: what is the probability that prize is in door 2, given that is not in 1?

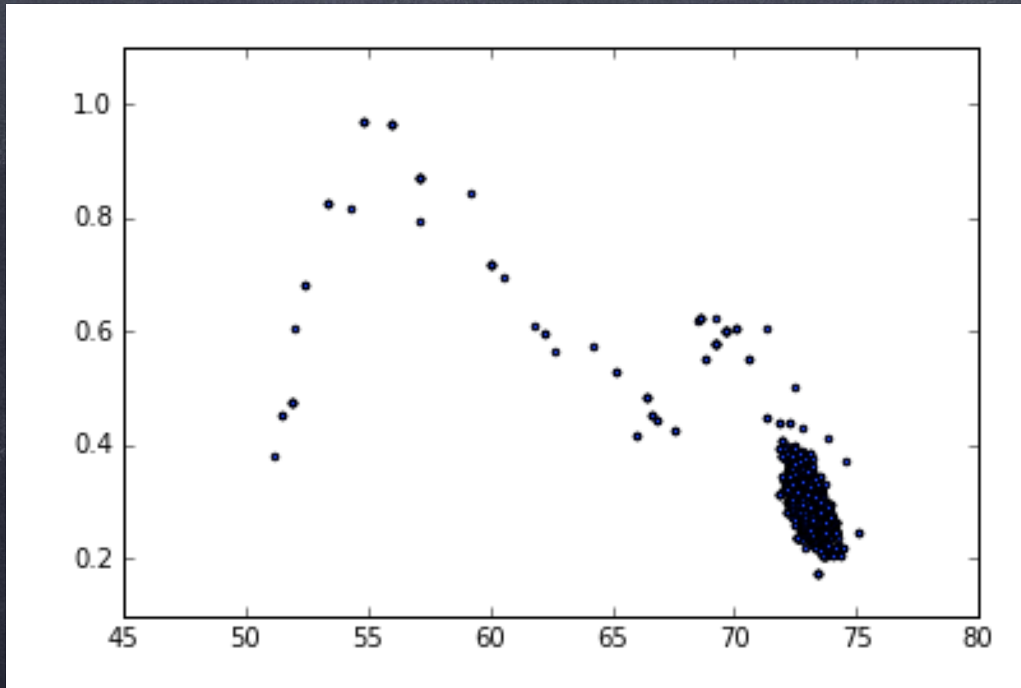
MonteCarlo Markov Chain for Bayesian Inference



$$P(\vec{\theta}\,|D) \propto L(D\,|\,\vec{\theta}\,) * P(\vec{\theta}\,)$$

$$\ln(P(\vec{\theta}\,|D)) \propto \ln(L(D\,|\,\vec{\theta}\,)) + \ln(P(\vec{\theta}\,))$$

Posterior and Likelihood depends on your model assumptions as well



Metropolis Algorithm

Draw random samples and accept them or reject them according to the Posterior

- Define an starting point for the parameters.
- Draw a sample parameter from a normal distribution centered in the starting point, and with some σ_0 dispersion.
- If the Posterior of a new sample, P(new), is higher than the old one, P(old) we accept the sample and save it. new—>old
- If the Posterior of a new sample is lower than the old one: draw a random number between 0-1, p(new)/p(old) is larger than such number then we accept it, not otherwise.
- Draw a new sample and start again….
- After many steps, look at the resultant distribution (the chains) of parameters, i.e., the Posterior…

The prior

- A common choice when you have no informations is to use flat priors within a given range.
- But you can use Gaussian, or another probability distribution if you have more information.
- For parameter inference is usual not worry about normalization factors in the prior, likelihood of posterior, but if we want to do model comparison, then it is more important.

MCMC

- Run different Chains starting in different positions of the parameters space. —> Walkers
- Give enough steps so that the Walkers forget where they started and arrived to the maximum Posterior.
- When should you stop? Different convergence criteria to discuss in next session…