# Bayesian Inference

Jan Carbonell

July 22, 2019

Academy AI

# Introduction

# Frequentist vs Bayesian thinking

There are two major interpretations to probabilities: **Bayesian** and **Frequentist**:

- **Frequentist:** probabilities represent long term frequencies with which events occur. Commonly, a frequentist approach is referred to as the *objective* approach since there is no expression of belief and/or prior events in it.
- **Bayesian:** probabilities are treated as an expression of belief. The idea behind Bayesian thinking is to keep updating the beliefs as more evidence is provided. Since this approach deals with belief, it is usually referred to as the *subjective* view on probability.

# Frequentist vs Bayesian thinking

- **Frequentist:** a frequentist can say that the probability of having tails from a coin toss is equal to 0.5 on the long run. Each new experiment, can be considered as one of an infinite sequence of possible repetitions of the same experiment. Frequentists will never say "I am 45% (0.45) sure that there is lasagna for lunch today", since this does not happen on the long run.
- **Bayesian:** it is perfectly reasonable for a Bayesian to say "I am 50% (0.5) sure that there is lasagna for lunch today". By combining **prior** beliefs, and current events (the **evidence**), one can compute the **posterior**, i.e. the probability that there is lasagna today.

# Bayesian thinking

**Bayesian inference** is an extremely powerful set of tools for modeling any random variable, such as the value of a regression parameter, a demographic statistic, a business KPI, or the part of speech of a word. We provide our understanding of a problem and some data, and in return get a quantitative measure of how certain we are of a particular fact. This approach to modeling uncertainty is particularly useful when:

- Data is limited;
- We're worried about overfitting;
- We have reason to believe that some facts are more likely than others, but that information is not contained in the data we model on;
- We're interested in precisely knowing how likely certain facts are, as opposed to just picking the most likely fact.

# Bayesian thinking

The table below enumerates some applied tasks that exhibit these challenges, and describes how Bayesian inference can be used to solve them:

| Problem | Bayesian Solution | Real-World Examples |
| --- | --- | --- |
| Our dataset is small, but there is related information available | Use information from related systems as priors | • Evaluating new marketing campaigns<br>• Predicting effects of sales events<br>• Opening a business in a new market |
| Our model is extremely flexible (high variance models), so we want to prevent overfitting | Use priors that prefer values close to 0 (same concept as regularization in machine learning) | • Building a sentiment model on customer surveys<br>• Churn models |
| Our data is grouped in some way (nested or non-nested), so we want to let information flow between groups as needed | Build probability models for the system where priors are tied to parent or sibling groups | • Item pricing models for ecommerce<br>• Modeling data by geographic region |
| We want to understand how likely any parameter value is, rather than just making a best guess | Evaluating the full posterior of variables/parameters of interest | • Evaluating the effect of programs on LTV<br>• Financial risk analysis<br>• Evaluating A/B tests |

# Bayesian inference

In the philosophy of decision making, Bayesian inference is closely related to the Bayesian view on probability, in the sense that it manipulates **priors**, **evidence**, and **likelihood** to compute the **posterior**.

Given some event B, what is the probability that event A occurs?

# Bayesian inference

This is answered by Bayes' famous formula:

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

With:

- $P(A \mid B)$ is the **posterior**, what we wish to compute.
- $P(B \mid A)$ is the **likelihood**. Assuming A occurred, how likely is B.
- $P(A)$ is the prior, how likely the event A is regardless of evidence.
- $P(B)$ is the evidence, how likely the evidence B is regardless of the event.

# Bayesian inference - an example

Think of A as some proposition about the world, and B as some data or evidence. For example, A represents the proposition that it rained today, and B represents the evidence that the sidewalk outside is wet:

$$p(rain \mid wet) = \frac{p(wet|rain) \cdot p(rain)}{P(wet)} = \frac{p(wet|rain) \cdot p(rain)}{p(wet|rain) \cdot p(rain) + p(wet|norain) \cdot p(norain)}$$

$p(rain \mid wet)$ asks, **"What is the probability that it rained given that it is wet outside?"** To evaluate this question, let's walk through the right side of the equation. Before looking at the ground, what is the probability that it rained, $p(rain)$? Think of this as the plausibility of an assumption about the world. We then ask how likely the observation that it is wet outside is under that assumption, $p(wet \mid rain)$? This procedure effectively updates our initial beliefs about a proposition with some observation, yielding a final measure of the plausibility of rain, given the evidence.

# Bayesian inference - an example

This procedure is the basis for Bayesian inference, where our **initial beliefs are represented by the prior distribution** $p(rain)$, and our **final beliefs are represented by the posterior distribution** $p(rain \mid wet)$. The denominator simply asks, **"What is the total plausibility of the evidence?"**, whereby we have to consider all assumptions to ensure that the posterior is a proper probability distribution.

Bayesians are uncertain about what is true (the value of a KPI, a regression coefficient, etc.), and use data as **evidence** that certain facts are *more likely than others*. **Prior** distributions reflect our *beliefs before seeing any data*, and **posterior** distributions reflect our *beliefs after we have considered all the evidence*. To unpack what that means and how to leverage these concepts for actual analysis, let's consider the example of evaluating new marketing campaigns.

# Example: genetic probabilities

# Example: genetic probabilities

Let's put Bayesian inference into action using a very simple example. I've chosen this example because it is one of the rare occasions where the **posterior can be calculated by hand**. We will show how data can be used to update our belief in competing hypotheses.

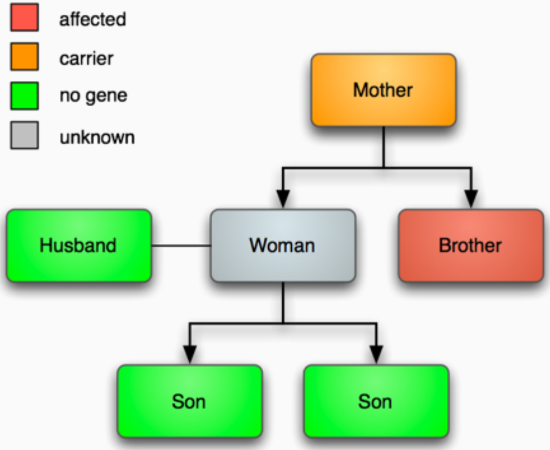
# Case study

**Hemophilia is a rare genetic disorder** that impairs the ability for the body's clotting factors to coagulate the blood in response to broken blood vessels. The disease is an **x-linked recessive** trait, meaning that there is only one copy of the gene in males but two in females, and the trait can be masked by the dominant allele of the gene.

This implies that males with 1 gene are affected, while females with 1 gene are unaffected, but carriers of the disease. **Having 2 copies of the disease is fatal**, so this genotype does not exist in the population.

## Case study

In this example, consider a woman whose mother is a carrier (because her brother is affected) and who marries an unaffected man. Let's now observe some data: the woman has **two consecutive (non-twin) sons who are unaffected**. We are interested in determining if the woman is a carrier.



# Case study - prior

To set up this problem, we need to set up our **probability model**. The unknown quantity of interest is simply an indicator variable $W$ that equals 1 if the woman is affected, and zero if she is not. We are interested in the probability that the variable equals one, given what we have observed:

$$P(W = 1 \mid s_1 = 0, s_w = 0)$$

Our prior information is based on what we know about the woman's ancestry: **her mother was a carrier**. Hence, the prior is $P(W = 1) = 0.5$. Another way of expressing this is in terms of the **prior odds**, or:

$$O(W = 1) = \frac{P(W=!)}{P(W=0)} = 1$$

# Case study - likelihood

Now for the **likelihood**, the form of this function is:

$$L(W \mid s_1 = 0, s_2 = 0)$$

This can be calculated as the probability of observing the data for any passed value for the parameter. For this simple problem, the likelihood takes only two possible values:

$$L(W = 1 \mid s_1 = 0, s_2 = 0) = 0.5 \cdot 0.5 = 0.25$$
$$L(W = 0 \mid s_1 = 0, s_2 = 0) = 1 \cdot 1 = 1$$

# Case study - Bayes formula

With all the pieces in place, we can now apply Bayes' formula to calculate the posterior probability that the woman is a carrier:

$$P(W = 1 \mid s_1 = 0, s_2 = 0) = \frac{L(W=1|s_1=0,s_2=0) \cdot P(W=1)}{L(W=1|s_1=0,s_2=0) \cdot P(W=1) + L(W=0|s_1=0,s_2=0) \cdot P(W=0)}$$

Hence, there is a 0.2 probability of the woman being a carrier.

# Case study - doing it in Python

Its a bit trivial, but we can code this in Python:

```
prior =0.5
p =0.5

L =lambda w, s: np.prod([(1-i, p**i *(1-p)**(1-i))[w] for i in
                                                    s])

s =[0,0]

post =L(1, s) *prior /(L(1, s) *prior +L(0, s) *(1 -prior))
print(post)
```

Output: 0.20000000000000001

# Case study - doing it in Python

Now, what happens if the woman has a **third unaffected child**? What is our estimate of her probability of being a carrier then?

Bayes' formula makes it easy to **update analyses with new information**, in a sequential fashion. We simply assign the posterior from the previous analysis to be the prior for the new analysis, and proceed as before:

```
print(L(1, [0]))
```

Output: 0.5

```
s =[0]
prior =post

print(L(1, s) *prior /(L(1, s) *prior +L(0, s) *(1 -prior)))
```

Output: 0.11111111111111112