

Being a Basic Bayesian

EDP 619 Week 9

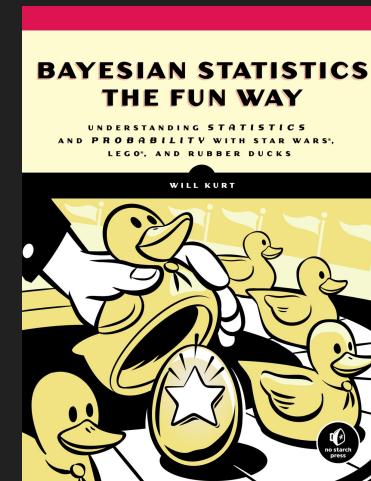
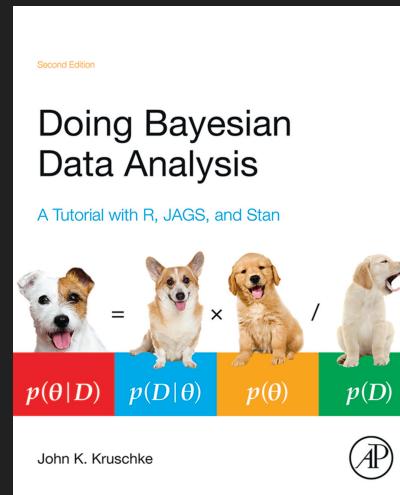
Dr. Abhik Roy

Note



We are going to try to keep this as light as possible, but you will have to think outside of your traditional training. For most of you, you'll likely see this as a new way of thinking and may even appear to be counterintuitive - however you may find this idea of iterative refining of one's beliefs is a fundamental principle of being human

If you are interested in learning more about the statistics, here is an absolutely biased selection of texts you may want to look at. While both employ R, the one on the right uses it significantly less than the text sitting on the left.



As of this writing, both texts are available as ebooks through WVU Libraries. Please click on the images above to access them



Bayesian Inference

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

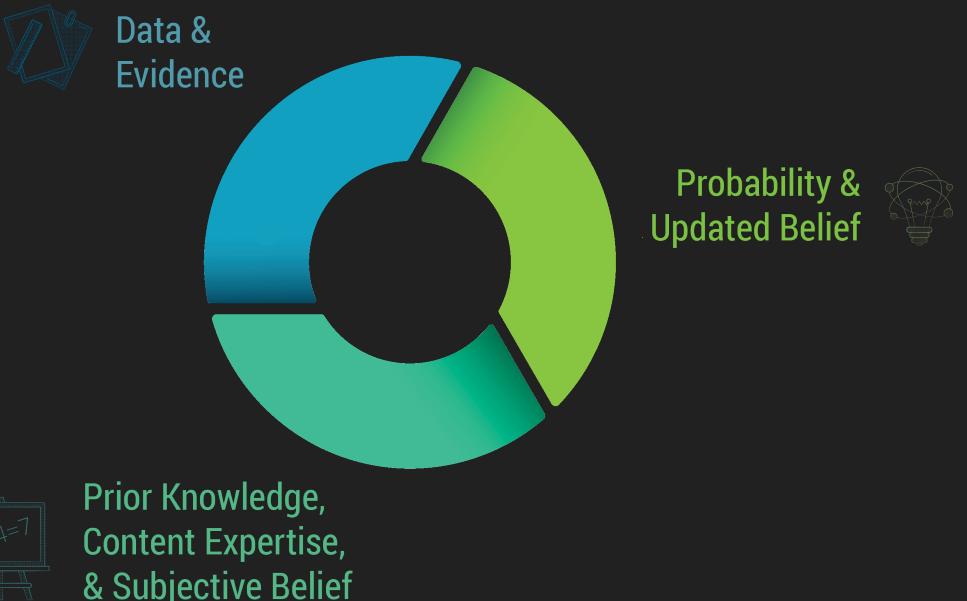
The Idea



1. Begin with

- | a hypothesis
- | a degree of belief (in the hypothesis)

So based on domain expertise or prior knowledge, we assign a non-zero probability to that hypothesis



2. Then

- | gather data

Collection is probability based and done without a purpose

3. Finally

- | update our initial beliefs (if needed)

If the data supports the hypothesis then the probability goes up, and if not then it goes down. ***Then recursively begin the cycle again with the updated probabilities***



Getting A Different Perspective

Is your understanding of *conditional probability* a bit fuzzy or possibly even non existent? That is understandable - you probably took statistics from a frequentist, that's ok! Click on the icons below to take a look at the two videos from the wonderful series Crash Course Statistics. I'm 95% sure you won't regret it!





Hopefully Some Review

$$P(A \text{ and } B)$$

means the probability of A and B happening

$$P(A | B)$$

means the probability of A happening if B happens

$$P(A \text{ and } B) = P(B \text{ and } A)$$

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

As long as both events happen, we DO NOT care about the order

One event relies on the other happening so we ABSOLUTELY care about the order



And Some More Review

$$P(\textcolor{blue}{A} \text{ and } \textcolor{red}{B})$$

has a standard formula regardless
if they are dependent or independent

$$P(\textcolor{blue}{A} \text{ and } \textcolor{red}{B}) = P(\textcolor{red}{B} \mid \textcolor{blue}{A}) \cdot P(\textcolor{blue}{A})$$

$$P(\textcolor{red}{B})$$

is the average
probability of a positive test result¹

$$P(\textcolor{red}{B}) = P(\textcolor{blue}{A} \text{ and } \textcolor{red}{B}) + P(\textcolor{red}{B} \mid \neg \textcolor{blue}{A}) \cdot (1 - P(\textcolor{blue}{A}))$$

¹ $\neg \textcolor{blue}{A}$ is one of many ways to write shorthand for the statement NOT $\textcolor{violet}{A}$. In an absolutely non confusing way, you may also see variants like $\sim \textcolor{violet}{A}$, $\tilde{\textcolor{blue}{A}}$, or $\overline{\textcolor{blue}{A}}$ just to name a few that all denote the same thing



Labels

Now that we have some understanding of what the parts of Bayes' Theorem are, let's take a look at the names of each

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

posterior
the probability of A being true
given B is true

likelihood
the probability of B being true
given A is true

prior
the probability of
 A being true (the
probability you
guess or know)

marginalization
the probability of B being true



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

$$P(\textcolor{blue}{A} | \textcolor{red}{B}) = \frac{P(\textcolor{blue}{A} \text{ and } \textcolor{red}{B})}{P(\textcolor{red}{B})}$$

$$P(\textcolor{red}{B} | \textcolor{blue}{A}) = \frac{P(\textcolor{red}{B} \text{ and } \textcolor{blue}{A})}{P(\textcolor{blue}{A})}$$

Some Algebra!

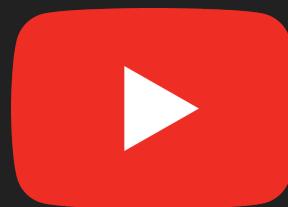
$$P(\textcolor{blue}{A} \text{ and } \textcolor{red}{B}) = P(\textcolor{blue}{A}) \cdot P(\textcolor{red}{B} | \textcolor{blue}{A})$$

$$P(\textcolor{red}{B} \text{ and } \textcolor{blue}{A}) = P(\textcolor{red}{B}) \cdot P(\textcolor{blue}{B} | \textcolor{blue}{A})$$



Getting A Different Perspective

Still a bit or all the way confused? Then try going through this incredible video from Josh Starmer by clicking on the icon below. Now I'm 98% sure you won't regret it!



Example





Screening

Let's say the chance of a test for infection coming back **positive** given a person was **bitten by a zombie** is

$$P(+ \mid \text{zombie bite}) = 0.90$$

This is known as as test **sensitivity**

And the chance of a test for infection coming back **negative** given a person was **NOT bitten by a zombie** is

$$P(- \mid \text{NO zombie bite}) = 0.95$$

This is known as as test **specificity**

So based on the two pieces of information above we can say the following about a test for infection coming back **positive** even though a person was **NOT bitten by a zombie**

$$\begin{aligned} P(+ \mid \text{NO zombie bite}) &= 1 - P(- \mid \text{NO zombie bite}) \\ &= 1 - 0.95 \\ &= 0.05 \end{aligned}$$

Remember that all test, especially medical tests have some error



Questions

Perspective of a person administering the test

Given a person was bitten by a zombie, what is the probability that their test is positive?

$$P(+ \mid \text{zombie bite}) = 0.90$$

Perspective of a person who is not infected

Given that their test is positive, what is the probability that person was bitten by a zombie?

$$P(\text{zombie bite} \mid +) = ?$$



Assumptions

Assume that for now zombies are pretty rare and represent only 0.1% of the population. This means that

$$P(\text{zombie bite}) = 0.001$$

This is known as a *prior*



The Math

1. From Bayes' theorem we have

$$P(\text{zombie bite} \mid +) = \frac{P(\text{zombie bite and } +)}{P(+)}$$

The numerator

$$\begin{aligned} P(\text{zombie bite and } +) &= P(\text{zombie bite}) \cdot P(+ \mid \text{zombie bite}) \\ &= 0.001 \cdot 0.90 \\ &= 0.0009 \end{aligned}$$

This follows

$$P(A \text{ and } B) = P(A) \cdot P(B \mid A)$$



The denominator

$$\begin{aligned}P(+) &= P(\text{zombie bite and } +) + P(+ \mid \text{NO zombie bite}) \cdot (1 - P(\text{zombie bite})) \\&= 0.0009 + 0.05 \cdot (1 - 0.0001) \\&= 0.5085\end{aligned}$$

The above follows

$$P(B) = P(A \text{ and } B) + P(B \mid \neg A) \cdot (1 - P(A))$$

2. We can take the previous values into Bayes' theorem to get

$$P(\text{zombie bite} \mid +) = \frac{0.0009}{0.05085} \\ = 0.02$$



implying that given a positive test result, there is only a 2% chance that they were bitten by a zombie



Does this seem counterintuitive?

You are definitely not alone, but this silly example is indicative a pretty important result because it parallels the nature of many realistic testing contexts such as HIV and DNA testing, criminal profiling, and even your standard run of the mill statistical significance testing

Now for the time being, forget about the result that you may be dismissing as garbage and try to make sense think about why the probability was so low. After you come to conclusion or get stuck, move on



Considerations

Whenever the condition of interest is very rare, having a test that finds all the true cases is still no guarantee that a positive result carries much information at all

Remember that zombies are pretty rare, only making up 0.1% of the population. As of this writing, there are about 7.93 billion people in the world. While 8 million zombies appears to be a lot, consider that in this scenario they would be spread out over the globe. Even if you restrict that number to the United States, the country alone has a population of about 333 million people making zombies about 2.4% of the total count of "people". or to put it another way, as of 2021 the number of zombies could almost equal the state population of Virginia

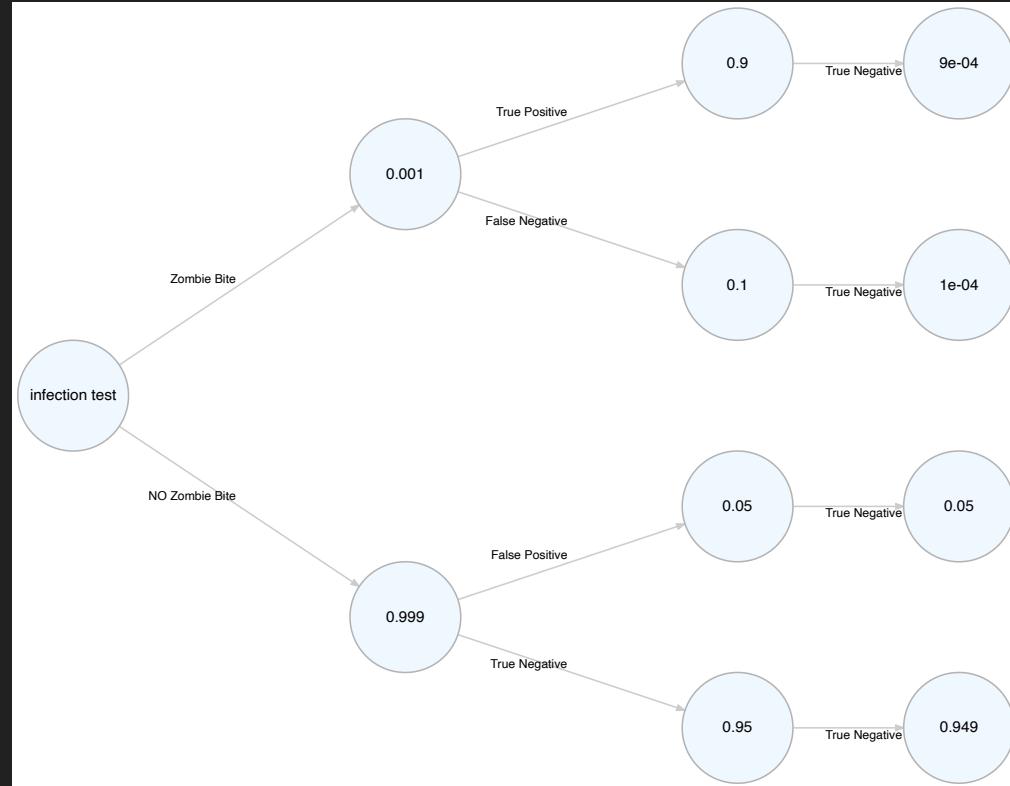
But why?

Well the reason may also seem counterintuitive but here you go:

most positive results are false positives, even when all the true positives are detected correctly

Now through testing and larger samples, if the prior probability changes then there is a good chance that the narrative above also changes with it

A Probability Tree of the Zombie "Problem"





Is it perfect?

No! If someone has a prior of 0 or 1, then it doesn't matter what you do, statistically speaking their mind is not going to change!

Also there are some issues with people's ability to use Bayesian approaches

BAYESIAN PRIOR CRISIS METHODS

Pro:

- Works well even under small data.
- Often very intuitive.

Con:

- Computationally costly in large data.
- A prior must be chosen.

Chris Albon



Did You Know?

In 2005 John Ioannidis, a professor at the Stanford School of Medicine, published a very famous and highly controversial paper named *Why Most Published Research Findings Are False*. He used Bayes' Theorem to establish a weak basis, though one did not exist prior to his publication. His calculation led to a broader argument and warnings that have been for the most part accepted by the broader research community



If You Torture Data Long Enough, It Will Confess to Anything

If you're interested in a broader review of how most research is false but is worth doing anyway, click on the icon below



One Last Thing About Bayes



We barely scratched the surface of Bayesian thinking. Unfortunately it is underutilized in the social sciences and education and that will become an issue as areas such as machine learning and the data sciences in general become prevalent. If you find this area interesting, I am happy to point you to resources that may be useful. In the meantime, you may wish to take a look at the papers below by hovering over the pills



Aczel et al. (2020)



Kruschke (2010)

Thats it!

If you have any questions, please reach out



This work is licensed under a
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License