

# module\_prob1\_lecture6

Fri, 1/21 12:31PM 18:13

## SUMMARY KEYWORDS

probability, .5, proportion, beliefs, student, chance, clicker question, toss, personal beliefs, rules, frequentist, head, quarantine, coin, assignments, interpretation, statcrunch, talk, bouncing, rolling

## SPEAKERS

Sumon Majumdar

---

### Sumon Majumdar 00:04

So in the last clip, I was talking about extending the addition rule to three sets. And it was given by this particular formula, which is brought for the probability of A union B, union C, right? So what's the probability of either A, or B, or C. So what I'm going to start out here is applying it to a particular example. And that will give you a hang of how to use use this rule. So let's, let's look at this example where there are 100 students in a class. And they could do projects A, B, or C. Right? So 30 students did a 40 day to be 20 days see, but there's some who also did both MB, some did B, and C, some DD, and C. And there's some really hard working students who did all three. So the question is, what's the probability that random student did at least one of the assignments, you cannot just simply add up 30 plus 40, plus 20. Because there are some students who did the same student who did B and C, the same student who might have done A and B, or may have done all three. So what we're really looking for is the probability of A union B, union C. So it's the probability that the student did, at least one of the assignments may be more so. So what we're going to make use of is obviously this particular formula. So if we're looking at this, right, so it's firstly the probability of A plus B plus C. So the probability of A so that's 30 over 100, right, so that's a probability of B is 40 over 100. C is 20 over 100. minus the probability of A intersection B, so those, these are the students who did both A and B. So that's this. So minus 10, over 100, minus those who did both B and C, that's 12, over 100. Minus those who did both ANC, that's this, that's 15 over 101, last term is left, this is two people who did all three, right, and that needs to be added in. So that's five over 100. So that's your probability, that random student did at least one of the assignments. So again, if we add everything up, right, so that's 30, 40, and 20, that's 90 minus 10, 12, that's 22. And 15, that's 37 plus five. So, so that's 95 minus 37, that's 58 over 100. Which is .42, sorry, .58. So this is the probability that a random student did at least one of the assignments, right, so 58% of the students here, did at least one of the assignments. So what I would like for you to do now is do a clicker question based on the same example. So yeah, please stop the video at this point, and attempt the clicker question. And once you're done again, resume the video. So hope you got the chance to do the clicker question, which was based on the same example. And what is asked is what the probability that random student did not submit any of the projects. So we don't really figured out what's the probability that the student did at least one of these projects in A B union C. So this looks at the complementary event, right? That the student didn't do any of them. So it's one minus this probability. So this we had already figured out as .58. So one

minus .58 is .42. Right? So the correct answer to the clicker question is P. So hope you're getting the hang of this idea of probability, at the end the rules behind probability, which help you compute the probability in a variety of different contexts. So let me so let me just change, tone a little bit. And talk a little bit about not so much about the particular rules, but what are the into what is the interpretation of probability? It's what does a number like .5, .6 or that? That ahead, if you if you toss a coin, what's the probability the probability that the head comes up is .5? What does that actually mean? So I want to talk a little bit about this interpretation of probability. Right? So one of the ideas and this is the classical idea of probability is that it gives you the, the relative chance or the relative frequency that something happens. So the probability that if you if you toss a coin and if the coin is unbiased, then the probability that a head comes up is half. That is point five. And the probability that a tail comes up is half, it's .5, right? So what does this mean? So the classical idea of probability is that if you repeat this many, many, many, many times, right? So if we toss this coin many, many, many, many times, then the fraction of times that the head comes up will be about half will be about point five, the fraction of times that the tail comes up will be about it doesn't mean that you cannot get 10 heads in a row, it doesn't mean that you may not get five tails, or six tails in a row. But if you do it, long, lot, lot, lot, lot. But then the proportion of times that you're going to get ahead is going to be about 50%. So let me show you. This is from a I did this in a in a website called statcrunch.com, where they have different applets, where you can play around with things like that. So this is where they where they have simulated the probability of head with a fair coin. And we have done 1000 tosses. Right? And this gives you the proportion of hence. Right? So see, this green line here that they have, this is the .5, right? So we know that the theoretical probability should be .5. Now if you do it, right, initial, right, so this is this blue line here, this gives the proportion of heads that's coming up in this tosses, as you do that same the initial This is the first 200 tosses that the 400 600 and so on. So in the initial cases, right, this number is bouncing around, right? So probably it started with with the tail, right, then the next one became a head, then you got a series of heads and so on, right? And this number is bouncing around. But as you are going more and more right past 200 400 600 800, it's still moving around a little bit. But it is slowly slowly coming closer and closer to point five. But there's still some bouncing around if you if you blow this up, right? So you can see that there is still some bouncing around. It's not exactly .5, you probably got a whole series of tails. Maybe you've got a whole series of heads, right? But as the number of coin tosses is becoming bigger and bigger, right, that proportion, is slowly going to point five. And this is the classical idea of, of probability. Same thing if you again, go to StatCrunch and do the rolling a die, what's the probability of a six coming up? Right, so you're going to die. So the probability of a six coming up is  $1/6$ . So this is like, point, what is it? .1667? Right? Again, if you roll this die 1000 times, right? Again, initially see, something else came up, right? You got some, some sixes, right? And this blue line is bouncing around. But as you're going more and more and more 200 400 600 800, this number is moving closer and closer to this .1667, the theoretical probability. So that's the idea of, of, that's the interpretation of probability, something which is frequentist, in the sense that if you do this event, a lot of times, the proportion of times that something will happen will be its probability. The same thing, if you go to a casino, right? In Las Vegas, casinos, the house is only it's supposed to be almost fair, right? The house maybe has a 1% higher chance of winning, then you but you go there, and you put money in the slot machine, maybe 10 times and you get nothing, then you say, Oh, these guys must be bias. But that's not true. Because 10 may be a small number, maybe if you do it a million times red, then you will see that the proportion of winnings of the house versus you is going to be it's about 51% for the house 49% for you. So that's the frequentist classical definition of probability. And this works well with flipping a coin rolling a die, things like this. The problem is that, especially in the social sciences, a lot of times you have to consider probability and think about events, where it's difficult to replicate it so many times. So for example, if so people take courses, and you will

always hear this thing, oh, I think I'm going to fail this course, or I'm going to do great in this course, or I'm going to get an A in this course. Right now. So my chances of getting an A in this course is 40%. Right? But the See, you don't take the course many, many, many, many times you take that course, only once. Right? So what does that mean? Right? So what does the 40% mean? Right? So if you took this course, many, many, many, many, many, many times, there's the probability that you will get an A 40%. Right? And even that's not true. Because see, if you take the same course, many, many times, we're going to get very, very good at it. And you're almost surely going to get an A not just 40% of the time. Right? So a lot of times we we in real life, right? So probability is not just frequent is it's also a personal belief. Right? So another example, Will Trump get reelected? Right? So we don't know, because it's not that Trump has stood for election many, many, many times. And we can know what's the probability that he's going to win or lose? Similarly, is his rules for international air travel going to change? Is it going to become back to pre pandemic times by the time December comes around? We don't know. Because that does hasn't happened a lot of times in the past, so we don't know what to base on. But people do have beliefs on that. Right? So many of these are personal beliefs. And the beliefs could differ across individuals. Some people may think, yeah, I think the very, very 80% chance that this quarantine restrictions for a travel, those are going to go away. Some people think No, I don't think so it's only a 10% chance. There's nothing wrong with them, because each of them have their personal beliefs. But those are also probabilities. Right? And the things but and what we try to do in formalizing this is we try to keep the same ideas that we did for whether tossing a coin rolling a die, even with personal beliefs, that the probabilities here should be between zero and one. Right? The probability of the samples based that should be one. And the probability of two disjoint sets probability of A union B, where there is nothing common between A and B, that should be just probability of A plus probability of B. So, so long as your beliefs satisfy these three axioms, these are valid probabilities, right? And these are as valid a probability as anything else. And a lot of times in the social sciences, we make decisions based on such beliefs. So for example, should I buy a ticket for a vacation to Mexico? Right? It depends on what my belief is about whether quarantine rules are going to change by December or not. Right? Same thing, if I make a decision about taking this course, or another course, it could be influenced by what I believe is my probability of doing well in this course versus another. Right? So there are at least these two interpretations of probability. And a lot of times, the people who we call so called experts, they try to build some kind of a model to put a little bit more structure on their beliefs. So especially like if you go to a stock investor, a specialist, right. So this is a person who has probably read out models, and has analyzed data and trying to get a better idea of what the chance of this stock going up or versus that stock going down. But at the end of the day, if part of it could be just his personal belief, right? So for example, one of the earlier examples we did was looking at what's the probability of it's snowing, or being very cold in Ottawa on on February 5th, and I looked at the past 10 years of data. And I was happy with that. But someone else may say, Look, you should have really looked at past 100 years of data. 10 years is nothing, right? Weather is very fickle. Someone else may have countered by saying, no, no, no, I think winters have changed so much weather has changed so much. You shouldn't even looking be looking at the past 10 years, you should have looked at just the past five years. Right? So even opinions could differ among experts. And why I say this is because these are all probabilities, but where they come from, right, that could depend on your personal judgment, your the data that you're looking at, and and your beliefs. So let me stop this clip here. And in the next clip, I'll be talking about independence and relationship between events.