PROFESSOR: Today's focus is probability and statistics. So let's start with probability. Let's look at probability for binary variables. What do you mean by a binary variable? It can take only two outcomes. So it can take only two values. For example, it could be 0 or 1, head or tail, on or off.

So we are going to call this variable A, for instance. So A could be H, or A is equal to T. But that could happen. That event could happen with a certain probability. So by that, I mean the probabilities, like we are expressing the belief that the particularly event could happen. So we could assign a value to that. That is the probability of A taking value H.

So here, the values of A and B-- sorry, here, the value of A can be either H or T, which means it has only two possible outcomes. That's why we call it a binary variable. However, P of A is equal to H can lie anywhere from 0 and 1, including 0 and 1.

AUDIENCE: They don't have to be even?

PROFESSOR: Sorry?

AUDIENCE: They don't have to be even?

PROFESSOR: Even?

AUDIENCE: Even chance, even probability, like the same.

PROFESSOR: Sorry, I didn't get your question.

AUDIENCE: Even though they're binary, don't you need be able to have the same probability?

PROFESSOR: OK, we'll look at that later. Like, this particular event can take a particular probability. And we'll look at that particular case later. But in general, a probability will always lie between 0 and 1. And it can take any value between 0 and 1 since the range it can take is continuous, sorry discrete.

However, the value the variable can take is going to be discrete. It can take only H or T. So that's why you call it a binary variable.

For example, take a deck of cards. Here, the value could be, for example if you consider only one particular suit, then it can be any one of those 13 values. So there, this variable is not binary. However, the probability of a particular event happening is always between 0 and 1.

Now, let's look at some probability, like what you asked earlier is whether they will be equal, whether the probably of head and tail can be equal. So let's represent the probability of A of H. This can be between 0 and 1. What is the probability of A not happening? So we call it by A bar. Given P of A, can you give me P of A bar?

AUDIENCE: 1 minus P of A.

PROFESSOR: 1 minus P of A. If there are two events happening, for example, you're throwing two coins, then we can consider their joint probabilities. So let's say we have a coin, A, and this coin, B. So this coin can take two values. And so this coin can take another two values. Sorry. We know A can take H with probability, say I assume it's unbiased, so it'll be 1/2. All these are going to be 1/2. What's the probability of HT?

So now, we are considering a joint event, P of A is equal to H and P of B is equal to T. So in probability, we represent it by something like this. P A- do you know what is that? P A intersection B, you want both events to happen. That will be P of A. And in this case, it's P of B. So we could simply say it's 1/4.

Why is this possible? It's because these two events are independent. The coin A getting head doesn't affect coin B getting a tail. So it doesn't have any influence. That's why these two events are independent. The dependent events are a bit

complex, to analyze. Let's skip them at the moment. So we know all these probabilities are going to be 1/4.

So we looked at a particular condition here. That is, A taking head and B taking tail. What about the condition, what about the case where either A or B takes a head? How can we represent that? So it will be something like A is equal to H or B is equal to H. Oh, probability at least 1, so by that, I can also represent something like this. OK, here, this is sufficient anyway.

So what are the possibility events? So these three events could give rise to this probability. It's better if you can represent this in a diagram. So let's go and represent this in a diagram. This is A and this is B getting, say, head. In one case, both can take head. That is, this particular condition, intersection we earlier looked at.

So what is this whole thing? That is, either A gets H or B gets H, which is this condition. We call it P of A union B. Ok.

Is there an efficient way of finding this rather than writing down all possible cases? Is there an efficient way of finding P of A union B? From high school maths, probably? No idea? OK. P of A union B is equal to P of A plus P of B minus P of A intersection B. Because if you consider P of A, you would have taken this full circle. When you take P of B, you would have taken this full circle. So which means you're counting this area twice. So here, we deduct it once. OK? Great. So this is the basics of the probability.

Now, actually we looked at two events, two joint events here. But we should have a formal way of looking at multiple events. So how can we do that? The first way is doing it by trees.

Let's say we represent the outcome of the first trial by a branch. We can represent the outcome of the second trial by another branch from these two previous branches. So this would be H HH HT TH TT. And we know this could happen with probability 1/2. So we know it's, again, 1/2, 1/2, 1/2, 1/2. So this is 1/4.

Suppose we want to do this for an outcome of throwing dice. Then, probably we would have 6 branches here. Which, again, forks into another 36 branches. So there should be another easier way. For that, we could use a second method call grid. We could simply put that in a diagram.

So this is the first trial. And this will be our second trial. So now, we can represent any possible outcome on this grid.

For example, can give you me an example where you throw the same number in both the trials? Then, what would be the layout of it in this grid? Throwing the same number in both the trials. Here's the first trial. This, the second. Then it would be the diagonal.

If you want to calculate the probability, do you know the probability is the ratio between the outcomes we expect over all possible outcomes? So here, we know there will be 6 instances in this highlighted area. Compare that, 36 to all possibilities. So it'll be simpler 6/36.

How can you find the probability of getting a cumulative total of, say, 6? Then again, it would be very simple. It could be 1, 5; 2, 4; 3, 3; 4, 2; 1, 5. All right? So it'll be 5 by 36. OK? So either by using trees or grid, you can easily find the probabilities.

Now, let's look at a few concrete examples. Let's see. Suppose we are throwing three coins. Then, what is the probability of one particular outcome in that trial, in all three trials? What is the probability, assuming that these are unbiased coins? What is the probability of one particular outcome? Because how many possible outcomes are there if you are throwing three coins?

Consider this tree. First, it splits into 2. Then, it splits into 4. Then? 8, all right? OK, so there are 8 possible outcomes.

So each outcome will have the probability 1/8. so what is the probability of heads appearing exactly twice? How can you do that? Of course, you can write the tree and count. What is the easier way of doing that? Since we know this count, since we know this probability of a particular event happening? How can we come up with the

probability of getting exactly 2 heads?

It could be head, head, or tail-- so this is by enumerating all the possible outcomes. So it could have been head, head, tail, where me put the tail only at the end. It could have been head, tail, head. Or it could have been tail, head, head. In these three cases, you're getting exactly 2 heads. So we are enumerating all possible outcomes. And we know each possible outcome will take the probability 1/8. So the total probability here is 3/8. OK? So this is one way of handling a probability question.

You can do that only because these are independent events. And you can sum them. We'll come to that later.

Suppose you are rolling two four-sided dice. And assuming they're fair, how many possible outcomes are there? Two four-sided dice, and assuming that each of them are fair-- that means unbiased-- how many possible outcomes are there?

Consider this tree. First, it branches into 4, OK? In the first trial, it's a four-sided dice, so there are 4 possible outcomes. So it branches into 4. Then, each branch will, in turn, fork into another 4 branches. So there are totally 16 outcomes.

So what is the probability of rolling a 2 and a 3? What is the probability of rolling a 2 and a 3? Not in a given order, not in the given order. Can anyone give the answer?

OK, let's see. So we have to roll a 2 and a 3. So which means it could have been 2, 3, or 3, 2. And we know the probability of each event is 1/16. So this will be 1/16. And this will be 1/16. So the total probability is 1/8.

What is the probability of getting the sum of the rolls an odd number? What is the probability of getting an odd number as sum of the rolls? Now, this is getting a bit tricky because now it's maybe a bit harder to enumerate all possible cases. So how can we do that?

There should be a short cut.

AUDIENCE: It can either be odd or even.

PROFESSOR: Sorry?

AUDIENCE: You can either get odd or even.

PROFESSOR: It can be either odd or even, right? So it will be 1/2. OK, there's another trick we might be able to use to get the answers quickly.

What is the probability of the first roll being equal to the second roll? In the same line, you can think. What is the probability of getting the first roll equal to the second roll? It's quite similar to this. Any ideas?

It's a four-sided dice. There are 4 possible outcomes. This is one case where it could be 1, 1, or it could be 2, 2, or 3, 3, or 4, 4. And if it's inside a dice, it would be n, right? So if it's n-sided dice, there and n possible outcomes desired, and totally n by n outcomes. So you get 1/n probability.

What is the probability of at least 1 roll equal to 4? At least 1 roll equal to 4? This is very interesting. These type of questions, you'll get in that Psets, I know. Probably in the quiz, too. What is the probability of getting at least 1 roll equal to 4?

OK, so what are the possible outcomes? First roll, could be a 4. And the second roll could be anything. Or it could be 4, and the first roll could have been anything. Or both could have been 4, but we would have considered that here, as well.

So what we had to do is we had to calculate this probability and this probability, add them, and deduct this, because this would have been double counted. It's quite like, this intersection. We want to remove that, and we want to find the union OK? So what is this probability?

Since we don't care about the second roll, we have to care only about the first roll, our first roll getting 4, which is 1/4. And this is 1/4 similarly. And this is 1/4 by 1/4, so 1/16. So it'll be 1/2 minus 1/16. And when you give the answers, if it's hard, you can just leave it like this.

So this is what we call giving the answers as formula instead of giving exact fractions. Because sometimes it might be hard to find the fraction. Suppose it's something like 1 over, say, 2 to the power 5 and a 3 to the 2, something like this. Or we'll say 5. You're not supposed to give the exact value in this amount or even the fractions. You can give such formulas. You can give something like this, too, to give the inverse probability of that not happening.

Let's see. Let's move into a little bit more complicated example. A pack of cards-- what is the probability of getting an ace? Anyone?

AUDIENCE: 1 out of 2?

PROFESSOR: 1 out of 2?

AUDIENCE: out of 52.

PROFESSOR: Not a particular-- an ace, yes, just ace.

AUDIENCE: Is it 4 out of 52?

PROFESSOR: 4/52, yes. Or if you consider one suit, it would have been like 1/13, right? You could have considered one suit, and out of-- OK. It's the same analysis, right? OK.

What is the probability of getting a specific card, which means, say, the ace of hearts? It's what she said, yeah, 1/52. What is the probability of not getting an ace?

AUDIENCE: [INAUDIBLE]?

PROFESSOR: Sorry?

AUDIENCE: 1 minus--

PROFESSOR: 1/13. OK, this is where me make you solve the inverse probability. OK, so that will come into play very often.

OK, now let's get into two decks of playing cards. OK, what is the sample size? What is the sample size of drawing cards from two decks of cards? Two cards,

actually. You're going to draw two cards from two different decks. Sorry? OK.

What is the sample size of drawing a card from one deck? There are 52 possible outcomes. So for each outcome here, we have 52 outcomes there, right? So it's 52 by 52. It's like the tree, but here, we have 52 branches. So eventually, you will have 52 by 52. This is where you can't enumerate all the possible cases. So you should have a way to find the final probability, OK?

So in this case, what is the probability of getting at least one ace? What's the probability of getting at least one ace? This is, again, similar to this case. Remember this diagram. It's called Venn diagram. Remember this. So what is the probability of getting at least one ace, which means you could have got the ace from the first deck, or the second deck, or both. But if you're getting from both, you have to deduct it because otherwise, you would have double counted it.

So getting an ace from the first deck is 1/13. Second deck, 1/13. Getting from both is 1/52 by 52. Sorry, 1/13 by 1/13. Sorry.

AUDIENCE: Are you adding them?

PROFESSOR: Yeah, that's what I explained earlier. You're doing two trials. You could have got the ace from here. And this could have been anything. You could have got the ace from here, and this could have been anything. You could have got an ace from both. So you should add these two probabilities because we need a case where at least one card is ace. But the problem is, this could have happened here and here. And so you will deduct it.

What is the probability of getting neither card-- what is the probability of neither card being an ace?

AUDIENCE: 1 minus that?

PROFESSOR: 1 minus this, exactly. OK, you're getting comfortable with the inverse probability now. What's the probability of two cards from the same suit? What is the probability of getting two cards from the same suit? Now, it's getting interesting. Two cards

from the same suit. So how can we think about this? Of course, you can enumerate all possible cases and count. We don't want to do that.

OK, you're going to use the grid here to visualize this. OK? It could have been a spades, or hearts, or clubs, or a diamond. So we want two cards of the same suit, right? So it's 4/16 possible outcomes. Do you see that? So see, we are using all the tools available at our disposal-- trees, grids, counting, Ven diagrams, inverse probability. Yeah, you should be able to do that to get the answers quickly because you could have actually done-- you could have done something like this, too. But it will take more time, right? So this will be a simpler way of visualizing things.

What is the probability of getting neither card a diamond nor a club? Neither card is diamond nor club. That is tricky. But since we have this grid, we can easily visualize that. So if neither card is diamond nor club, then it could have been only these two values, right? Which is, again, 4/16. So there are 4 possible cases. OK?

So what is the summary? What is the take home message here? In probability, the probability of the belief, or the way of expressing the belief, of a particular event happening. Now, there could be several possible outcomes. Out of those possible outcomes, you have a certain number of desired outcomes. How can you find that?

You can either enumerate all of them. You can put them in a tree, or you can put them in a grid. Or you can use some sort of Venn diagram and come up with some sort of analysis.

Here, we start with our belief that the coin is unbiased, or we have a fair chance of drawing any card from the deck of cards. So we have all these unbiased beliefs, or beliefs about the characteristics of each trial. So we start from that.

Then, we find the probability of a particular event happening in a certain number of trials. But what if you don't have the knowledge about the coin? What if you don't know whether it's fair or not? What if you don't know P of A is equal to H is equal to 1/2? Suppose you don't know that. Suppose it's P. How can you find it?

What you could do is you could simulate this. You can throw coin several times and

count the total number of heads you get, OK? So it could be n of heads over n trial will give you the P, right? This is a way of finding the probabilities through a certain number of trials. It's like simulating the experiments. It's called Monte Carlo simulation. And using that, we try to find a particular parameter of the model.

You know how they actually found the value of pi at the beginning, pi? It's again using a Monte Carlo simulation. What you could do is for a given radius, you can actually check whether it lies within a circle or not. You can simulate the Monte Carlo simulation. And given this radius, you can come up with a particular location at random and check whether it's within this boundary or not, OK?

So then, you know the outcome. You know the outcomes, right? So suppose this is n_a, And the total outcome is n_t. This gives you the area, right? We know this is r-squared, and this is pi r-squared. Sorry. When this is 4 r-squared, this is 2r, right? So using this, you can easily calculate pi.

So now, since we are going to come up with these parameters through repeating the trials, we need to have a standardized way of finding these parameters. We can't simply say this, right? Take this example.

You know this MIT shuttle right? A shuttle arriving at the right time, or the time difference between the arrival and the actual quoted time can be plotted in a graph. So if you put that it is spread around 0, right? Probably, or we hope so. OK?

Now, from this, we can see that actually the mean of this simulation will give you the expected difference in the time, the expected difference in the arrival time from the actual quoted time. And we hope this expectation to be 0. We call that mean. Means is taking the average.

But this distribution might actually give you some information, some extra information, as well. That is, how well we can actually believe this, how much we can rely on this. If the spread is greater, something like this, then probably you might actually not trust the system, right? Although the mean is 0, it's going to come early or late, right? Which means it's useless.

Similarly, in this case, we have a spread around mean 0. But if you take the score, the marks you get for 600, it could be something like this. It's not centered around 0, right? Hopefully. It's probably, say, 50. Then, we actually want the spread to be small or large? We want the spread to be large because we want to distinguish the levels, right? The students' level of understanding. 600.

Anyway, so the spread determines what is the variation percent in their distribution of the scores? We measure that by a variable called standard deviation. In this case, this particular sample will be different from its mean by a particular value, right? We can express that as x_i minus its mean. Let's call the mean mu. So this would be the difference.

Standard deviation is summing up all the differences. But the problem is, when you sum up the differences, it'll be 0, right? The total summation of the differences will be 0 if that's how you get the mean because if you expand this, it'll be something like this, right? Which will be n mu. Should be equal to 0.

So we have to sum, or actually take the differences into account. So, let's square this. So now, it will no longer be 0. Now, this gives 0, the differences. It's the squared sum of the differences averaged across all the samples. We call this variance. And the square root of variance is standard deviation. OK?

Now, having a standard deviation-- so we know the standard deviation tells you how spread the distribution is. But can we actually rely only on the standard deviation to determine the consistency of some event? Can we? Probably not.

Suppose take two examples, one is the scores, 50. And suppose the standard deviation is minus 10, plus 10, OK? So the standard deviation is 10 here. Suppose it lies in this form.

Consider another example, the weight, the weight of the people, like say at MIT. And suppose it's centered around 150. Now, if the standard deviation is, say, 10, then the standard deviation 10 here and the standard deviation 10 here don't convey the same message, OK? So we need to have a different way of expressing

the consistency of a distribution.

So we represent it by coefficient of variation, which is equal to the standard deviation divided by mean. Now here, it will be 10/150. Here, it will be 10/50. So we know this is more consistent than this. The weights of the students at MIT, it's more consistent than the marks you might get, or you get, for 600. It might be true.

Now, what is for the use of the standard deviation? How can we use that? Let's look at this graph where suppose the mean is 0 and the standard deviation is, say, 5. Consider another example where standard deviation is 10. It might have been like this, OK?

Now, before that, let me sort of digress a little bit so I can explain this better. We can take the outcome of a particular event as a sample in our distribution. So suppose you're throwing a die. So you get an outcome. You can represent that outcome as a distribution, OK?

So here, there's x, which can take 1 to, say, 6. And we can represent x_i as a sample point in our distribution. So I don't know, it might be uniform, probably, we hope. So it's with 1/6 probability, we always take one of these values. OK.

But this might not be the case with all events. OK, so what I'm trying to say here is you can actually represent the outcome of the trial in the distribution. Or you can also represent the probability of something happening in a distribution. How does it work?

OK, in this case, we throw our dice. We get an outcome. We go and put it in the x-axis. It could be between 1 and 6. And it takes this distribution.

In addition, what you could do is you could have, say, 100 trials. So you throw a coin. You take 100 trials. You get the mean, you get the probability of getting a head. And you have that mean, right?

So probability of getting a head for 100 trials, say, 0.51. You do another 100 trials, you got another one. So you have now another distribution. So there's a distribution

of probabilities. So you can have a distribution of probabilities, or you can have a distribution for the events. We handle these two cases in the p-set. So probably you should be able to distinguish those two.

Anyway, so here in this particular example, let's take this as our mu. Let's take this as our standard deviation. And for the first distribution, let's take the standard deviation to be 5. When the standard deviation is great, it's going to be more spread. It's going to be more distributed than the former.

So here, say the standard deviation is 10. The standard deviation is a way of expressing how many items, how many samples are going to lie between those particular boundaries. So for a normal distribution, we know the exact area, exact probability of things happening. If there's no mu, we know within the first standard deviation, there will be 68% of events lie in that area. Within two standard deviations-- OK, one standard deviation, 68%. Two standard deviations on either side, it's going to be 95%. Three standard deviations, it's going to be 99%.

So suppose you conducted so many trials. And you get the values. And in the distribution, suppose mu, mean, is 10, and the standard deviation is, say, 1. So now, with 99% confidence, we can say then the outcome of the next trial is going to be between what? 7 and 13, right? So this is where finding the distribution and standard deviation helps us giving a confidence interval, expressing our belief of that particular event happening.

We will look at a few examples because you might need this in your p-set. So this particular function you have already seen in the lecture. But we need to understand this particular part.

Suppose you have a probability of something happening. Suppose you estimated the probability of something happening. Suppose you're given the coin is biased, OK? Sorry, unbiased. So we know p of H is equal to 1/2. How can we simulate an outcome? How can you simulate an outcome and see whether it's a head or a tail with this particular probability?

We do that by calling this function, random.random(), which is going to give you a random value between 0 and 1. And you're going to check whether it's below this or not. If it's below this, we can take it as head. If it's not, it's tail. And this will happen with probability 1/2, because the random function is going to return a value between 0 and 1 with equal probabilities. It's uniform probabilities. So to simulate a head or tail, you call that function. You write the expression like that, OK?

Then, if you consider this example, for a certain number of flips, we simulate the event. And we count the number of heads we obtain. And also from that, you can calculate the number of tails as well. If you know the total flips, you know the number of tails.

Using that, we are taking two ratios. Now, the ratio between the heads and tails, and the difference between heads and tails. We are doing this for certain number of trials. And we're going to take the mean and standard deviation of these trials, OK?

So here in our distribution, what are we considering? What is going to build our distribution here? The ratios, right? The ratios of the events. And we simulated certain number of trials to get those events, OK? Only if you simulate certain number of trials, you can actually summarize the outcome of the events in mean and standard deviation. This is exactly like the difference in the times of the bus arriving and the quoted times.

Let's check this example. Let's plot this and see. It's going to take a while. OK, that's another thing I want to explain here because since you're going to be going to plot-- we are going to use PyLab extensively and plot graphs. You'll need to put a title and labels to all the plots you're generating. Plus, you can use this text to actually put the text in the graph. We will show that in a while.

Plus-- here, sorry. If you want to change the axis to log-log scale, you can call this comma at the end after calling the plot. Because you might sometimes need to change the axis to log scale in x and y-axis.

So this is the mean, heads versus tails. And if you can see it, the mean tends to be

1 when we have a large number of flips. So to get the consistency, we need to simulate large number of trials. Then only it will tend to be close to the mean, OK? This is sort of a way of checking the evolution of the series by actually doing it for a certain number of flips at every time.

So it's quite like a scatter plot. A scatter plot is like plotting the outcomes of our experiments. Suppose it's x1 and x2 in a graph. So we are going to say-- so for example, suppose you have a variable, and the variable causes an outcome-- a probability of the coin flip, so p of H. And it can result in a certain number of heads appearing, say n of H.

Now, you can do a scatter plot between these two variables. And it will be probably a spread. But we know that if you increase the probability of heads, the number of heads is going to increase as well. So it would be probably something like this.

From this, we can assume that it's linear or something like that. But the scatter plot is actually representing the outcomes of the trial versus some other variable in the graph and visualize it. And let me show the last graph, and we'll be done with that.

So this, again, we actually know, instead of putting a scatter plot, we're actually giving the distribution as a histogram and printing a text box in the graph. This might be useful if you want to display something on your graph. I guess we will be uploading the code to the site. So you can check the code if you want later, OK? Sure. See you next week.