

Bayes' rule

Giorgio Corani

Bayesian Data analysis and Probabilistic
Programming

- Chap. 2 of *Bayes Rules! An Introduction to Applied Bayesian Modeling*
 - <https://www.bayesrulesbook.com/chapter-2.html>

- According to a study, 40% of the articles taken from a certain website are fake news and 60% are real.
- The usage of an exclamation point in the title is however uncommon in real article.
 - 26.67% of fake news titles use an exclamation point
 - only 2.22% of real news titles use it.

Bayes' rule

- Given that an article contains the exclamation point, which is its probability of being fake news?

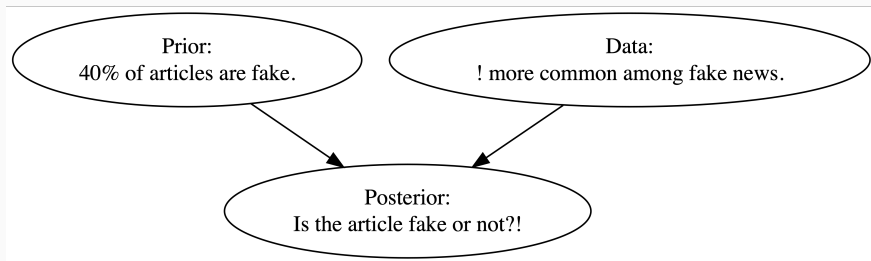


Figure 1: Bayesian knowledge-building for whether or not the article is fake

Prior probability model

- Consider the random variable A (article) with possible states $\{real, fake\}$.
- Prior probability, i.e., before having seen the article:
 - $P(article = fake) = 0.4$
 - $P(article = real) = 0.6$

A further random variable

- Variable B denotes whether the article title contains or not an explanation point.
- Its possible states are $\{!, \sim!\}$

Conditional probability

- The exclamation point in the title is more compatible with fake news than with real news.
- If an article is fake, there is 26.67% chance it uses exclamation points in the title.
- This is the conditional probability $P(!|fake)$
- The conditional probability $P(!|real)$ is instead 0.0222.

Conditional vs unconditional probability

- Let A and B be two events.
- $P(A)$: unconditional (or marginal) probability of A . It measures the probability of observing A , without any knowledge of B .
- $P(A|B)$: conditional probability of A given B : probability of observing A since B occurred.

- A is not observed.
- Comparing $P(A|B)$ vs $P(A)$ reveals how much the observation of B informs us about the most likely value of A .
- $P(A|B)$ can be larger, smaller or equal to $P(A)$.

- Probability of joining an orchestra, given that we practice clarinet every day:

$$P(\text{orchestra} \mid \text{practice}) > P(\text{orchestra})$$

- Probability of getting the flu given that we wash thoroughly our hands:

$$P(\text{flu} \mid \text{wash hands}) < P(\text{flu})$$

$$P(A|B) \neq P(B|A)$$

- Roughly 100% of puppies are adorable:

$$P(\text{adorable} \mid \text{puppy}) = 1$$

- But an adorable object is not necessarily a puppy:

$$P(\text{puppy} \mid \text{adorable}) < 1$$

$$P(B|A) > P(B) \text{ or } P(B|A) < P(B)?$$

- A = you just finished reading a book from a certain author and you enjoyed it!
- B = you will also enjoy the newest novel from the same author.

$P(B|A) > P(B)$ **or** $P(B|A) < P(B)$?

- A = it's 0 degrees Celsius tonight.
- B = tomorrow it will be very warm.

$$P(B|A) > P(B) \text{ or } P(B|A) < P(B)?$$

A woman is the mother of two children.

- A = the first child is a boy
- B = the first child is a girl

Independent events

- A and B are independent if the occurrence of B doesn't tell us anything about the occurrence of A :

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

- For instance:
 - event A : rider Q wins the next motoGP race
 - event B : my coin lands tails
 - Since the coin does not affect the performance of rider Q, marginal and conditional probability of A are the same.

Notation for random variables

- We denote random variable with uppercase letters and their specific values by lowercase letters.
- $P(Y)$: probability distribution for variable Y , i.e, a table containing as many elements as the possible values of Y .
 - Also called a probability mass function (pmf) if Y is discrete.
- Specific values are denote by lowercase; for instance, y .
- $p(y)$ denotes the probability of $P(Y = y)$, where $\sum_y p(y) = 1$.

- The likelihood measures the relative compatibility of data A with the different values of B .
- $P(A|b)$: probability of the different values of A , after we observe $B = b$.

Prior probabilities and likelihoods of fake news.

	fake	real
prior probability	$P(\text{fake})=0.40$	$P(\text{real})=0.6$
likelihood	$P(! \mid \text{fake})=0.26$	$P(! \mid \text{real})=0.02$

- Prior probabilities add up to 1 but the likelihoods do not.
- The likelihood function is not a probability function; instead it is a way to compare the relative compatibility of the observation (title contains !) with the different states of article A (fake or real).

Joint probability of the fake news and exclamation point

	fake	real
!		
~!		

- The joint probability $P(a, b)$ is the probability of observing both $A = a$ and $B = b$.
- It is defined as

$$P(a, b) = P(a|b)P(b)$$

Joint probability of the fake news and exclamation point

$$\begin{aligned}P(\text{fake}, !) &= P(\text{fake})P(! \mid \text{fake}) \\&= 0.4 \cdot 0.2667 = 0.1067\end{aligned}$$

	fake	real
!	0.1067	
~!		

Joint probability of fake news, no exclamation point

$$\begin{aligned}P(\text{fake}, !)&= P(\text{fake})P(! \mid \text{fake}) \\&= 0.4 \cdot (1 - 0.2667) = 0.2933\end{aligned}$$

	fake	real
contains !	0.1067	
does not contain !	0.2933	

Filling up the joint table

$$\begin{aligned}P(\text{real}, !) &= P(\text{real})P(! \mid \text{fake}) \\ &= 0.6 \cdot 0.0222 = 0.0133\end{aligned}$$

	fake	real
contains !	0.1067	0.0133
does not contain !	0.2933	

Filling up the joint table

$$\begin{aligned}P(\text{real}, \text{no !}) &= P(\text{real})P(\text{no !} \mid \text{fake}) \\&= 0.6 \cdot (1 - 0.0222) \\&= 0.6 \cdot 0.9778 = 0.5867\end{aligned}$$

	fake	real
contains !	0.1067	0.0133
does not contain !	0.2933	0.5867

Marginal distribution and marginal probability

- The marginal distribution of A is obtained by summing the joint distribution over all states of B .

	fake	real
!	0.1067	0.0133
\sim !	0.2933	0.5867
	0.4	0.6

- The marginal distribution of A is {fake=0.4; real=0.6}.
- The marginal probability $P(A = fake)$ is $P(fake, !) + P(fake, \sim !)$.

Marginal distribution and marginal probability

- The marginal distribution of A is obtained by summing the joint distribution over all states of B .

	fake	real
!	0.1067	0.0133
\sim !	0.2933	0.5867

- The marginal distribution of B is $\{! = 0.12; \sim ! = 0.88\}$.

Computing marginal probabilities

- If instead of the joint, you have $P(A)$ and $P(B|A)$, the marginal probability $P(B = b)$ is $P(b) = \sum_a P(b, a) = \sum_a P(b | a)P(a)$.

$$\begin{aligned}P(!) &= P(\text{real}, !) + P(\text{fake}, !) \\&= P(\text{real})P(! | \text{real}) + P(\text{fake})P(! | \text{fake}) \\&= 0.4 \cdot 0.0222 + 0.6 \cdot 0.2667 \\&= 0.1067 + 0.1600 = 0.2667\end{aligned}$$

Sum rule

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

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

Joint and conditional probability

- The joint probability $P(a, b)$, shorthand for $P(A = a, B = b)$, is computed by weighting the conditional probability $P(a|b)$ by the marginal probability $P(b)$:

$$P(a, b) = P(a|b)P(b)$$

- When A and B are independent, $P(A, B) = P(A)P(B)$.
- Dividing both sides by $P(B)$, and assuming $P(B) \neq 0$, yields the conditional probability of A given B :

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

- The conditional probability $P(a, b)$ compares the probability that $A = a, B = b$ occur together with the overall probability that b , $P(b)$.

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

- The **conditional probability** is:

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

Posterior probability that the article is fake

- We want the posterior probability of the article being fake given that it uses exclamation points.

$$P(fake | !) = \frac{P(fake)P(! | fake)}{P(!)} = \frac{0.4 \cdot 0.2667}{0.12} = 0.889$$

.

$$\begin{aligned} p(Y = y \mid X = x) &= p(y \mid x) &= \frac{p(x, y)}{p(x)} \\ &= \frac{p(x \mid y) \cdot p(y)}{\sum_{y'} p(x \mid y) \cdot p(y)} \end{aligned}$$

Using Bayes' theorem

- Unobserved variable H
- Observable variable Y .

Define: * the prior probability $P(H)$ * 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 $Y = y$:

$$p(h|y) = \frac{p(h)p(y | h)}{p(y)} = \frac{p(h)p(y | h)}{\sum_h p(y, h)}$$
$$p(y) > 0$$

- $p(H)$ 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

- $p(Y | h)$ represents the distribution of the possible outcome Y given $H = h$.
- 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 to the observed value.
- The likelihood is not a probability distribution; 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 :

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

- 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 independent of H .
- Using Bayes' rule to update a distribution of a quantity of interest H , given the observed data, is a Bayesian inference, or posterior inference.

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

The diagnostic test

The test T can be either:

- positive
- negative

We want to determine the probability distribution $p(H|T = \textit{positive})$.

Test performance

Assume the conditional probability of the test outcome, given an infected person, to be:

- $p(\text{positive} \mid \text{infected})$: the probability of a positive test for an infected persons is 87.5%.

	test negative	test positive
infected	0.125	.875

Test performance

Assume the conditional probability of the test outcome, given an healthy person, to be:

- $p(\text{negative} \mid \text{healthy})$, i.e., the probability of a negative test for a healthy patient, 97.5%

	test negative	test positive
infected	0.125	.875

Prior $P(H)$

- The *prevalence* is the percentage of persons affected by the disease.
- The covid prevalence in New York City 2020 was 10%.

	$H=\text{healthy}$	$H=\text{infected}$
probability	0.9	.1

Posterior probability of being infected, given a positive test

$$p(\textit{infected} \mid \textit{positive}) = \frac{\overbrace{P(\textit{infected})}^{\text{prior: prevalence}} \overbrace{P(\textit{positive} \mid \textit{infected})}^{\text{likelihood}}}{\underbrace{P(\textit{positive})}_{\text{prob of observing a positive test: marginal likelihood}}}$$

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

$$\begin{aligned}P(\text{positive}) &= P(\text{positive, infected}) + P(\text{positive, healthy}) \\&= 0.875 \times 0.1 + 0.025 \times 0.9 \\&= 0.11\end{aligned}$$

$$\begin{aligned}P(H = 1|Y = 1) &= \frac{P(\text{infected})P(\text{positive} \mid \text{infected})}{P(\text{positive})} \\&= \frac{0.1 \times 0.875}{0.11} \\&= 0.795\end{aligned}$$

- The posterior probability of being healthy is:

$$p(\text{healthy} \mid \text{positive}) = 1 - p(\text{infected} \mid \text{positive}) = 0.205$$

Discussion

- The positive outcome of the test increases your probability of being infected from 0.1 to 0.795.

Exercise

- Work out the probability of being infected if you test negative
 - 0.014

Yet another exercise

- We have two coins:
 - the first coin lands heads or 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?

Since the coins are randomly chosen, the prior is:

$$P(\text{fair}) = P(\text{rigged}) = 0.5$$

The likelihood is:

$$P(\text{head} \mid \text{fair}) = 0.5$$

$$P(\text{head} \mid \text{rigged}) = 1$$

The posterior probability of the coin being rigged is:

$$\begin{aligned} & P(\text{rigged} \mid \text{head}) \\ &= \frac{P(\text{rigged})P(\text{head} \mid \text{rigged})}{P(\text{head})} \\ &= \frac{P(\text{rigged})P(\text{head} \mid \text{rigged})}{P(\text{head})} \\ &= \frac{0.5 \times 1}{P(\text{head})} \end{aligned}$$

Computing the denominator

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

where

$$\begin{aligned} P(\text{head}, \text{rigged}) &= P(\text{head}|\text{rigged})P(\text{rigged}) \\ &= 1 \times 0.5 \end{aligned}$$

$$\begin{aligned} P(\text{head}, \text{fair}) &= P(\text{head}|\text{fair})P(\text{fair}) \\ &= 0.5 \times 0.5 \end{aligned}$$

$$P(\text{head}) = 0.5 \times 1 + 0.5 \times 0.5 = 0.75$$

to be filled...

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

formula da aggiornare!!

$$\begin{aligned} p(H = 1|Z = 1, Y = 1) &= \\ &= \frac{P(H = 1)P(Y = 1|H = 1)}{P(Y = 1, H = 1) + P(Y = 1, H = 0)} \\ &= \frac{P(H = 1)P(Y = 1|H = 1)}{P(Y = 1|H = 1)P(H = 1) + P(Y = 1, H = 0)} \\ &= \frac{0.1 \times 0.875}{0.1 \times +0.025 \times 0.9} \\ &= 0.795 \end{aligned}$$

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

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?