

Table of Contents

- [References](#)
- [Alcholic Problem](#)
- [Clinic Test problem](#)
- [Genetic Defect Problem](#)
- [Cancer Problem](#)
- [Cookie problem](#)
- [M&M problem](#)
- [Elvis Presley Problem](#)
- [Blood type problem](#)
- [Smoker Problem](#)
- [Monty Hall Problem](#)

References

- <https://www.statisticshowto.datasciencecentral.com/bayes-theorem-problems/>
- [Alan Downey Bayes problems](#)
- [Bayes Theorem and some of the mysteries it has solved](#)

Alcholic Problem

You might be interested in finding out a patient's probability of having liver disease if they are an alcoholic. "Being an alcoholic" is the test (kind of like a litmus test) for liver disease.

- A could mean the event "Patient has liver disease." Past data tells you that 10% of patients entering your clinic have liver disease. $P(A) = 0.10$.
- B could mean the litmus test that "Patient is an alcoholic." Five percent of the clinic's patients are alcoholics. $P(B) = 0.05$.
- You might also know that among those patients diagnosed with liver disease, 7% are alcoholics. This is your $B|A$: the probability that a patient is alcoholic, given that they have liver disease, is 7%.

Bayes' theorem tells you:

$$P(A|B) = (0.07 * 0.1)/0.05 = 0.14$$

In other words, if the patient is an alcoholic, their chances of having liver disease is 0.14 (14%). This is a large increase from the 10% suggested by past data. But it's still unlikely that any particular patient has liver disease.

Clinic Test problem

Another way to look at the theorem is to say that one event follows another. Above I said "tests" and "events", but it's also legitimate to think of it as the "first event" that leads to the "second event." There's no one right way to do this: use the terminology that makes most sense to you.

In a particular pain clinic, 10% of patients are prescribed narcotic pain killers. Overall, five percent of the clinic's patients are addicted to narcotics (including pain killers and illegal substances). Out of all the people prescribed pain pills, 8% are addicts. If a patient is an addict, what is the probability that they will be prescribed pain pills?

Step 1: Figure out what your event "A" is from the question. That information is in the italicized part of this particular question. The event that happens first (A) is being prescribed pain pills. That's given as 10%.

Step 2: Figure out what your event "B" is from the question. That information is also in the italicized part of this particular question. Event B is being an addict. That's given as 5%.

Step 3: Figure out what the probability of event B (Step 2) given event A (Step 1). In other words, find what $P(B|A)$ is. We want to know "Given that people are prescribed pain pills, what's the probability they are an addict?" That is given in the question as 8%, or .8.

Step 4: Insert your answers from Steps 1, 2 and 3 into the formula and solve.

$$P(A|B) = P(B|A) * P(A) / P(B) = (0.08 * 0.1) / 0.05 = 0.16$$

The probability of an addict being prescribed pain pills is 0.16 (16%).

Solution

$$P(N) = 10/100$$

$$P(A) = 5/100$$

$$P(A|N) = 8/100$$

$$\begin{aligned} P(N|A) &= P(A|N) * P(N) / P(A) \\ &= 8 * 10 / 5 = 16\% \text{ answer} \end{aligned}$$

Genetic Defect Problem

1% of people have a certain genetic defect.

90% of tests for the gene detect the defect (true positives).

9.6% of the tests are false positives.

If a person gets a positive test result, what are the odds they actually have the genetic defect?

The first step into solving Bayes' theorem problems is to assign letters to events:

A = chance of having the faulty gene. That was given in the question as 1%.

That also means the probability of not having the gene ($\sim A$) is 99%.

X = A positive test result.

So:

$P(A|X)$ = Probability of having the gene given a positive test result.

$P(X|A)$ = Chance of a positive test result given that the person actually has the gene.

That was given in the question as 90%.

$p(X|\sim A)$ = Chance of a positive test if the person doesn't have the gene. That was given in the question as 9.6%

Now we have all of the information we need to put into the equation:

$$P(A|X) = (.9 * .01) / (.9 * .01 + .096 * .99) = 0.0865 \text{ (8.65\%)}$$

The probability of having the faulty gene on the test is 8.65%.

Concept

Bayes' theorem problems can be figured out without using the equation (although using the equation is probably simpler). But if you can't wrap your head around why the equation works (or what it's doing), here's the non-equation solution for the same problem in #1 (the genetic test problem) above.

Step 1: Find the probability of a true positive on the test. That equals people who actually have the defect (1%) * true positive results (90%) = .009.

Step 2: Find the probability of a false positive on the test. That equals people who don't have the defect (99%) * false positive results (9.6%) = .09504.

Step 3: Figure out the probability of getting a positive result on the test. That equals the chance of a true positive (Step 1) plus a false positive (Step 2) = .009 + .09504 = .10404.

Step 4: Find the probability of actually having the gene, given a positive result. Divide the chance of having a real, positive result (Step 1) by the chance of getting any kind of positive result (Step 3) = .009/.10404 = 0.0865 (8.65%).

Solution

```
P(D) = 1/100    defective gene
P(T|D) = 90/100 True Positive TP means if there are 100 defects, test will
find 90 of them.
P(T|~D) = 9.6/100 False Positive FP means test shows +ve for NOT defective
person.
```

```
There are 1 defective gene person. And, he becoming +ve is 90 %. This
means 90 * 1
```

```
There are 99 not defective gene person. And, he becoming +ve is 9.6%. This
means 9.6 * 99
```

$$P(D|T) = \frac{P(T|D) * P(D)}{P(T|D) * P(D) + P(T|\sim D) * P(\sim D)}$$

$$= \frac{90 * 1}{90 * 1 + 9.6 * 99}$$

Cancer Problem

Q. Given the following statistics, what is the probability that a woman has cancer if she has a positive mammogram result?

One percent of women over 50 have breast cancer.

Ninety percent of women who have breast cancer test positive on mammograms.

Eight percent of women will have false positives.

Step 1: Assign events to A or X. You want to know what a woman's probability of having cancer is, given a positive mammogram. For this problem, actually having cancer is A and a positive test result is X.

Step 2: List out the parts of the equation

(this makes it easier to work the actual equation):

$$P(A)=0.01$$

$$P(\sim A)=0.99$$

$$P(X|A)=0.9$$

$$P(X|\sim A)=0.08$$

Step 3: Insert the parts into the equation and solve. Note that as this is a medical test, we're using the form of the equation from example #2:

$$(0.9 * 0.01) / ((0.9 * 0.01) + (0.08 * 0.99)) = 0.10.$$

The probability of a woman having cancer, given a positive test result, is 10%.

Aliter:

The main difference with this form of the equation is that it uses the probability terms intersection(\cap) and compliment (c). Think of it as shorthand: it's the same equation, written in a different way.

In order to find the probabilities on the right side of this equation, use the multiplication rule:

$$P(B \cap A) = P(B) * P(A|B)$$

The two sides of the equation are equivalent, and $P(B) * P(A|B)$ is what we were using when we solved the numerator in the problem above.

$$P(B) * P(A|B) = 0.01 * 0.9 = 0.009$$

For the denominator, we have $P(Bc \cap A)$ as part of the equation. This can be (equivalently) rewritten as $P(Bc)P(A|Bc)$. This gives us:

$$P(Bc)P(A|Bc) = 0.99 * 0.08 = 0.0792.$$

Inserting those two solutions into the formula, we get:

$$0.009 / (0.009 + 0.0792) = 10\%.$$

Cookie problem

The first one is a warm-up problem. I got it from Wikipedia (but it's no longer there): Suppose there are two full bowls of cookies. Bowl #1 has 10 chocolate chip and 30 plain cookies, while bowl #2 has 20 of each. Our friend Fred picks a bowl at random, and then picks a cookie at random. We may assume there is no reason to believe Fred treats one bowl differently from another, likewise for the cookies. The cookie turns out to be a plain one. How probable is it that Fred picked it out of Bowl #1? First the hypotheses: A: the cookie came from Bowl #1 B: the cookie came from Bowl #2

And the priors: $P(A) = P(B) = 1/2$

The evidence: E: the cookie is plain

And the likelihoods: $P(E|A)$ = prob of a plain cookie from Bowl #1 = $3/4$ $P(E|B)$ = prob of a plain cookie from Bowl #2 = $1/2$

Plug in Bayes's theorem and get $P(A|E) = 3/5$

You might notice that when the priors are equal they drop out of the BT equation, so you can often skip a step.

M&M problem

This one is also an urn problem, but a little trickier. The blue M&M was introduced in 1995. Before then, the color mix in a bag of plain M&Ms was (30% Brown, 20% Yellow, 20% Red, 10% Green, 10% Orange, 10% Tan). Afterward it was (24% Blue, 20% Green, 16% Orange, 14% Yellow, 13% Red, 13% Brown). A friend of mine has two bags of M&Ms, and he tells me that one is from 1994 and one from 1996. He won't tell me which is which, but he gives me one M&M from each bag. One is yellow and one is green. What is the probability that the yellow M&M came from the 1994 bag? Hypotheses: A: Bag #1 from 1994 and Bag #2 from 1996 B: Bag #2 from 1994 and Bag #1 from 1996

Again, $P(A) = P(B) = 1/2$.

The evidence is: E: yellow from Bag #1, green from Bag #2

We get the likelihoods by multiplying the probabilities for the two M&M:

$$P(E|A) = (0.2)(0.2) \quad P(E|B) = (0.1)(0.14)$$

For example, $P(E|B)$ is the probability of a yellow M&M in 1996 (0.14) times the probability of a green M&M in 1994 (0.1).

Plugging the likelihoods and the priors into Bayes's theorem, we get $P(A|E) = 40 / 54 \sim 0.74$

By introducing the terms Bag #1 and Bag #2, rather than "the bag the yellow M&M came from" and "the bag the green came from," I avoided the part of this problem that can be tricky: keeping the hypotheses and the evidence straight.

Elvis Presley Problem

This one is from one of my favorite books, David MacKay's Information Theory, Inference, and Learning Algorithms: Elvis Presley had a twin brother who died at birth. What is the probability that Elvis was an identical twin? To answer this one, you need some background information: According to the Wikipedia article on twins: Twins are estimated to be approximately 1.9% of the world population, with monozygotic twins making up 0.2% of the total---and 8% of all twins."

There are several ways to set up this problem; I think the easiest is to think about twin birth events, rather than individual twins, and to take the fact that Elvis was a twin as background information.

So the hypotheses are A: Elvis's birth event was an identical birth event B: Elvis's birth event was a fraternal twin event

If identical twins are 8% of all twins, then identical birth events are 8% of all twin birth events, so the priors are

$$P(A) = 8\% \quad P(B) = 92\%$$

The relevant evidence is E: Elvis's twin was male

So the likelihoods are $P(E|A) = 1$ $P(E|B) = 1/2$

Because identical twins are necessarily the same sex, but fraternal twins are equally likely to be opposite sex (or, at least, I assume so). So

$$P(A|E) = 8/54 \sim 0.15.$$

The tricky part of this one is realizing that the sex of the twin provides relevant information!

Blood type problem

Also from MacKay's book: Two people have left traces of their own blood at the scene of a crime. A suspect, Oliver, is tested and found to have type O blood. The blood groups of the two traces are found to be of type O (a common type in the local population, having frequency 60%) and of type AB (a rare type, with frequency 1%). Do these data (the blood types found at the scene) give evidence in favour [sic] of the proposition that Oliver was one of the two people whose blood was found at the scene? For this problem, we are not asked for a posterior probability; rather we are asked whether the evidence is incriminating. This depends on the likelihood ratio, but not the priors.

The hypotheses are X: Oliver is one of the people whose blood was found Y: Oliver is not one of the people whose blood was found

The evidence is E: two blood samples, one O and one AB

We don't need priors, so we'll jump to the likelihoods. If X is true, then Oliver accounts for the O blood, so we just have to account for the AB sample:

$$P(E|X) = 0.01$$

If Y is true, then we assume the two samples are drawn from the general population at random. The chance of getting one O and one AB is

$$P(E|Y) = 2(0.6)(0.01) = 0.012$$

Notice that there is a factor of two here because there are two permutations that yield E.

So the evidence is slightly more likely under Y, which means that it is actually exculpatory! This problem is a nice reminder that evidence that is consistent with a hypothesis does not necessarily support the hypothesis.

Smoker Problem

I like this problem because it doesn't provide all of the information. You have to figure out what information is needed and go find it. According to the CDC, ``Compared to nonsmokers, men who smoke are about 23 times more likely to develop lung cancer and women who smoke are about 13 times more likely.'' If you learn that a woman has been diagnosed with lung cancer, and you know nothing else about her, what is the probability that she is a smoker? I find it helpful to draw a tree:

If y is the fraction of women who smoke, and x is the fraction of nonsmokers who get lung cancer, the number of smokers who get cancer is proportional to $13xy$, and the number of nonsmokers who get lung cancer is proportional to $x(1-y)$.

Of all women who get lung cancer, the fraction who smoke is $13xy / (13xy + x(1-y))$.

The x 's cancel, so it turns out that we don't actually need to know the absolute risk of lung cancer, just the relative risk. But we do need to know y , the fraction of women who smoke. According to the CDC, y was 17.9% in 2009. So we just have to compute

$$13y / (13y + 1-y) \sim 74\%$$

This is higher than many people guess.

Monty Hall Problem

6. Next, a mandatory Monty Hall Problem. First, here's the general description of the scenario, from Wikipedia: Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say Door A but the door is not opened, and the host, who knows what's behind the doors, opens Door B, which has a goat. He then says to you, "Do you want to pick Door C?" Is it to your advantage to switch your choice? The answer depends on the behavior of the host when the car is behind Door A. In this case the host can open either B or C. Suppose he chooses B with probability p and C otherwise. What is the probability that the car is behind Door A (as a function of p)?

The hypotheses are A: the car is behind Door A B: the car is behind Door B C: the car is behind Door C

And the priors are $P(A) = P(B) = P(C) = 1/3$

The likelihoods are $P(E|A) = p$, because in this case Monty has a choice and chooses B with probability p , $P(E|B) = 0$, because if the car were behind B, Monty would not have opened B, and $P(E|C) = 1$, because in this case Monty has no choice.

Applying Bayes's Theorem, $P(A|E) = p / (1+p)$

In the canonical scenario, $p=1/2$, so $P(A|E) = 1/3$, which is the canonical solution. If $p=0$, $P(A|E) = 0$, so you can switch and win every time (when Monty opens B, that it). If $p=1$, $P(A|E) = 1/2$, so in that case it doesn't matter whether you stick or switch.

When Monty opens C, $P(A|E) = (1-p) / (2-p)$

[Correction: the answer in this case is not $(1-p) / (1+p)$, which what I wrote in a previous version of this article. Sorry!].