Bayes' rule

Giorgio Corani

```
{r      setup,      include=FALSE      }
#        library(reticulate)        #
use_condaenv("bda") #
```

```{python, echo=FALSE}
# import numpy as np
# import pandas as pd
# import arviz as az
# import scipy.stats as stats
# from matplotlib import pyplot as plt
```

```{python, echo=FALSE} #
az.style.use('arviz-darkgrid') #
```

Objectives

- To learn Bayes theorem
- Learn how probabilities are interpreted in Bayesian statistics
- Understand prior, posterior, and predictive posterior distributions

Introduction

Most statistics courses teach *classical* (also called *frequentist*) statistics.

Bayesian statistics

- We instead focus on Bayesian methods.
- We will assign a prior distribution to every unknown parameter.
- We will update the prior distribution given the data using Bayes' theorem.

Notation

- We denote random variable with uppercase letters and their specific values by lowercase letters.
- For instance $p(Y = y)$, shortened as $p(y)$.
- Hence $p(y)$ denote the probability of Y taking the specific value y and we have $\sum_y p(y) = 1$.
- Instead $P(Y)$ is the probability distribution for variable Y, i.e, a table containing as many elements as the possible values of Y.

Conditional probability $p(Y = y \mid X = x)$

- The **conditional probability** is:

$$p(Y = y|X = x) = \underbrace{p(y \mid x)}_{\text{shortcut notation}} = \frac{p(x, y)}{p(x)}$$

Product rule

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

- If X and Y are independent:
  - $p(y \mid x) = p(y)$
  - $p(x|y) = p(x)$
  - $p(x, y) = p(x)p(y).$

Sum rule

- Given the joint $p(X, Y)$
- The marginal probability of $X = x$ is given by the **sum rule**

$$p(x) = \sum_y p(x, y)$$

Bayes' rule

■ Hence we derive **Bayes' rule**:

$$p(y \mid x) = p(x, y)/p(x)$$
$$= \frac{p(x \mid y) \cdot p(y)}{\sum_{y'} p(x \mid y) \cdot p(y)}$$

Bayes' theorem

- Let us consider an unobserved variable H and an observable variable Y. Our model needs:
  - the prior $p(H)$
  - the conditional distribution $p(Y|H)$
- Having observed $Y = y$, we obtain the posterior distribution of H via Bayes' rule. That is, we *condition* on $Y = y$:

$$p(h|y) = \frac{p(h)p(y \mid h)}{p(y)} = \frac{p(h)p(y \mid h)}{\sum_h p(y, h)}$$

$$p(y) > 0$$

This follows from the identity

Prior

- $p(H)$ is the prior distribution of H; it represents how likely are the different values of H, according to our beliefs, before we see any data.
- If H is a discrete variable with k possible values, $p(H)$ is also called a probability mass function (*pmf*).

Likelihood

- $p(Y \mid h)$ represents the distribution of the possible outcome Y given $H = h$. This is also called the *observation distribution*.
- When we have a specific observation y, we get the function $p(y|h)$, the likelihood.
- The likelihood is a function of h, since y is fixed, but it is not a probability distribution, since it does not sum to one.

Marginal likelihood

- The denominator of Bayes' rule is a normalizing constant, referred to as the marginal likelihood.
- It marginalizes the likelihood over the states of the unobserved variable H:

$$p(y) = \sum_h p(h)p(y \mid h) = \sum_h p(h, y)$$

Posterior

Normalizing the joint distribution by computing $p(H = h, Y = y)/p(Y = y)$ for each h gives the posterior distribution $p(H = h|Y = y)$; this represents our new belief state about the possible values of H.

- We can summarize Bayes rule as :

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{marginal likelihood}}$$

- The symbol $\propto$ denotes *proportional to*: we ignore the denominator, which is a constant

Testing for Covid 19 (Murphy, Sec 3.2.1)

- You decide to take a diagnostic test to check if you have contracted Covid. You want to make inference about your health H:
  - H=1: you are infected
  - H=0: you are healthy

The diagnostic test

The outcome of the test is:

- T=1: positive(infection)
- T=0: negative (healthy)

We want to determine $p(H|t)$, where t is the test outcome.

Likelihood

- $p(T = 1 \mid H = 1)$: the probability of a positive test given the infection (87.5% for our test). This is also called *true positive rate (TPR)*.
- $p(T = 0 \mid H = 0)$, i.e., the probability of a negative test given that the patient is healthy (0.975 for the test).
  This is also called *true negative rate.*

Truth/Test	Y=0	Y=1
healthy	0.975	0.025
infected	0.125	.875

Prior $P(H)$

- $p(H = 1)$ is the prevalence of the disease.
- We set $p(H = 1)$ = 0.1 (i.e., 10%), i.e., the prevalence in New York City in Spring 2020.

Posterior probability of being sick, given a positive test

$$p(H = 1 | Y = 1) = \frac{\overbrace{P(H = 1)}^{\text{prior: prevalence}} \overbrace{P(Y = 1 | H = 1)}^{\text{likelihood}}}{\underbrace{P(Y = 1)}_{\text{prob of the data: marginal likelihood}}}$$

Marginal Likelihood

■ Overall probability of having a positive test, i.e.,
the probability of testing positive while infected
plus the probability of testing positive while
healthy.

$$p(Y = 1) = P(Y = 1, H = 1) + P(Y = 1, H = 0)$$
$$= 0.1 \times + 0.025 \times 0.9$$
$$= 0.00225$$

Posterior

$$p(H = 1|Y = 1) = \frac{P(H = 1)P(Y = 1|H = 1)}{P(Y = 1)}$$

$$= \frac{0.1 \times 0.875}{0.00225}$$

$$= 0.795$$

- The posterior probability of being negative is:

$$p(H = 0|Y = 1) = 1 - p(H = 1|Y = 1) = 0.205$$

Your turn

There is still a 0.205 probability of being negative.
You then perform a second more powerful test Z,
with the following characteristics:

Truth/Test	Z=0	Z=1
healthy	0.99	0.01
infected	0.05	.95

Which is the probability of being positive, if also the
second test is positive?

Posterior

Our prior now incorporates the outcome of the first test.

- $p(H = 0)$=0.205
- $p(H = 1)$=0.795

Bayes' rule is written as in the previous case.

However the prior is already conditional on Y.

$$p(H = 1|Z = 1, Y = 1) = \frac{P(H = 1|Y = 1)P(Z = 1|Y}{P(Z = 1)}$$

$$(1)$$

$$(2)$$

We assume the two tests to be independent given

Another exercise

- We have two coins:
    - the first coins lands heads of tails with equal probability
    - the second coin is rigged and always lands heads.
- We take one coin at random and we get heads. What is the probability that this coin is the rigged one?

Solution

Since the coins are randomly chosen, the prior is:
$$P(\text{fair}) = P(\text{rigged}) = 0.5$$
The likelihood is:
$$P(\text{head} \mid \text{fair}) = 0.5$$
$$P(\text{head} \mid \text{rigged}) = 1$$

Solution - 2

The posterior is:

$$P(\text{rigged} \mid \text{head})$$

$$= \frac{P(\text{rigged})P(\text{head} \mid \text{rigged})}{P(\text{head})}$$

$$= \frac{P(\text{rigged})P(\text{head} \mid \text{rigged})}{P(\text{head, rigged}) + P(\text{head, fair})}$$

$$= \frac{P(\text{rigged})P(\text{head} \mid \text{rigged})}{P(\text{rigged})P(\text{head} \mid \text{rigged}) + P(\text{fair})P(\text{head} \mid \text{fair})}$$

$$= \frac{0.5 \times 1}{0.5 \times 1 + 0.5 \times 0.5)} = \frac{0.5}{0.5 \times 1 + 0.5 \times 0.5)} = \frac{2}{3}$$

Your turn

- What would be the posterior probability of having the rigged coin, if we observe a

formula da aggiornare!!

$$p(H = 1 | Z = 1, Y = 1) =$$

$$= \frac{P(H = 1)P(Y = 1 | H =}{P(Y = 1, H = 1) + P(Y =}$$

$$= \frac{P(H = 1)P(}{P(Y = 1 | H = 1)P(H = 1) +}$$

$$= \frac{0.1 \times 0.875}{0.1 \times +0.025 \times 0.9}$$

$$= 0.795$$

We used Bayes' rule to sequentially update the
probability of being positive, as new observations

Possible exercises - 1

*There are three doors, behind one of them a goat, behind the other two, a coupon to have tea with Eduardo Feinmann.
*You are asked to choose one of them. Once the host of the program has been chosen (which he knows what is behind each door), he opens one of the doors that contains the coupon to spend some time with Eduardo Feinmann. At that moment the presenter warns you that you have the possibility of changing doors or keeping the door you initially chose. Which is the best option?

Possible exercises - 2

A woman is the mother of two children. Assume
that the *a priori* girls and boys are equally probable.
* The woman says she has a boy. What is the
probability that one of the children is a girl?

- Blood is found at a crime scene.
- The blood is of a type that is only present in 1% of the population.

Possible exercises - 3a

The prosecutor states: "If the accused were innocent, the probability that he had the same blood type found at the crime scene would be 1%, and 99% if he were guilty, therefore the probability is that he is guilty".

- This reasoning is incorrect, explain why.

The defense attorney states: "The crime occurred in a city of 500,000 inhabitants, so 5,000 people have that blood type, therefore the defendant only has a 1/5,000 chance of being responsible.

- This reasoning is also incorrect, explain why.