Bayesian Inference in Processing Experimental Data

David Gordo Gómez
Universidad Autónoma de Madrid
Nico Giangiacomi
University of Bologna
HASCO Summer School 2014

August 1^{st} 2014

Summary

1. Introduction
   - Bayesian vs. Frequentist
   - Standard Rules of Probability
   - Bayes' Theorem
2. Priors
   - Importance of Priors
   - Choice of Priors
   - Conjugate Priors
   - Coin Flipping
3. Conclusions

Bayesian vs. Frequentist

Frequentist approach

- Ratio between the number of times an event happens in a run and the total number of trials.
- The experiment has to be done a large number of times (ideally infinite).
- It assumes that the process happens with the same probability every time.
- Very easy to apply once the conditions are satisfied.

Bayesian vs. Frequentist

Bayesian approach

- Probabilities have the intuitive meaning of the degree of belief that an event will occur.
- It has a subjective nature (Probabilities depend on our state of knowledge).
- Leads to the basic laws of probability, once the principle of *coherence* is introduced.
- Classical results are obtained (e.g. Maximum-Likelihood principle) under certain assumptions, that now are under control.

Bayesian vs. Frequentist

Belief and Coherence

- There is no single way to derive the basic rules of probability within the Bayesian approach.
- The concept of *degree of belief* and the principle of *coherence* are the usual guide to many Bayesians.
- The *degree of belief* can be easily explained considering a bet; the larger the degree of belief, the more will be the quantity of money that the better will pay to receive the same amount of money.
- The principle of coherence states that a person should be ready to accept bets in either direction, with odd ratios calculated from those values of probability.

Bayesian vs. Frequentist

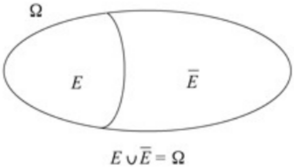
Let us clarify this with an example.

If a person states that the probability of Brazil to win the World Cup is 2/3 (*degree of belief*) he should accept a bet in which he puts 2€ and the opponent 1€.

Also, he should accept a bet against Brazil winning in which he puts 1€ and the opponent 2€ (principle of *coherence*).

Standard Rules of Probability

- Axiom 1: $0 \leq P(E) \leq 1$.
- Axiom 2: $P(\Omega) = 1$ (Certain Event).
- Axiom 3: $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ if $E_1 \cap E_2 = \emptyset$.



From the axioms the following properties can be derived:

- $P(E) = 1 - P(\overline{E})$
- $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$
- $P(E_1 \cap E_2) = P(E_1, E_2) = P(E_1|E_2)P(E_2)$

Bayes' Theorem

Bayes' Theorem is easily derived from the previous properties:

$$P(E \cap H) = P(E|H)P(H) = P(H \cap E) = P(H|E)P(E)$$

$$\Rightarrow P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

- The new condition E (experiment) alters our belief in H (hypothesis) using as an updating factor the probability of E given H.
- Symmetric treatment of expressions concerning hypothesis and observations.

Bayes' Theorem

Several Hypothesis

Things became more interesting when we consider a set of hypothesis H_j that all together are:

1. $\cup_j H_j = \Omega \Rightarrow \sum_j P(H_j) = 1$ (Exhaustive).
2. $H_j \cap H_k = \emptyset$ if $j \neq k$ (Exclusive).

For the complete class H the probability of event E is obtained by the summation over all possible hypothesis that can produce E.

$$P(E) = \sum_j P(E, H_j) = \sum_j P(E|H_j)P(H_j)$$

General Case

In the general case, we have several observations or experiments E_i and we have to include our current state of knowledge, generalizing the formula:

$$P(H_j|E_i, I) = \frac{P(E_i|H_j, I)P(H_j|I)}{P(E_i|I)}$$

The denominator $P(E_i|I)$ can be difficult to evaluate, so it is usually written in terms of the conditional probability $P(E_i|H_j, I)$, making evident that it is just a **NORMALIZATION FACTOR**.

$$P(H_j|E_i, I) = \frac{P(E_i|H_j, I)P(H_j|I)}{\sum_k P(E_i|H_k, I)P(H_k|I)}$$

Importance of Priors

Importance of Priors

$$P(H_j|E_i, I) \propto P(E_i|H_j, I)\ P(H_j|I)$$

Posterior Likelihood Prior

- Priors are seen by those critical of the Bayesian approach as the major weakness of it.
- Bayesian supporters see them as one of the advantages, because they explicitly admit the existence of prior information.
- Powerful because they allow to deal with realistic situations in which previous information can be taken into account.
- Crucial because we need them to make probability inversions via Bayes' Theorem.
- There are not prescriptions for the choice of priors, it is a highly debated issue among Bayesians.

Choice of Priors

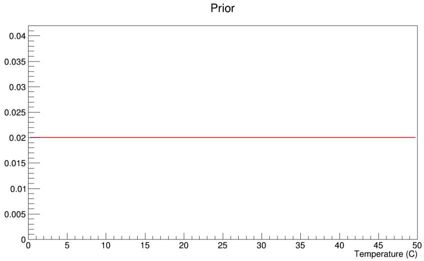
Choice of Priors

- The prior should tell us where the probability is concentrated, but without taking too seriously the details.
- Priors have to contain physical bounds (e.g. mass$\geq$ 0). Remember that where the prior is zero the posterior will be zero.
- It is important not to fully suppress unexpected possibilities (to accommodate '*surprises*').
- What matters is the gross value of the pdf and the influence on the posteriors.
- Mathematically convenient!.

Choice of Priors

If we consider the outcome of measuring the temperature at this room with a digital thermometer with one degree resolution, there are some values of the thermometer display you are more confident to read.

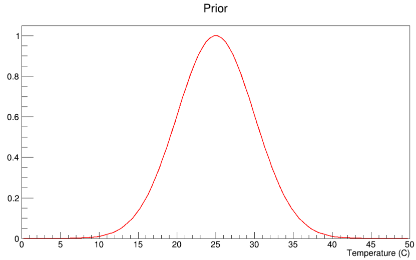
A flat distribution is not a good prior, it doesn't capture this information. It can be the prior for a person with no knowledge.



Choice of Priors

If we consider the outcome of measuring the temperature at this room with a digital thermometer with one degree resolution, there are some values of the thermometer display you are more confident to read.

It can be modeled better using a Gaussian distribution, in which we can introduce our degree of belief.



Conjugate Priors

Conjugate Priors

- Solving analytically a problem is often difficult. Modeling priors has been traditionally a compromise between a realistic model and mathematical simplicity.
- One can choose as a prior a function such as the posterior belongs to the same family as the prior. In that case it is called a **CONJUGATE PRIOR**.
- For example, given a Gaussian likelihood and choosing a Gaussian prior, the posterior is still a Gaussian. The Gaussian distribution is auto-conjugate.

Conjugate Priors

Useful Conjugate Priors

In this table are shown the most used conjugate priors with their respective likelihoods and posteriors:

likelihood $p(x \mid \theta)$	conjugate prior $p_0(\theta)$	posterior $p(\theta \mid x)$
$\text{Normal}(\theta, \sigma)$	$\text{Normal}(\mu_0, \sigma_0)$	$\text{Normal}(\mu_1, \sigma_1)$
$\text{Binomial}(N, \theta)$	$\text{Beta}(r, s)$	$\text{Beta}(r+n, s+N-n)$
$\text{Poisson}(\theta)$	$\text{Gamma}(r, s)$	$\text{Gamma}(r+n, s+1)$
$\text{Multinomial}(\theta_1, \ldots, \theta_k)$	$\text{Dirichlet}(\alpha_1, \ldots, \alpha_k)$	$\text{Dirichlet}(\alpha_1+n_1, \ldots, \alpha_k+n_k)$

Coin Flipping

Coin Flipping

Suppose we want to know if a coin is fair tossing it several times; i.e., we must infer p the probability of head.

Frequentist approach:

$$p = \lim_{N \to +\infty} \frac{h}{N}$$

Bayesian approach:

$$P(p|h,N) = \frac{P(h|N,p)P(p)}{\int dp\, P(h|N,p)P(p)} \propto P(h|N,p)P(p)$$

$h =$ number of heads, $N =$ number of trials

Coin Flipping

Likelihood: $P(h|N,p) = \binom{N}{h} p^h (1-p)^{N-h}$ Binomial!

Conjugate prior: $B(p;r,s) = \frac{p^{r-1}(1-p)^{s-1}}{\beta(r,s)}$

Where $\beta(r,s)$ is Euler's Beta function that can be written as:

$$\beta(r,s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} = \frac{(r-1)!(s-1)!}{(r+s-1)!}$$

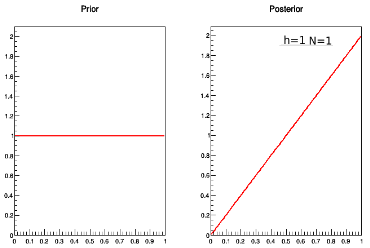
Mean value and variance are:

$$E(p) = \frac{r}{r+s}, \sigma^2(p) = E(p)^2 \frac{s}{r} \frac{1}{1+r+s}$$

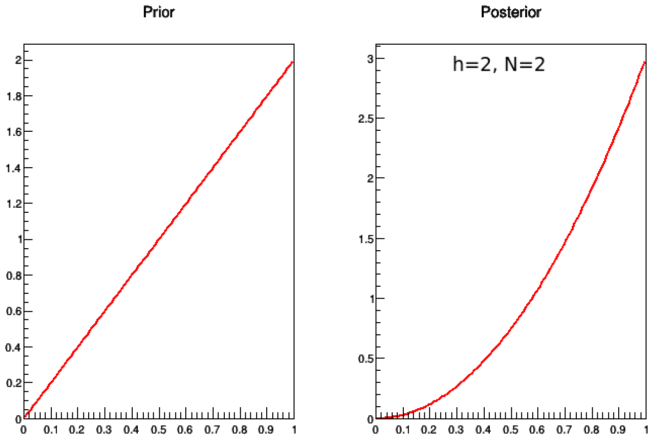
For $r = s = 1$ we have a flat distribution!

Coin Flipping

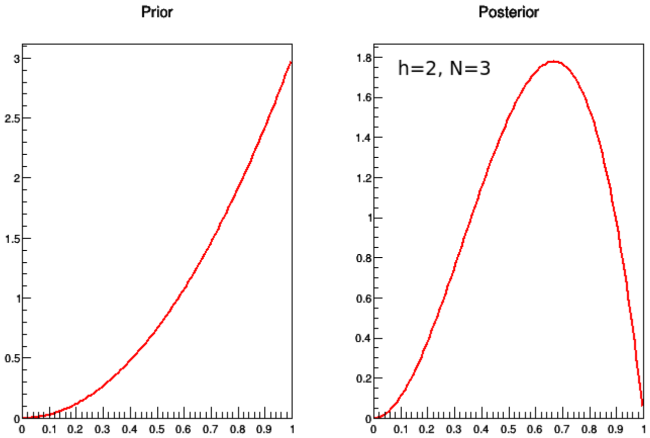
Every coin toss updates our knowledge, changing the prior for the next flip.
If we start with no preference about the value of p we can use a flat prior $(0 \leq p \leq 1)$.



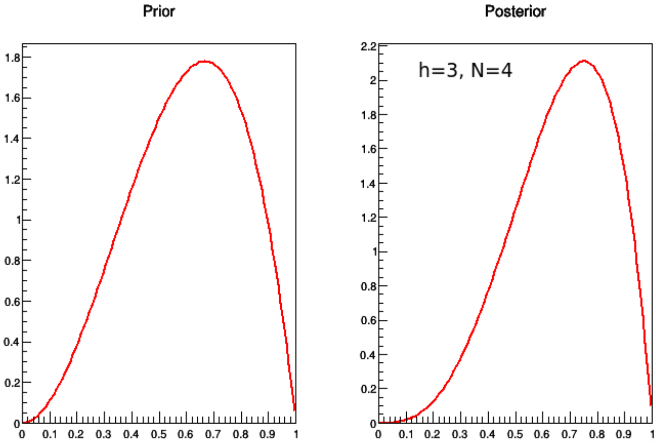
Coin Flipping



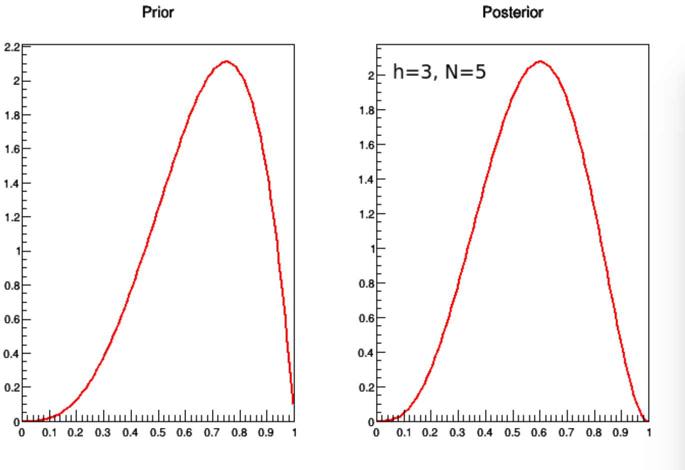
Coin Flipping



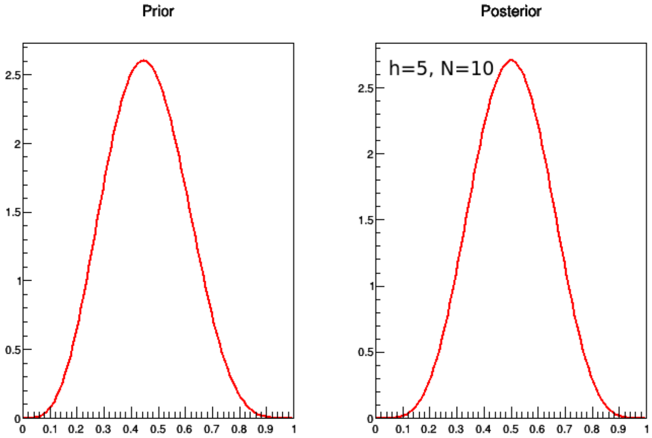
Coin Flipping



Coin Flipping



Coin Flipping



Conclusions

- Bayesian statistics is very close to the way of reasoning of physicists.
- It generalized frequentist statistics, demonstrating classical results under well determined assumptions, and including the subjective nature of probability.
- Although still debatable,priors are one of the most powerful features of the Bayesian approach. They allow us to include physical bounds and to update our state of knowledge after each experiment.
- Usually, after a large number of experiments the dependence on the prior is lost.
- Its choice has to be a compromise between being realistic and simplicity (with computers and MC integration this is not too important).

Conclusions