$(.10)(.995) + (.02)(.005)$

$= .0996$ or $\boxed{9.96\%}$

~~P(S) = .05~~ P(P|S~) =
~~P(S~) = .99~~ P(P~|S~) =
P(S) = .005
P(S~) = .995
P(P|S) = .98
P(P~|S) = .02

Assignments

- Now it's time to compute some probabilities. Keep track of your work in a Google document or markdown file that you can share with your mentor.

Drill set 1

1. Calculate the probability of flipping a balanced coin four times and getting each pattern: HTTH, HHHH and TTHH.
2. If a list of people has 24 women and 21 men, then the probability of choosing a man from the list is 21/45. What is the probability of not choosing a man?
3. The probability that Bernice will travel by plane sometime in the next year is 10%. The probability of a plane crash at any time is .005%. What is the probability that Bernice will be in a plane crash sometime in the next year?
4. A data scientist wants to study the behavior of users on the company website. Each time a user clicks on a link on the website, there is a 5% chance that the user will be asked to complete a short survey about their behavior on the website. The data scientist uses the survey data to conclude that, on average, users spend 15 minutes surfing the company website before moving on to other things. What is wrong with this conclusion?

When you're done, the solution for this drillset can be found here.

Drill set 2

Now it's time to use Bayes' rule to compute some conditional probabilities. First look over the numbers and estimate each of the four probabilities, using your intuition. Then, calculate the probabilities using Bayes' rule. Keep track of your work in a Google document or markdown file that you can share with your mentor.

A diagnostic test has a 98% probability of giving a positive result when applied to a person suffering from Thripshaw's Disease, and 10% probability of giving a (false) positive when applied to a non-sufferer. It is estimated that 0.5 % of the population are sufferers. Suppose that the test is now administered to a person whose disease status is unknown. Calculate the probability that the test will:

1. Be positive → P(P) = P(P|S)P(S) + P(P|S~)P(S~) = P(P∩S) + P(P∩S~)
2. Correctly diagnose a sufferer of Thripshaw's - 98%. → is this P(P|S) or P(S|P)?
3. Correctly identify a non-sufferer of Thripshaw's - 90%. → is this P(~P|S~) or P(S~|P~)?
4. Misclassify the person => P(P∩S~) + P(P~∩S) = P(P|S~)P(S~) + P(P~|S)P(S)

Were your intuitions on the mark, or way off? If your statistical intuition is leading you astray, you aren't alone. According to Nobel-prize winning Daniel Kahneman, humans simply are not good intuitive statisticians. That fact has two strong implications for you as you prepare for a career in data science:

1. Just because your statistical intuition is wrong does not mean you're a bad statistician. We're all in this boat. Don't give up on yourself. And,
2. Just because you easily intuit the answer to a statistical question does not mean you're right. Don't trust your intuition. Check your work.

Discuss your answer to each of these questions with your mentor. If you want more Bayes practice, try these exercises.

After giving it a try yourself, you can find a solution here.

DS1.1 - $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \boxed{1/16}$ for all of them

DS1.2 - $\boxed{24/45} = 1 - P(M) = 1 - \frac{21}{45} = P(W) = P(M\sim)$

DS1.3 - $P(T) = .10 \quad P(L) = .00005$

$P(T \cap L) = P(T) \cdot P(L) = .10 \times .00005 = .000005$ or $\boxed{.0005\%}$

DS1.4 - The problem with this is that users who click on more links are likely to spend more time on the website, which means these users will be prompted for the survey more often - making for a biased sample.

DS2.1 - $P(p)$ = probability of positive result

$P(S)$ = probability of suffering patient

Bayes: $P(A|B) = \dfrac{P(B|A) \cdot P(A)}{P(B)}$

$P(p|s) = .98$

$P(p|s\sim) = .10$

$P(p\sim|s\sim) = \boxed{.9}$

$P(s) = .005$

$P(s\sim) = .995$

$$P(P) = P(p|s) \cdot p(s) + P(p|s\sim) \cdot P(s\sim)$$
$$= (.98) \cdot (.005) + (.10)(.995)$$
$$= .1044 \text{ or } \boxed{10.44\%}$$

DS2.2 $\boxed{98\%} = P(p|s)$ No $\Rightarrow P(s|p) = \dfrac{P(P|s) \cdot P(s)}{P(p)} = \dfrac{(.98)(.005)}{.1044} = \boxed{4.69\%}$

DS2.3 $P(p\sim|s\sim) = 1 - .10 = \boxed{90\%}$ NOPE

~~DS2.3~~ $P(s\sim|p\sim) = \dfrac{P(p\sim|s\sim)P(s\sim)}{P(p\sim)} = \dfrac{(.9)(.995)}{1-.1044} = \boxed{99.99\%}$

DS2.4 $\Rightarrow P(s|p\sim) + P(s\sim|p)$

$\cdot P(s)$, $\cdot P(s\sim)$,

$$P(s|p\sim) = \frac{P(p\sim|s) \cdot p(s)}{P(p\sim)} = \frac{(.02)(.005)}{1-.1044} = .000112$$

$$P(s\sim|p) = \frac{P(p|s\sim) \cdot p(s\sim)}{P(p)} = \frac{(.10)(.995)}{.1044} = .953065$$

~~P(P|s~~ $P(s|p)$.

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

$1 - \left(P(s \cap P) + P(s\sim \cap P\sim)\right) = 1 - (P(s)P(p|s) + P(s\sim)P(p\sim|s\sim))$

$P(P \cap s) + P(P\sim \cap s\sim)$

$1 - ((.005)(.98) + (.995)(.9))$

$P(P)P(s|P) + P(P\sim)P(s\sim|P\sim)$

$1 - (.0049 + .8955)$

$(.1044)(.0469) + (.8956)(.9999)$

$.0996$ or $\boxed{9.96\%}$

1. Aircraft emergency locator transmitter

An aircraft emergency locator transmitter (ELT) is a device designed to transmit a signal in the case of a crash. The Altigauge Manufacturing Company makes 80% of the ELTs, the Bryant Company makes 15% of them, and the Chartair Company makes the other 5%. The ELTs made by Altigauge have a 4% rate of defects, the Bryant ELTs have a 6% rate of defects, and the Chartair ELTs have a 9% rate of defects (which helps to explain why Chartair has the lowest market share).

$P(A)=.8 \quad P(B)=.15 \quad P(C)=.05 \quad P(D|A)=.04 \quad P(D|B)=.06 \quad P(D|C)=.0$

If an ELT is randomly selected from the general population of all ELTs, find the probability that it was made by the Altigauge Manufacturing Company. $\boxed{80\%}$ $P(A)=.80$

If a randomly selected ELT is then tested and is found to be defective, find the probability that it was made by the Altigauge Manufacturing Company.

$$P(A|D)=\frac{P(D|A)P(A)}{P(D)}=\frac{(.04)(.8)}{.0455}=\boxed{70.33\%}$$

$$P(D)=P(A\cap D)+P(B\cap D)+P(C\cap D)$$
$$=P(D|A)P(A)+P(D|B)P(B)+P(D|C)P(C)$$
$$=(.04)(.8)+(.06)(.15)+(.09)(.05)=\underline{0.0455}$$

2. Ingrowing toenail

You go to see the doctor about an ingrowing toenail. The doctor selects you at random to have a blood test for swine flu, which for the purposes of this exercise we will say is currently suspected to affect 1 in 10,000 people in Australia. The test is 99% accurate, in the sense that the probability of a false positive is 1%. The probability of a false negative is zero. You test positive. What is the new probability that you have swine flu?

$P(P\sim|S)=0$

Now imagine that you went to a friend's wedding in Mexico recently, and (for the purposes of this exercise) it is know that 1 in 200 people who visited Mexico recently come back with swine flu. Given the same test result as above, what should your revised estimate be for the probability you have the disease?

over →

$P(S)=1/10000 \quad \cancel{P(P|S)=.99} \quad P(P|S\sim)=.01$

$P(S\sim)=9999/10000$

$\cancel{P(P\sim|S\sim)=0}$

$$P(S|P)=\frac{P(P|S)P(S)}{P(P)}$$

$$P(P)=P(P\cap S)+P(P\cap S\sim)$$
$$=P(P|S)P(S)+P(P|S\sim)P(S\sim)$$
$$=.010098$$

$P(P|S)=1$ not .99

$$=\frac{(.99)(1/10000)}{.010098}=.0098 \text{ or } \boxed{.98\%} \text{ close}$$

$P(P|S)=1-P(P\sim|S)=1-0=1$

$P(P|S)=1$

$P(P\sim|S\sim)=.99$

$P(P\sim|S)=0$

$P(S) = 1/200 \qquad P(S\sim) = 199/200$

$\boxed{P(P|S) = 1 \text{ not } .99}$

$$P(P) = P(P|S)P(S) + P(P|S\sim)P(S\sim)$$
$$= (.99)(.005) + (.01)(.995)$$
$$= .0149$$

$$P(S|P) = \frac{P(P|S)P(S)}{P(P)} = \frac{.00495}{.0149} = \boxed{33.22\%}$$

Close

3. Trip to Las Vegas

Imagine that, while in Mexico, you also took a side trip to Las Vegas, to pay homage to the TV show CSI. Late one night in a bar you meet a guy who claims to know that in the casino at the Tropicana there are two sorts of slot machines: one that pays out 10% of the time, and one that pays out 20% of the time [note these numbers may not be very realistic]. The two types of machines are coloured red and blue. The only problem is, the guy is so drunk he can't quite remember which colour corresponds to which kind of machine. Unfortunately, that night the guy becomes the vic in the next CSI episode, so you are unable to ask him again when he's sober

Next day you go to the Tropicana to find out more. You find a red and a blue machine side by side. You toss a coin to decide which machine to try first; based on this you then put the coin into the red machine. It doesn't pay out. How should you update your estimate of the probability that this is the machine you're interested in? What if it had paid out - what would be your new estimate then?

$P(R \text{ or } B)$

$P(P\sim | R \text{ or } B) = .8$

$P(P\sim | R \text{ or } B) = .9$

$P(P | R \text{ or } B) = .2$

$P(P | R \text{ or } B) = .1$

(I'd be more confident with more data...)

If it doesn't pay out, it is more likely the 10% machine, if it does pay out it is more likely the 20% machine.

4. Genetic defects

1% of people have a certain genetic defect. $P(D) = .01$
$P(D\sim) = .99$

90% of tests for the gene detect the defect (true positives). $P(P|D) = .9$
$P(P\sim|D) = .1$

9.6% of the tests are false positives. $P(P|D\sim) = .096$
$P(P\sim|D\sim) = .904$

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

$$P(D|P) = \frac{P(P|D)P(D)}{P(P)} = \frac{(.9)(.01)}{P(P|D\sim)P(D\sim) + P(P|D)P(D)}$$

$$= \frac{.009}{.10404} = .0865$$

or

$\boxed{8.65\%}$

5. A test for cancer

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

$P(C|P)?$

she must be over 50 then...

One percent of women over 50 have breast cancer. – does this help?

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

Eight percent of women will have false positives.

~~$P(C) = 1/50$~~

$P(C) = 1/100$

$P(\sim C) = 99/100$

$P(P|C) = .9$

$P(P\sim|C) = .1$

$P(P|C\sim) = .08$

$P(P\sim|C\sim) = .92$

$$P(C|P) = \frac{P(P|C)P(C)}{P(P|C)P(C) + P(P|C\sim)P(C\sim)}$$

$$= \cancel{\frac{.018}{.018 + .0784}} = \cancel{\boxed{18.67\%}}$$

$$\frac{(.9)(1/100)}{(.9)(1/100) + (.08)(99/100)} = \frac{.009}{.0882} = \boxed{10.204\%}$$