VINA NGUYEN: Hi. Are you guys all ready? All set? OK. So my name is Vina Nguyen. I'm at MIT, going to be a senior next year. I'm studying electrical engineering and computer science. That's just one major-- half of both. And I'm teaching probability because I thought it was an intro class that I really liked a lot. OK, hold on.

[DOOR CLOSING]

Yeah. Sorry. I speak kind of soft, so if you can't hear me, just tell me, and I'll try to speak louder. OK. So there's a few announcements I have to go over. First, you need to be here because you want to be here, not because your parents want to be here. So if you want to leave, leave. But I want you to take it because you want to take it. That's the best way to learn.

During registration you also got your sheet with the blank schedule. I gave you that sticker. So you can put it on there. And everyone did that? OK. So next week, you should come to this room directly. So don't come to Lobby 13. Just come directly here. OK. So on your registration sheet, they also told you about a server called Caroline. So for my class, on the sticker you need to type in the password, which is frog something. And I'm going to post these slides up there, or any lecture notes I had in case you missed a class, or you wanted to review something.

There is chat and forums, but I work full time during the week. So I probably won't have time to always chat with you guys. But my email is listed on the syllabus if you do have questions. I always check my email, but I don't necessarily have time to chat. So email is good if you have questions. Fourth announcement, there is a lunch-time activities form you also got. You need to get that signed. Make sure you turn it in next week, because HSSP won't take it the week after next week. And if you have any other questions about administrative stuff, go to the HSSP office, because I'm just a teacher here to teach and not here to manage paperwork.

OK. So now that that's done, I'm going to start. OK. So as you all know, this is probability. I told you a little bit about myself, but I also want to know a little bit about you guys. I have never been to Boston except for college, so I'm not entirely sure what it's like. If you could go around and tell me your name, and what grade you are, and what school, then-- start with you.

AUDIENCE: I'm Margaret [INAUDIBLE]. I'm going to ninth grade, and I'm going to [INAUDIBLE] High School.

VINA NGUYEN: OK. That way.

AUDIENCE: [INAUDIBLE]. And I'm going to junior year. And to Wellesley.

AUDIENCE: My name is Iris. And I'm also going to Wellesley. And I'm going to be in 12th grade.

AUDIENCE: I'm [INAUDIBLE]. I'm going to be in 12th grade. I'm related to her. And yeah, so I go to Wellesley.

AUDIENCE: I'm Amy, And I'm going to ninth grade at Wellesley High School.

AUDIENCE: Hi. I'm [INAUDIBLE], and I'm going to be a junior at [INAUDIBLE] High School.

AUDIENCE: I'm [INAUDIBLE], and I'm going to be in ninth grade, and I go to [INAUDIBLE] Middle School.

AUDIENCE: I'm [INAUDIBLE], and I'm going to eighth grade, also, in Stonybrook.

AUDIENCE: My name's Kevin, and I'm going to seventh grade. And I'm going to Boston Latin Academy next year.

AUDIENCE: I'm Diana, and I'm going to junior year also. And I'm going to [INAUDIBLE] High School.

AUDIENCE: I'm Andrew. I'm going into seventh grade. I'm going to Cherry Hill.

AUDIENCE: I'm Ben. I'm going into sophomore year at Kingsbury Oxford.

AUDIENCE: I'm Pierre, and I'm going to 11th grade at [INAUDIBLE].

AUDIENCE: I'm Tina. I'm going to be a junior at [INAUDIBLE] High.

AUDIENCE: Me?

VINA NGUYEN: Yeah.

AUDIENCE: I'm [INAUDIBLE]. I'm going to seventh grade, And I'm going with Wellesley Middle School.

AUDIENCE: I'm Alan I'm going to eighth grade, and I'm going to Boston Latin School.

AUDIENCE: I'm James. I'm going to eighth grade, and I go to [INAUDIBLE] School.

VINA NGUYEN: OK So I guess we have a pretty good range. Shh. Thank you. So I'm sure you guys know this is all for beginners. If you already know this stuff, you have no reason to be here. This lecture, I'm probably going to use mostly PowerPoint. If I do some things, I'll use the chalkboard. And we'll work out problems, either as a class, or I'll hand out stuff. And there is no homework. So and please feel free to ask questions. I don't want to move on unless you guys understand it. It's like math. Everything later builds up on what you knew before.

OK. So why should we study probability? Well first, you want to model the uncertain. It's easy to be like, oh, I don't know anything about this, so I can't decide anything. I can't estimate anything. But with probability, you can at least get a sense of what's going on with the world. And it's not just that you want to know what's going on, but you also want to decide based on it.

And then for the third point especially, there's also studies out there in the news, they make up some random statistic-- I mean, not always, but it could sound that way-- but if you have probability background, you can at least take a more intellectual approach to it. You can't just take it as is. You need to question other factors, or maybe there's a certain way of sampling that they did wrong. So this would help you to understand more what the studies are confirming or suggesting.

Some examples are like, what's the weather like tomorrow, sun, rain, how much percent? What are the chances of a drug working-- depending on the factors of the person, their age, their gender, their medical background-- how successful will this drug be? Third kind, what kind of customer will buy my product? Maybe based on their buying behavior or their demographics, how likely will this product sell to them? How profitable will it be?

Fourth one, should I buy a lottery ticket? Will two help? And then there's also bio-applications, like whether your child will be a boy or a girl. Hopefully you guys aren't there yet. So now you know why we should learn probability, you should know there's two actually different definitions for it. The first one is frequency probability, which is the more physical. Like, if you repeat something over and over again, how often will the result happen? So how likely is a certain event that you know is the same every time going to happen?

And the second one is Bayesian probability, which we will get to later, and that's a measure more of how sure you are that something will happen given the evidence. So in Bayesian probability, unlike frequency probability, you can't repeat something over and over again. An example of that would be like, how likely am I going to get an A given that I attended all the classes, or I participated, or that-- you can't repeat that experiment over and over again. It's a measure of belief. So is that clear to everyone? OK.

So we're going to look at frequency probability a little bit more. I'm sure you guys know that there's a 50% chance that heads will come up-- or tails will come up-- when you flip a coin. But how exactly are you going to measure that? I mean, you could flip it once, or you could flip it twice, or you could flip it three times, or you could flip it a lot of times and then figure out the ratio of heads to tails. So we're going to do not necessarily an experiment. I have a coin, and we're going to flip it a large number of times. And I want you to observe the percent of heads that comes up after each time.

So you want to take the average of heads that comes out after one time, after two times, after 10 times, et cetera. And what I want you to do is observe what happens initially, so after one or three times, and then what happens after a while. So what I actually have here is an Excel sheet. If you can see on the left column, it's the number of trials. The second one is a 1 or 0, 1 being heads and 0 being tails. And then after each time, it calculates the ratio of heads.

So the way we graph this-- can everyone see this? Yeah. OK. So you know a probability goes from 0 to 100. So that would be 0% heads and then 100%. And you're expecting around 50%, right? So that's 50, 10, 20, 30, 40, 60, 70. OK. And then your x-axis is the number of trials, so the number of times you flip your coin. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. OK, we have around 15 people, but probably go up to 20 times. OK. So I have a coin, and I want each of you to flip it and tell me if it comes up heads or tails. And then I'll use the Excel chart to calculate it. OK.

[COIN LANDING]

AUDIENCE: Heads.

VINA NGUYEN: Heads. OK. So right now, your ratio is 100% because one out of one heads. Give it to the next person behind you.

AUDIENCE: Tails.

VINA NGUYEN: Tails? OK, so 50%. Next person.

AUDIENCE: Tails.

VINA NGUYEN: Tails. OK. So now you're at 33%, just like around here.

AUDIENCE: Tails.

VINA NGUYEN: 25%.

[COIN LANDING]

AUDIENCE: Heads.

VINA NGUYEN: Heads. 40%.

AUDIENCE: Tails.

VINA NGUYEN: Tails. It's 33%.

AUDIENCE: Tails.

VINA NGUYEN: Tails. 29%.

AUDIENCE: Tails.

VINA NGUYEN: 25%.

AUDIENCE: Heads.

VINA NGUYEN: Heads. 33%.

[COIN LANDING]

AUDIENCE: Tails.

VINA NGUYEN: 30%.

AUDIENCE: Tails.

VINA NGUYEN: Tails. 27%.

AUDIENCE: Heads.

VINA NGUYEN: Heads. 33%. Was that the last person? OK. Oh, go around. OK.

AUDIENCE: Yeah. We've got to go too.

AUDIENCE: Tails.

VINA NGUYEN: Tails. 31%.

AUDIENCE: Heads.

VINA NGUYEN: Heads?

AUDIENCE: Tails.

VINA NGUYEN: 36%.

AUDIENCE: Heads.

VINA NGUYEN: Heads. 40%. OK. So oh, not done yet.

AUDIENCE: Of course.

AUDIENCE: Heads.

AUDIENCE: Heads.

VINA NGUYEN: 44%.

AUDIENCE: Tails.

VINA NGUYEN: Tails.

AUDIENCE: Anybody else?

VINA NGUYEN: Anyone else? No.

AUDIENCE: No.

VINA NGUYEN: Thank you. OK. So you have a pretty small group. If you were to take just that experiment, you would assume that heads showed up about 40% of the time. That's not true, right? So you need to have a larger sample. So in this one, you guys can see that I auto-generated the same experiment, but 500 times, which we can't do today. It'd take way too long. But if you can see at the beginning, which is kind of what we did, it varies a lot. The 20 here would be like this part, where it's still very variable. But after a very long number of times, it will eventually converge to 50%.

So this is actually the law of large numbers-- the explanation for that we can cover later-- but this is just an example of how you would figure out what a frequency probability is. And you need to know how many times you need to do that. So this law is only relevant for large numbers, hence, large number. So does that make sense to everyone? OK.

Before we do probability, one of the very important things is set theory. I can't remember if high school taught you guys that. Do you guys know what set theory is, just a little bit basic understanding? OK. I'm just going to go through it just in case for you guys who need the refresher. So a set is a collection of objects. For example, there's outcomes of the die. You can have 1, 2, 3, all the way to 6. And each object in a set is called an element. And each element needs to be unique.

So if you have a collection, say, of 1, 1, 2, 2, that's not a set. It has to reduce to 1 and 2. There's different kinds of sets. You can have your empty set, which is represented by the 0, the cross. There is a set with an infinite number of elements. An example of this is a set of integers. It can go on negative 1, 0, 1, 2, et cetera, et cetera. So even though it's not a finite number of objects, it's still considered a set.

You can have subsets. So if you have a set H, then H is a subset of G if every element in H is in G. So if H was a set of 1 and 2, then that's a subset of G as well. And if every element in G is a subset of H, then that means they're equal. And then you have your universal set, which is symbolized by the omega there. And that means it contains all the elements possible in your problem context.

So this is a more graphical look at it. So assuming that your universal set for this context is all numbers, then you have 1,053.6-- or any number-- and then your set G would be all the integers. And then H, which has 1 and 2, would be your subset of G. So G encompasses H. And does that make sense to everyone? There are set operations you can perform. First one is complement of S, which, oh, S is just a random set called S. So complement of S means all the elements that are not in S.

So the red here means that those are the shaded areas that are not in S but are in the universal set. Is that clear to everyone? OK. And then the second one is union of sets, which means all elements in S or T or both. And that's symbolized by that U right there. So if you had a set that had 1 and 2, and then T had 2 and 3, then your union of sets is 1, 2, and 3. So you don't count 2 twice, because elements of the set have to be unique. Right? OK.

And then your intersection of sets means all the elements of both S and T. So given that example I had before, then your intersection would only be 2. So does that makes sense, everyone? OK. So I just want to make sure you guys get it. Can anyone tell me what the first one is?

AUDIENCE: The system is giving me shade. Do you want the shade?

VINA NGUYEN: Yeah.

AUDIENCE: OK, so complement of S.

VINA NGUYEN: Mm-hmm. And the second one? Yeah?

AUDIENCE: Complement of T?

VINA NGUYEN: You sure? Oh, sorry. The second one is this one, right? Yeah.

AUDIENCE: OK.

VINA NGUYEN: So complement of T would mean this area and that area, right? But because T is shaded, it can't mean not T. Can anyone help her?

AUDIENCE: Is it complement of S plus union of S and T?

VINA NGUYEN: Yep. Can anyone see that? I'll just write the answers. OK. So your first one was SC. Can everyone see this? Can everyone see this? OK. So you said, complement of S, union T, right? Does everyone see why that's right? Your complement of S would be, like,

AUDIENCE: Kind of an intersection.

VINA NGUYEN: Hm?

AUDIENCE: Yeah, it's kind of an intersection.

VINA NGUYEN: Oh, OK. Well, it's actually a union. Sorry. Wait.

AUDIENCE: Intersection of--

VINA NGUYEN: Complement S, union T. Yeah, it's--

AUDIENCE: It's an intersection, which is what I meant to say. But I said union.

VINA NGUYEN: Wait, sorry?

AUDIENCE: It's intersection, but I said union by accident. So I think it was union.

VINA NGUYEN: It's union.

AUDIENCE: Yeah.

VINA NGUYEN: Yes. Because if you said union and T, then none of this could count. Right?

AUDIENCE: Oh.

VINA NGUYEN: Yeah.

AUDIENCE: Oh.

VINA NGUYEN: Is that clear for everyone?

AUDIENCE: No.

VINA NGUYEN: OK. So you have your SC, right? Plus T. It can't-- OK. Does that make sense? Actually, S was this part too. Sorry. OK. So the third one?

AUDIENCE: Intersection of ST--

VINA NGUYEN: Comp--

AUDIENCE: Comp-- wait.

VINA NGUYEN: Comp-- Complement. Right?

AUDIENCE: OK.

VINA NGUYEN: OK. Did everyone hear that? He said it was intersection of S and T complement, so complement of the intersection of S and T. I'll do that again. So S union T complement. Does that make sense to everyone? OK. And another way you can write that is complement of S and complement of T. Does everyone see that? Yeah? OK. So 4?

AUDIENCE: T?

VINA NGUYEN: Yes. And 5?

AUDIENCE: Union, wait. Union set of T, complement of T?

VINA NGUYEN: Complement of T--

AUDIENCE: With the union set of S and T?

VINA NGUYEN: Close.

AUDIENCE: Complement of T intersection of S?

VINA NGUYEN: Yes. Yeah. Wait. Yeah. OK. So that would be-- right? OK. Does everyone see that? OK. And the sixth one?

AUDIENCE: Complement of the union of S and T?

VINA NGUYEN: Mm-hmm. And given this example, can anyone tell me a different way to write the sixth one? Oh.

AUDIENCE: Complement of S and the union as a complement of T?

VINA NGUYEN: This one?

AUDIENCE: Intersect.

VINA NGUYEN: Intersect.

AUDIENCE: And the intersection.

VINA NGUYEN: Yeah. OK. Does everyone see that? OK. So that is a very abstract thing, but it's very fundamental or everything else doesn't make sense. OK. So probability models, your sample space is like a set, right? You need to know what are all the possible outcomes. So that would be your universal set. It has to be exhaustive. You can't leave out any events or your probabilities will not be correct. And none of the events can overlap.

So every result that can happen has to be uniquely defined within your context. And then the events are a subset of your sample space. They don't necessarily have to be like probability of die being 1. They could be probability of your result being even. So it doesn't necessarily have to be just one. It can be a couple of events that can happen. And you also need probabilities. So that's your quantitative measure of the problem.

So if you were to model rolling just one die, what would your sample space be? Anyone?

AUDIENCE: It would be numbers 1 through 6.

VINA NGUYEN: Yup. And what kind of events would you have? Like, event that you roll a 1? Is there any other kind of event?

AUDIENCE: In general you can conclude--

VINA NGUYEN: Yup. Or it could be, like, an event that you roll greater than 4, or something like that. And then you need probabilities for that. So even number would be like one half. Greater than 4 would be like one third. Or probably of 1 would be one sixth. So what if you needed to do two die? Then how would you represent what that problem space is? Anyone?

AUDIENCE: Doesn't it depend on what event you're talking about? If you wanted to do sum, then it'd be 2, 12. But if you wanted to do pairs, then it'd be one--

VINA NGUYEN: So that would be events. But in terms of sample space, you only can have certain results that come out from your die, which would be the combination of numbers. Right?

AUDIENCE: Oh, so we're talking about confirmation as-- because couldn't you also base it off, like, what's the chance that the numbers are the same?

VINA NGUYEN: Right. So that would be an event, but not part of your sample space. Because your sample space is the actual, physical results that could happen, like a 1 and a 1, a 1 and a 3. And then how you interpret that is an event.

AUDIENCE: OK.

VINA NGUYEN: So I guess I'd kind of given it, but-- Where's my chalk? Hm?

AUDIENCE: It would both be the same--

AUDIENCE: Wouldn't it be the same?

VINA NGUYEN: It would not be the same, because you're rolling two, right?

AUDIENCE: The solution--

VINA NGUYEN: Yeah.

AUDIENCE: So the expert's--

VINA NGUYEN: Yeah.

AUDIENCE: It's gonna be one--

VINA NGUYEN: Right. So I'll get to that later. But the more basic thing-- can you guys see this? I don't--

AUDIENCE: Kind of.

VINA NGUYEN: Kind of. I'll just-- so you were right about representing it. But for what it actually is, your sample space would be like 1, 1. Right? That would be the roll of your first die, the role is your second die.

AUDIENCE: Oh, so you'll--

VINA NGUYEN: Yeah, just really basic. Yeah. Like that. So that's your sample space.

AUDIENCE: Oh.

VINA NGUYEN: Yeah.

AUDIENCE: Do you write it on--

VINA NGUYEN: I'm not going to. I can assume that you guys can figure that out. But is that clear how that's separate from an actual event? Yeah? OK. Right So that would be your sample space. But like what you mentioned, what's your name?

AUDIENCE: Mine?

VINA NGUYEN: Yeah.

AUDIENCE: [INAUDIBLE]

VINA NGUYEN: OK. What [INAUDIBLE] mentioned is that the events-- the probability that your sum is 12, that's an event. But that's not part of your sample space. Is that clear how that's different? Or the probability that you get doubles-- so do you guys understand difference between your physical sample space and the different kind of events you can get from it?

AUDIENCE: Is the sample space include exact values that you actually can get it?

VINA NGUYEN: Your exact outcome. The bare bone. And then your events would be how you interpret it. Is that good? OK. And the probability is just like what's the probability this will happen? What's the probability that will happen? I know this is one sixth. I will figure that out later. OK. So going back to what you said, how you represent this is pretty much up to you. The standard way is what you mentioned. What were you saying, again?

AUDIENCE: The table where that has one that sits on top. And this--

VINA NGUYEN: Right. So you're talking about this. Right? Is that what you're--

AUDIENCE: Yeah.

VINA NGUYEN: Yeah. OK. So this would be your first die.

AUDIENCE: OK.

VINA NGUYEN: And this would be your second. One, two-- OK. So that's one way to represent your outcomes, right? So the probability of doubles would be these diagonals. Does everyone see that? OK. And then 12-- oh, there's just one. OK. So that's easy. Is there another way to represent sample space? I just left that there so you guys can draw. Is there another way to represent how your experiment will progress?

AUDIENCE: Graph?

VINA NGUYEN: Graph? How would you graph it?

AUDIENCE: Like 1 through 6 on the x-axis, and--

VINA NGUYEN: So it's kind of like this, except it's like a grid. You would just put it here. Right? So, yes, you're right. But it's like this already. Is there another way? How much time we have? Well, what if you did a tree? So you roll it once. You get a 1. You get a 2. You get a 3, 4, 5, 6. And then you would just expand this-- 1, 2. Right? So that's just another way to keep in mind. For this example, it grows exponentially. So it's not the best way. But it is a good way for certain contexts. So does everyone see how that works? OK.

And we ended kind of short, because I wasn't sure how long this would take. So I want to make sure you guys know why we study probability. There's lots of reasons out there. There's two different definitions. I know that's not entirely intuitive. So it's either how often something happens given a number of repeatable experiments, or the second definition, how much you believe that something will happen given the evidence.

And then basic set theory, you guys seem to pretty much know that. I wasn't sure how much knowledge you guys had. But if you do have any questions, don't hesitate to email me. I can go over that with you. And probability models are important. So you can take a certain puzzle or problem context in your mind and graph it out in something that you can actually work with.

So does anyone have questions or anything about the class, probability, HSSB? Was this too fast or too slow? I'm not really sure the knowledge you guys have beforehand. No?

AUDIENCE: I think it's a good pace.

VINA NGUYEN: Good pace? OK. Next time I'll fill up the time. Sorry. I wasn't sure how long registration was going to take.

AUDIENCE: Could we go over set operations again?

VINA NGUYEN: Yup. OK. So do you have a certain question or just want to go over it?

AUDIENCE: No. Just in general.

VINA NGUYEN: Just in general? OK. So do you know what a universal set is? All the possible things that can happen is a universal set. And if you have a set-- so say your set is all the combinations that are doubles. So S would be like 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6. So this would be set S. And then complement of S is anything that's not an S. So that would mean anything here that's not a double. Does that make sense?

So a complement of S is anything that's not in S but within your universal problem context. Good? OK. So union of sets means that if you have two different kinds of sets, it's everything in S and T. So T could be that you rolled a 6 in either one of these. So this would be T. And this would be T. Can you see that? I'm sorry it's a little messy. So your union would be this block, this block, this block, this block, this block, this block, everything here, and everything here. So that would be your union.

And like I said before, even though you might have an element in both sets, when you do write out your set, you only count it once, because every element has to be unique. OK?

AUDIENCE: Does T stand for something or just like a random name?

VINA NGUYEN: T is just a random name. You can be like, set A, set B, set XYZ. I just did it because it's the next letter after S. Is that good? OK. So your intersection of sets is anything that is in both. And given this example, the only thing that's in both would be 6, 6. Right? So your intersection would be just 6, 6 equals S intersect with T. OK? Is that clear?

Did you guys have any other questions about these things?

AUDIENCE: I want to know about that.

VINA NGUYEN: OK. Go ahead.

AUDIENCE: Does it mean-- could you give us an example for it using probability in a situation?

VINA NGUYEN: OK. So there was a pretty simple example I gave you about how confident am I that I'm going to get an A or something, right? But another basic example is like say you notice that your neighbor's grass was wet when you came home. And that's an event, but you're not sure whether it was because of their sprinkler, or because it was thunder-storming this morning, or because-- I don't know-- there was a flood.

So you have to take the probability of each of those previous events, like, what's the probability that he has a sprinkler? If it's 0, then you can say that's less likely unless he got one today or something. Or if you live in California, the probability of a thunderstorm is lower, so you also can get a sense for how likely it was a thunderstorm or not. Right?

I don't about floods in California, but it's kind of like that. You have previous probabilities that you know about that are related to your current problem, but you can't exactly know. But you can quantify it given the previous evidence. Does that make sense?

AUDIENCE: Yes.

VINA NGUYEN: OK. Anybody else? Or any other non set questions or set questions? No? OK. Well, I guess you guys can just finish off the cookies. And if you do have questions, then you can come to me right now after class, or you can email me. My name is Vina, again. And I hope you guys have a good rest of the day. Oh, yeah. You guys should go back to Lobby 13 at 3:00 so you can choose your labs class. OK? Yup.