Updating Your Beliefs with Bayes (e.g. how it can help you see whatâ€™s behind you)

Crash Course: Statistics
https://youtube.com/watch?v=oZCskBpHWyk
https://nerdfighteria.info/v/oZCskBpHWyk

Hi, I'm Adriene Hill, and Welcome back to Crash Course Statistics.

We ended the last episode by talking about Conditional Probabilities which helped us find the probability of one event, given that a second event had already happened. But now I want to give you a better idea of why this is true and how this formula--with a few small tweaks--has revolutionized the field of statistics.

INTRO In general terms, Conditional Probability says that the probability of an event, B, given that event A has already happened, is the probability of A and B happening together, Divided by the probability of A happening - that's the general formula, but let's give you a concrete example so we can visualize it. Here's a Venn Diagram of two events, An Email containing the words "Nigerian Prince" and an Email being Spam. So I get an email that has the words "Nigerian Prince" in it, and I want to know what the probability is that this email is Spam, given that I already know the email contains the words "Nigerian Prince." This is the equation.

Alright, let's take this part a little. On the Venn Diagram, I can represent the fact that I know the words "Nigerian Prince" already happened by only looking at the events where Nigerian Prince occurs, so just this circle. Now inside this circle I have two areas, areas where the email is spam, and areas where it's not.

According to our formula, the probability of spam given Nigerian Prince is the probability of spam AND Nigerian Prince which is this region... where they overlap…divided by Probability of Nigerian Prince which is the whole circle that we're looking at. Now...if we want to know the proportion of times when an email is Spam given that we already know it has the words "Nigerian Prince", we need to look at how much of the whole Nigerian Prince circle that the region with both Spam and Nigerian Prince covers. And actually, some email servers use a slightly more complex version of this example to filter spam.

These filters are called Naive Bayes filters, and thanks to them, you don't have to worry about seeing the desperate pleas of a surprisingly large number of Nigerian Princes. The Bayes in Naive Bayes comes from the Reverend Thomas Bayes, a Presbyterian minister who broke up his days of prayer, with math. His largest contribution to the field of math and statistics is a slightly expanded version of our conditional probability formula.

Bayes Theorem states that: The probability of B given A, is equal to the Probability of A given B times the Probability of B all divided by the Probability of A You can see that this is just one step away from our conditional probability formula. The only change is in the numerator where P(A and B) is replaced with P(A B)P(B).

While the math of this equality is more than we'll go into here, you can see with some venn-diagram-algebra why this is the case. In this form, the equation is known as Bayes' Theorem, and it has inspired a strong movement in both the statistics and science worlds. Just like with your emails, Bayes Theorem allows us to figure out the probability that you have a piece of spam on your hands using information that we already have, the presence of the words "Nigerian Prince".

We can also compare that probability to the probability that you just got a perfectly valid email about Nigerian Princes. If you just tried to guess your odds of an email being spam based on the rate of spam to non-spam email, you'd be missing some pretty useful information--the actual words in the email! Bayesian statistics is all about UPDATING your beliefs based on new information.

When you receive an email, you don't necessarily think it's spam, but once you see the word Nigerian you're suspicious. It may just be your Aunt Judy telling you what she saw on the news, but as soon as you see "Nigerian" and "Prince" together, you're pretty convinced that this is junkmail. Remember our Lady Tasting Tea example... where a woman claimed to have superior taste buds ...that allowed her to know--with one sip--whether tea or milk was poured into a cup first?

When you're watching this lady predict whether the tea or milk was poured first, each correct guess makes you believe her just a little bit more. A few correct guesses may not convince you, but each correct prediction is a little more evidence she has some weird super-tasting tea powers. Reverend Bayes described this idea of "updating" in a thought experiment.

Say that you're standing next to a pool table but you're faced away from it, so you can't see anything on it. You then have your friend randomly drop a ball onto the table, and this is a special, very even table, so the ball has an equal chance of landing anywhere on it. Your mission--is to guess how far to the right or left this ball is.

You have your friend drop another ball onto the table and report whether it's to the left or to the right of the original ball. The new ball is to the right of the original, so, we can update our belief about where the ball is. If the original is more towards the left, than most of the new balls will fall to the right of our original, just because there's more area there.

And the further to the left it is, the higher the ratio of new rights to lefts Since this new ball is to the right, that means there's a better chance that our original is more toward the left side of the table than the right, since there would be more "room" for the new ball to land. Each ball that lands to the right of the original is more evidence that our original is towards the left of the table. But, if we get a ball landing on the left of our original, then we know the original is not at the very left edge.

Again, Each new piece of information allows us to change our beliefs about the location of the ball, and changing beliefs is what Bayesian statistics is all about. Outside thought experiments, Bayesian Statistics is being used in many different ways, from comparing treatments in medical trials, to helping robots learn language. It's being used by cancer researchers, ecologists, and physicists.

And this method of thinking about statistics...updating existing information with what's come before...may be different from the logic of some of the statistical tests that you've heard of--like the t-test. Those Frequentist statistics can sometimes be more like probability done in a vacuum. Less reliant on prior knowledge.

When the math of probability gets hard to wrap your head around, we can use simulations to help see these rules in action. Simulations take rules and create a pretend universe that follows those rules. Let's say you're the boss of a company, and you receive news that one of your employees, Joe, has failed a drug test.

It's hard to believe. You remember seeing this thing on YouTube that told you how to figure out the probability that Joe really is on drugs given that he got a positive test. You can't remember exactly what the formula is...but you could always run a simulation.

Simulations are nice, because we can just tell our computer some rules, and it will randomly generate data based on those rules. For example, we can tell it the base rate of people in our state that are on drugs, the sensitivity (how many true positives we get) of the drug test... and specificity (how many true negatives we get). Then we ask our computer to generate 10,000 simulated people and tell us what percent of the time people with positive drug tests were actually on drugs.

If the drug Joe tested positive for--in this case Glitterstim--is only used by about 5% of the population, and the test for Glitterstim has a 90% sensitivity and 95% specificity, I can plug that in and ask the computer to simulate 10,000 people according to these rules. And when we ran this simulation, only 49.2% of the people who tested positive were actually using Glitterstim. So I should probably give Joe another chance...or another test.

And if I did the math, I'd see that 49.2% is pretty close since the theoretical answer is around 48.6%. Simulations can help reveal truths about probability, even without formulas. They're a great way to demonstrate probability and create intuition that can stand alone or build on top of more mathematical approaches to probability.

Let's use one to demonstrate an important concept in probability that makes it possible to use samples of data to make inferences about a population: the Law of Large Numbers. In fact we were secretly relying on it when we used empirical probabilities--like how many times I got tails when flipping a coin 10 times--to estimate theoretical probabilities--like the true probability of getting tails. In its weak form, Law of Large Numbers tells us that as our samples of data get bigger and bigger, our sample mean will be 'arbitrarily' close to the true population mean.

Before we go into more detail, let's see a simulation and if you want to follow along or run it on your own - instructions are in the description below. In this simulation we're picking values from a new intelligence test--from the normal distribution, that has a mean of 50 and a standard deviation of 20. When you have a very small sample size, say 2, your sample means are all over the place.

You can see that pretty much anything goes, we see means between 5 and 95. And this makes sense, when we only have two data points in our sample, it's not that unlikely that we get two really small numbers, or two pretty big numbers, which is why we see both low and high sample means. Though we can tell that a lot of the means are around the true mean of 50 because the histogram is the tallest at values around 50.

But once we increase the sample size, even to just 100 values, you can see that the sample means are mostly around the real mean of 50. In fact all of the sample means are within 10 units of the true population mean. And when we go up to 1000, just about every sample mean is very very close to the true mean.

And when you run this simulation over and over, you'll see pretty similar results. The neat thing is that the Law of Large numbers applies to almost any distribution as long as the distribution doesn't have an infinite variance. Take the uniform distribution which looks like a rectangle.

Imagine a 100-sided die, every single value is equally probable. Even the sample means that are selected from a uniform distribution get closer and closer to the true mean of 50.. The law of large numbers is the evidence we need to feel confident that the mean of the samples we analyze is a pretty good guess for the true population mean.

And the bigger our samples are, the better we think the guess is! This property allows us to make guesses about populations, based on samples. It also explains why casinos make money in the long run over hundreds of thousands of payouts and losses, even if the experience of each person varies a lot.

The casino looks at a huge sample--every single bet and payout--whereas your sample as an individual is smaller, and therefore less likely to be representative. Each of these concepts can help us another way ...another way to look at the data around us. The Bayesian framework shows us that every event or data point can and should "update" your beliefs but it doesn't mean you need to completely change your mind.

And simulations allow us to build upon these observations when the underlying mechanics aren't so clear. We are continuously accumulating evidence and modifying our beliefs everyday, adding today's events to our conception of how the world works. And hey, maybe one day we'll all start sincerely emailing each other about Nigerian Princes.

Then we're gonna have to do some belief-updating. Thanks for watching. I'll see you next time.