PROFESSOR: OK, so welcome to 6.041/6.431, the class on probability models and the like. I'm John Tsitsiklis. I will be teaching this class, and I'm looking forward to this being an enjoyable and also useful experience. We have a fair amount of staff involved in this course, your recitation instructors and also a bunch of TAs, but I want to single out our head TA, Uzoma, who is the key person in this class. Everything has to go through him. If he doesn't know in which recitation section you are, then simply you do not exist, so keep that in mind.

All right. So we want to jump right into the subject, but I'm going to take just a few minutes to talk about a few administrative details and how the course is run. So we're going to have lectures twice a week and I'm going to use old fashioned transparencies. Now, you get copies of these slides with plenty of space for you to keep notes on them. A useful way of making good use of the slides is to use them as a sort of mnemonic summary of what happens in lecture. Not everything that I'm going to say is, of course, on the slides, but by looking them you get the sense of what's happening right now. And it may be a good idea to review them before you go to recitation.

So what happens in recitation? In recitation, your recitation instructor is going to maybe review some of the theory and then solve some problems for you. And then you have tutorials where you meet in very small groups together with your TA. And what happens in tutorials is that you actually do the problem solving with the help of your TA and the help of your classmates in your tutorial section.

Now probability is a tricky subject. You may be reading the text, listening to lectures, everything makes perfect sense, and so on, but until you actually sit down and try to solve problems, you don't quite appreciate the subtleties and the difficulties that are involved. So problem solving is a key part of this class. And tutorials are extremely useful just for this reason because that's where you actually get the practice of solving problems on your own, as opposed to seeing someone else who's solving them for you.

OK but, mechanics, a key part of what's going to happen today is that you will turn in your schedule forms that are at the end of the handout that you have in your hands. Then, the TAs will be working frantically through the night, and they're going to be producing a list of who goes into what section. And when that happens, any person in this class, with probability 90%, is going to be happy with their assignment and, with probability 10%, they're going to be unhappy.

Now, unhappy people have an option, though. You can resubmit your form together with your full schedule and constraints, give it back to the head TA, who will then do some further juggling and reassign people, and after that happens, 90% of those unhappy people will become happy. And 10% of them will be less unhappy.

OK. So what's the probability that a random person is going to be unhappy at the end of this process? It's 1%. Excellent. Good. Maybe you don't need this class. OK, so 1%. We have about 100 people in this class, so there's going to be about one unhappy person. I mean, anywhere you look in life, in any group you look at, there's always one unhappy person, right? So, what can we do about it?

All right. Another important part about mechanics is to read carefully the statement that we have about collaboration, academic honesty, and all that. You're encouraged, it's a very good idea to work with other students. You can consult sources that are out there, but when you sit down and write your solutions you have to do that by setting things aside and just write them on your own. You cannot copy something that somebody else has given to you.

One reason is that we're not going to like it when it happens, and then another reason is that you're not going to do yourself any favor. Really the only way to do well in this class is to get a lot of practice by solving problems yourselves. So if you don't do that on your own, then when quiz and exam time comes, things are going to be difficult.

So, as I mentioned here, we're going to have recitation sections, that some of them are for 6.041 students, some are for 6.431 students, the graduate section of the class. Now undergraduates can sit in the graduate recitation sections. What's going to happen there is that things may be just a little faster and you may be covering a problem that's a little more advanced and is not covered in the undergrad sections. But if you sit in the graduate section, and you're an undergraduate, you're still just responsible for the undergraduate material. That is, you can just do the undergraduate work in the class, but maybe be exposed at the different section.

OK. A few words about the style of this class. We want to focus on basic ideas and concepts. There's going to be lots of formulas, but what we try to do in this class is to actually have you understand what those formulas mean. And, in a year from now when almost all of the formulas have been wiped out from your memory, you still have the basic concepts. You can understand them, so when you look things up again, they will still make sense. It's not the plug and chug kind of class where you're given a list of formulas, you're given numbers, and you plug in and you get answers.

The really hard part is usually to choose which formulas you're going to use. You need judgment, you need intuition. Lots of probability problems, at least the interesting ones, often have lots of different solutions. Some are extremely long, some are extremely short. The extremely short ones usually involve some kind of deeper understanding of what's going on so that you can pick a shortcut and use it. And hopefully you are going to develop this skill during this class.

Now, I could spend a lot of time in this lecture talking about why the subject is important. I'll keep it short because I think it's almost obvious. Anything that happens in life is uncertain. There's uncertainty anywhere, so whatever you try to do, you need to have some way of dealing or thinking about this uncertainty. And the way to do that in a systematic way is by using the models that are given to us by probability theory.

So if you're an engineer and you're dealing with a communication system or signal processing, basically you're facing a fight against noise. Noise is random, is uncertain. How do you model it? How do you deal with it?

If you're a manager, I guess you're dealing with customer demand, which is, of course, random. Or you're dealing with the stock market, which is definitely random. Or you play the casino, which is, again, random, and so on. And the same goes for pretty much any other field that you can think of.

But, independent of which field you're coming from, the basic concepts and tools are really all the same. So you may see in bookstores that there are books, probability for scientists, probability for engineers, probability for social scientists, probability for astrologists. Well, what all those books have inside them is exactly the same models, the same equations, the same problems. They just make them somewhat different word problems.

The basic concepts are just one and the same, and we'll take this as an excuse for not going too much into specific domain applications. We will have problems and examples that are motivated, in some loose sense, from real world situations. But we're not really trying in this class to develop the skills for domain-specific problems. Rather, we're going to try to stick to general understanding of the subject.

OK. So the next slide, of which you do have in your handout, gives you a few more details about the class. Maybe one thing to comment here is that you do need to read the text. And with calculus books, perhaps you can live with a just a two page summary of all of the interesting formulas in calculus, and you can get by just with those formulas. But here, because we want to develop concepts and intuition, actually reading words, as opposed to just browsing through equations, does make a difference.

In the beginning, the class is kind of easy. When we deal with discrete probability, that's the material until our first quiz, and some of you may get by without being too systematic about following the material. But it does get substantially harder afterwards. And I would keep restating that you do have to read the text to really understand the material.

OK. So now we can start with the real part of the lecture. Let us set the goals for today. So probability, or probability theory, is a framework for dealing with uncertainty, for dealing with situations in which we have some kind of randomness. So what we want to do is, by the end of today's lecture, to give you anything that you need to know how to set up what does it take to set up a probabilistic model. And what are the basic rules of the game for dealing with probabilistic models? So, by the end of this lecture, you will have essentially recovered half of this semester's tuition, right?

So we're going to talk about probabilistic models in more detail-- the sample space, which is basically a description of all the things that may happen during a random experiment, and the probability law, which describes our beliefs about which outcomes are more likely to occur compared to other outcomes. Probability laws have to obey certain properties that we call the axioms of probability. So the main part of today's lecture is to describe those axioms, which are the rules of the game, and consider a few really trivial examples.

OK, so let's start with our agenda. The first piece in a probabilistic model is a description of the sample space of an experiment. So we do an experiment, and by experiment we just mean that just something happens out there. And that something that happens, it could be flipping a coin, or it could be rolling a dice, or it could be doing something in a card game.

So we fix a particular experiment. And we come up with a list of all the possible things that may happen during this experiment. So we write down a list of all the possible outcomes. So here's a list of all the possible outcomes of the experiment. I use the word "list," but, if you want to be a little more formal, it's better to think of that list as a set.

So we have a set. That set is our sample space. And it's a set whose elements are the possible outcomes of the experiment. So, for example, if you're dealing with flipping a coin, your sample space would be heads, this is one outcome, tails is one outcome. And this set, which has two elements, is the sample space of the experiment.

OK. What do we need to think about when we're setting up the sample space? First, the list should be mutually exclusive, collectively exhaustive. What does that mean?

Collectively exhaustive means that, no matter what happens in the experiment, you're going to get one of the outcomes inside here. So you have not forgotten any of the possibilities of what may happen in the experiment. Mutually exclusive means that if this happens, then that cannot happen. So at the end of the experiment, you should be able to point out to me just one, exactly one, of these outcomes and say, this is the outcome that happened.

OK. So these are sort of basic requirements. There's another requirement which is a little more loose. When you set up your sample space, sometimes you do have some freedom about the details of how you're going to describe it. And the question is, how much detail are you going to include?

So let's take this coin flipping experiment and think of the following sample space. One possible outcome is heads, a second possible outcome is tails and it's raining, and the third possible outcome is tails and it's not raining. So this is another possible sample space for the experiment where I flip a coin just once. It's a legitimate one. These three possibilities are mutually exclusive and collectively exhaustive.

Which one is the right sample space? Is it this one or that one? Well, if you think that my coin flipping inside this room is completely unrelated to the weather outside, then you're going to stick with this sample space. If, on the other hand, you have some superstitious belief that maybe rain has an effect on my coins, you might work with the sample space of this kind. So you probably wouldn't do that, but it's a legitimate option, strictly speaking.

Now this example is a little bit on the frivolous side, but the issue that comes up here is a basic one that shows up anywhere in science and engineering. Whenever you're dealing with a model or with a situation, there are zillions of details in that situation. And when you come up with a model, you choose some of those details that you keep in your model, and some that you say, well, these are irrelevant. Or maybe there are small effects, I can neglect them, and you keep them outside your model. So when you go to the real world, there's definitely an element of art and some judgment that you need to do in order to set up an appropriate sample space.

So, an easy example now. So of course, the elementary examples are coins, cards, and dice. So let's deal with dice. But to keep the diagram small, instead of a six-sided die, we're going to think about the die that only has four faces. So you can do that with a tetrahedron, doesn't really matter. Basically, it's a die that when you roll it, you get a result which is one, two, three or four.

However, the experiment that I'm going to think about will consist of two rolls of a dice. A crucial point here-- I'm rolling the die twice, but I'm thinking of this as just one experiment, not two different experiments, not a repetition twice of the same experiment. So it's one big experiment. During that big experiment various things could happen, such as I'm rolling the die once, and then I'm rolling the die twice.

OK. So what's the sample space for that experiment? Well, the sample space consists of the possible outcomes. One possible outcome is that your first roll resulted in two and the second roll resulted in three. In which case, the outcome that you get is this one, a two followed by three. This is one possible outcome.

The way I'm describing things, this outcome is to be distinguished from this outcome here, where a three is followed by two. If you're playing backgammon, it doesn't matter which one of the two happened. But if you're dealing with a probabilistic model that you want to keep track of everything that happens in this composite experiment, there are good reasons for distinguishing between these two outcomes. I mean, when this happens, it's definitely something different from that happening. A two followed by a three is different from a three followed by a two.

So this is the correct sample space for this experiment where we roll the die twice. It has a total of 16 elements and it's, of course, a finite set.

Sometimes, instead of describing sample spaces in terms of lists, or sets, or diagrams of this kind, it's useful to describe the experiment in some sequential way. Whenever you have an experiment that consists of multiple stages, it might be useful, at least visually, to give a diagram that shows you how those stages evolve. And that's what we do by using a sequential description or a tree-based description by drawing a tree of the possible evolutions during our experiment.

So in this tree, I'm thinking of a first stage in which I roll the first die, and there are four possible results, one, two, three and four.and 4. And, given what happened, let's say in the first roll, suppose I got a one. Then I'm rolling the second dice, and there are four possibilities for what may happen to the second die. And the possible results are one, tow, three and four again.

So what's the relation between the two diagrams? Well, for example, the outcome two followed by three corresponds to this path on the tree. So this path corresponds to two followed by a three. Any path is associated to a particular outcome, any outcome is associated to a particular path.

And, instead of paths, you may want to think in terms of the leaves of this diagram. Same thing, think of each one of the leaves as being one possible outcome. And of course we have 16 outcomes here, we have 16 outcomes here.

Maybe you noticed the subtlety that I used in my language. I said I rolled the first dice and the result that I get is a two. I didn't use the word "outcome." I want to reserve the word "outcome" to mean the overall outcome at the end of the overall experiment.

So "2, 3" is the outcome of the experiment. The experiment consisted of stages. Two was the result in the first stage, three was the result in the second stage. You put all those results together, and you get your outcome. OK, perhaps we are splitting hairs here, but it's useful to keep the concepts right.

What's special about this example is that, besides being trivial, it has a sample space which is finite. There's 16 possible total outcomes. Not every experiment has a finite sample space.

Here's an experiment in which the sample space is infinite. So you are playing darts and the target is this square. And you're perfect at that game, so you're sure that your darts will always fall inside the square. So, but where exactly your dart would fall inside that square, that itself is random. We don't know what it's going to be. It's uncertain.

So all the possible points inside the square are possible outcomes of the experiment. So a typical outcome of the experiment is going to a pair of numbers, x,y, where x and y are real numbers between zero and one. Now there's infinitely many real numbers, there's infinitely many points in the square, so this is an example in which our sample space is an infinite set. OK, so we're going to revisit this example a little later.

So these are two examples of what the sample space might be in simple experiments. Now, the more important order of business is now to look at those possible outcomes and to make some statements about their relative likelihoods. Which outcome is more likely to occur compared to the others? And the way we do this is by assigning probabilities to the outcomes.

Well, not exactly. Suppose that all you were to do was to assign probabilities to individual outcomes. If you go back to this example, and you consider one particular outcome-- let's say this point-- what would be the probability that you hit exactly this point to infinite precision? Intuitively, that probability would be zero. So any individual point in this diagram in any reasonable model should have zero probability. So if you just tell me that any individual outcome has zero probability, you're not really telling me much to work with.

For that reason, what instead we're going to do is to assign probabilities to subsets of the sample space, as opposed to assigning probabilities to individual outcomes. So here's the picture. We have our sample space, which is omega, and we consider some subset of the sample space. Call it A. And I want to assign a number, a numerical probability, to this particular subset which represents my belief about how likely this set is to occur.

OK. What do we mean "to occur?" And I'm introducing here a language that's being used in probability theory. When we talk about subsets of the sample space, we usually call them events, as opposed to subsets. And the reason is because it works nicely with the language that describes what's going on.

So the outcome is a point. The outcome is random. The outcome may be inside this set, in which case we say that event A occurred, if we get an outcome inside here. Or the outcome may fall outside the set, in which case we say that event A did not occur.

So we're going to assign probabilities to events. And now, how should we do this assignment? Well, probabilities are meant to describe your beliefs about which sets are more likely to occur versus other sets. So there's many ways that you can assign those probabilities. But there are some ground rules for this game.

First, we want probabilities to be numbers between zero and one because that's the usual convention. So a probability of zero means we're certain that something is not going to happen. Probability of one means that we're essentially certain that something's going to happen. So we want numbers between zero and one.

We also want a few other things. And those few other things are going to be encapsulated in a set of axioms. What "axioms" means in this context, it's the ground rules that any legitimate probabilistic model should obey. You have a choice of what kind of probabilities you use. But, no matter what you use, they should obey certain consistency properties because if they obey those properties, then you can go ahead and do useful calculations and do some useful reasoning.

So what are these properties? First, probabilities should be non-negative. OK? That's our convention. We want probabilities to be numbers between zero and one. So they should certainly be non-negative. The probability that event A occurs should be a non-negative number.

What's the second axiom? The probability of the entire sample space is equal to one. Why does this make sense? Well, the outcome is certain to be an element of the sample space because we set up a sample space, which is collectively exhaustive. No matter what the outcome is, it's going to be an element of the sample space. We're certain that event omega is going to occur. Therefore, we represent this certainty by saying that the probability of omega is equal to one.

Pretty straightforward so far. The more interesting axiom is the third rule. Before getting into it, just a quick reminder.

If you have two sets, A and B, the intersection of A and B consists of those elements that belong both to A and B. And we denote it this way. When you think probabilistically, the way to think of intersection is by using the word "and." This event, this intersection, is the event that A occurred and B occurred. If I get an outcome inside here, A has occurred and B has occurred at the same time. So you may find the word "and" to be a little more convenient than the word "intersection."

And similarly, we have some notation for the union of two events, which we write this way. The union of two sets, or two events, is the collection of all the elements that belong either to the first set, or to the second, or to both. When you talk about events, you can use the word "or." So this is the event that A occurred or B occurred. And this "or" means that it could also be that both of them occurred.

OK. So now that we have this notation, what does the third axiom say? The third axiom says that if we have two events, A and B, that have no common elements-- so here's A, here's B, and perhaps this is our big sample space. The two events have no common elements. So the intersection of the two events is the empty set. There's nothing in their intersection. Then, the total probability of A together with B has to be equal to the sum of the individual probabilities. So the probability that A occurs or B occurs is equal to the probability that A occurs plus the probability that B occurs.

So think of probability as being cream cheese. You have one pound of cream cheese, the total probability assigned to the entire sample space. And that cream cheese is spread out over this set. The probability of A is how much cream cheese sits on top of A. Probability of B is how much sits on top of B. The probability of A union B is the total amount of cream cheese sitting on top of this and that, which is obviously the sum of how much is sitting here and how much is sitting there.

So probabilities behave like cream cheese, or they behave like mass. For example, if you think of some material object, the mass of this set consisting of two pieces is obviously the sum of the two masses. So this property is a very intuitive one. It's a pretty natural one to have.

OK. Are these axioms enough for what we want to do? I mentioned a while ago that we want probabilities to be numbers between zero and one. Here's an axiom that tells you that probabilities are non-negative. Should we have another axiom that tells us that probabilities are less than or equal to one? It's a desirable property. We would like to have it in our hands.

OK, why is it not in that list? Well, the people who are in the axiom making business are mathematicians and mathematicians tend to be pretty laconic. You don't say something if you don't have to say it. And this is the case here. We don't need that extra axiom because we can derive it from the existing axioms.

Here's how it goes. One is the probability over the entire sample space. Here we're using the second axiom. Now the sample space consists of A together with the complement of A. OK? When I write the complement of A, I mean the complement of A inside of the set omega. So we have omega, here's A, here's the complement of A, and the overall set is omega.

OK. Now, what's the next step? What should I do next? Which axiom should I use? We use axiom three because a set and the complement of that set are disjoint. They don't have any common elements. So axiom three applies and tells me that this is the probability of A plus the probability of A complement. In particular, the probability of A is equal to one minus the probability of A complement, and this is less than or equal to one.

Why? Because probabilities are non-negative, by the first axiom.

OK. So we got the conclusion that we wanted. Probabilities are always less than or equal to one, and this is a simple consequence of the three axioms that we have. This is a really nice argument because it actually uses each one of those axioms. The argument is simple, but you have to use all of these three properties to get the conclusion that you want.

OK. So we can get interesting things out of our axioms. Can we get some more interesting ones? How about the union of three sets? What kind of probability should it have?

So here's an event consisting of three pieces. And I want to say something about the probability of A union B union C. What I would like to say is that this probability is equal to the sum of the three individual probabilities. How can I do it?

I have an axiom that tells me that I can do it for two events. I don't have an axiom for three events. Well, maybe I can massage things and still be able to use that axiom. And here's the trick. The union of three sets, you can think of it as forming the union of the first two sets and then taking the union with the third set. OK? So taking unions, you can take the unions in any order that you want.

So here we have the union of two sets. Now, ABC are disjoint, by assumption or that's how I drew it. So if A, B, and C are disjoint, then A union B is disjoint from C. So here we have the union of two disjoint sets. So by the additivity axiom, the probability of that the union is going to be the probability of the first set plus the probability of the second set.

And now I can use the additivity axiom once more to write that this is probability of A plus probability of B plus probability of C. So by using this axiom which was stated for two sets, we can actually derive a similar property for the union of three disjoint sets. And then you can repeat this argument as many times as you want. It's valid for the union of ten disjoint sets, for the union of a hundred disjoint sets, for the union of any finite number of sets. So if A1 up to An are disjoint, then the probability of A1 union An is equal to the sum of the probabilities of the individual sets.

OK. Special case of this is when we're dealing with finite sets. Suppose I have just a finite set of outcomes. I put them together in a set and I'm interested in the probability of that set. So here's our sample space. There's lots of outcomes, but I'm taking a few of these and I form a set out of them.

This is a set consisting of, in this picture, three elements. In general, it consists of k elements. Now, a finite set, I can write it as a union of single element sets. So this set here is the union of this one element set, together with this one element set together with that one element set. So the total probability of this set is going to be the sum of the probabilities of the one element sets.

Now, probability of a one element set, you need to use the brackets here because probabilities are assigned to sets. But this gets kind of tedious, so here one abuses notation a little bit and we get rid of those brackets and just write probability of this single, individual outcome. In any case, conclusion from this exercise is that the total probability of a finite collection of possible outcomes, the total probability is equal to the sum of the probabilities of individual elements.

So these are basically the axioms of probability theory. Or, well, they're almost the axioms. There are some subtleties that are involved here.

One subtlety is that this axiom here doesn't quite do the job for everything we would like to do. And we're going to come back to this at the end of the lecture. A second subtlety has to do with weird sets.

We said that an event is a subset of the sample space and we assign probabilities to events. Does this mean that we are going to assign probability to every possible subset of the sample space? Ideally, we would wish to do that. Unfortunately, this is not always possible.

If you take a sample space, such as the square, the square has nice subsets, those that you can describe by cutting it with lines and so on. But it does have some very ugly subsets, as well, that are impossible to visualize, impossible to imagine, but they do exist. And those very weird sets are such that there's no way to assign probabilities to them in a way that's consistent with the axioms of probability.

OK. So this is a very, very fine point that you can immediately forget for the rest of this class. You will only encounter these sets if you end up doing doctoral work on the theoretical aspects of probability theory. So it's just a mathematical subtlety that some very weird sets do not have probabilities assigned to them. But we're not going to encounter these sets and they do not show up in any applications.

OK. So now let's revisit our examples. Let's go back to the die example. We have our sample space. Now we need to assign a probability law. There's lots of possible probability laws that you can assign. I'm picking one here, arbitrarily, in which I say that every possible outcome has the same probability of 1/16.

OK. Why do I make this model? Well, empirically, if you have well-manufactured dice, they tend to behave that way. We will be coming back to this kind of story later in this class. But I'm not saying that this is the only probability law that there can be. You might have weird dice in which certain outcomes are more likely than others. But to keep things simple, let's take every outcome to have the same probability of 1/16.

OK. Now that we have in our hands a sample space and the probability law, we can actually solve any problem there is. We can answer any question that could be posed to us. For example, what's the probability that the outcome, which is this pair, is either 1,1 or 1,2. We're talking here about this particular event, 1,1 or 1,2. So it's an event consisting of these two items.

According to what we were just discussing, the probability of a finite collection of outcomes is the sum of their individual probabilities. Each one of them has probability of 1/16, so the probability of this is 2/16.

How about the probability of the event that x is equal to one. x is the first roll, so that's the probability that the first roll is equal to one. Notice the syntax that's being used here. Probabilities are assigned to subsets, to sets, so we think of this as meaning the set of all outcomes such that x is equal to one.

How do you answer this question? You go back to the picture and you try to visualize or identify this event of interest. x is equal to one corresponds to this event here. These are all the outcomes at which x is equal to one. There's four outcomes. Each one has probability 1/16, so the answer is 4/16.

OK. How about the probability that x plus y is odd? OK. That will take a little bit more work.

But you go to the sample space and you identify all the outcomes at which the sum is an odd number. So that's a place where the sum is odd, these are other places, and I guess that exhausts all the possible outcomes at which we have an odd sum. We count them. How many are there? There's a total of eight of them. Each one has probability 1/16, total probability is 8/16.

And harder question. What is the probability that the minimum of the two rolls is equal to 2? This is something that you probably couldn't do in your head without the help of a diagram. But once you have a diagram, things are simple.

You ask the question. OK, this is an event, that the minimum of the two rolls is equal to two. This can happen in several ways. What are the several ways that it can happen? Go to the diagram and try to identify them.

So the minimum is equal to two if both of them are two's. Or it could be that x is two and y is bigger, or y is two and x is bigger. OK. I guess we rediscover that yellow and blue make green, so we see here that there's a total of five possible outcomes. The probability of this event is 5/16.

Simple example, but the procedure that we followed in this example actually applies to any probability model you might ever encounter. You set up your sample space, you make a statement that describes the probability law over that sample space, then somebody asks you questions about various events. You go to your pictures, identify those events, pin them down, and then start kind of counting and calculating the total probability for those outcomes that you're considering.

This example is a special case of what is called the discrete uniform law. The model obeys the discrete uniform law if all outcomes are equally likely. It doesn't have to be that way. That's just one example of a probability law.

But when things are that way, if all outcomes are equally likely and we have N of them, and you have a set A that has little n elements, then each one of those elements has probability one over capital N since all outcomes are equally likely. And for our probabilities to add up to one, each one must have this much probability, and there's little n elements. That gives you the probability of the event of interest.

So problems like the one in the previous slide and more generally of the type described here under discrete uniform law, these problems reduce to just counting. How many elements are there in my sample space? How many elements are there inside the event of interest? Counting is generally simple, but for some problems it gets pretty complicated. And in a couple of weeks, we're going to have to spend the whole lecture just on the subject of how to count systematically.

Now the procedure we followed in the previous example is the same as the procedure you would follow in continuous probability problems. So, going back to our dart problem, we get the random point inside the square. That's our sample space. We need to assign a probability law. For lack of imagination, I'm taking the probability law to be the area of a subset.

So if we have two subsets of the sample space that have equal areas, then I'm postulating that they are equally likely to occur. The probably that they fall here is the same as the probability that they fall there. The model doesn't have to be that way. But if I have sort of complete ignorance of which points are more likely than others, that might be the reasonable model to use.

So equal areas mean equal probabilities. If the area is twice as large, the probability is going to be twice as big. So this is our model.

We can now answer questions. Let's answer the easy one. What's the probability that the outcome is exactly this point? That of course is zero because a single point has zero area. And since this probability is equal to area, that's zero probability.

How about the probability that the sum of the coordinates of the point that we got is less than or equal to 1/2? How do you deal with it? Well, you look at the picture again, at your sample space, and try to describe the event that you're talking about. The sum being less than 1/2 corresponds to getting an outcome that's below this line, where this line is the line where x plus y equals to 1/2. So the intercepts of that line with the axis are 1/2 and 1/2.

So you describe the event visually and then you use your probability law. The probability law that we have is that the probability of a set is equal to the area of that set. So all we need to find is the area of this triangle, which is 1/2 times 1/2 times 1/2, half, equals to 1/8.

OK. Moral from these two examples is that it's always useful to have a picture and work with a picture to visualize the events that you're talking about. And once you have a probability law in your hands, then it's a matter of calculation to find the probabilities of an event of interest. The calculations we did in these two examples, of course, were very simple.

Sometimes calculations may be a lot harder, but it's a different business. It's a business of calculus, for example, or being good in algebra and so on. As far as probability is concerned, it's clear what you will be doing, and then maybe you're faced with a harder algebraic part to actually carry out the calculations. The area of a triangle is easy to compute. If I had put down a very complicated shape, then you might need to solve a hard integration problem to find the area of that shape, but that's stuff that belongs to another class that you have presumably mastered by now.

Good, OK. So now let me spend just a couple of minutes to return to a point that I raised before. I was saying that the axiom that we had about additivity might not quite be enough. Let's illustrate what I mean by the following example.

Think of the experiment where you keep flipping a coin and you wait until you obtain heads for the first time. What's the sample space of this experiment? It might happen the first flip, it might happen in the tenth flip. Heads for the first time might occur in the millionth flip.

So the outcome of this experiment is going to be an integer and there's no bound to that integer. You might have to wait very much until that happens. So the natural sample space is the set of all possible integers.

Somebody tells you some information about the probability law. The probability that you have to wait for n flips is equal to two to the minus n. Where did this come from? That's a separate story. Where did it come from? Somebody tells this to us, and those probabilities are plotted here as a function of n.

And you're asked to find the probability that the outcome is an even number. How do you go about calculating that probability? So the probability of being an even number is the probability of the subset that consists of just the even numbers. So it would be a subset of this kind, that includes two, four, and so on.

So any reasonable person would say, well the probability of obtaining an outcome that's either two or four or six and so on is equal to the probability of obtaining a two, plus the probability of obtaining a four, plus the probability of obtaining a six, and so on. These probabilities are given to us. So here I have to do my algebra. I add this geometric series and I get an answer of 1/3. That's what any reasonable person would do.

But the person who only knows the axioms that they posted just a little earlier may get stuck. They would get stuck at this point. How do we justify this?

We had this property for the union of disjoint sets and the corresponding property that tells us that the total probability of finitely many things, outcomes, is the sum of their individual probabilities. But here we're using it on an infinite collection. The probability of infinitely many points is equal to the sum of the probabilities of each one of these. To justify this step we need to introduce one additional rule, an additional axiom, that tells us that this step is actually legitimate.

And this is the countable additivity axiom, which is a little stronger, or quite a bit stronger, than the additivity axiom we had before. It tells us that if we have a sequence of sets that are disjoint and we want to find their total probability, then we are allowed to add their individual probabilities. So the picture might be such as follows.

We have a sequence of sets, A1, A2, A3, and so on. I guess in order to fit them inside the sample space, the sets need to get smaller and smaller perhaps. They are disjoint. We have a sequence of such sets. The total probability of falling anywhere inside one of those sets is the sum of their individual probabilities.

A key subtlety that's involved here is that we're talking about a sequence of events. By "sequence" we mean that these events can be arranged in order. I can tell you the first event, the second event, the third event, and so on. So if you have such a collection of events that can be ordered as first, second, third, and so on, then you can add their probabilities to find the probability of their union.

So this point is actually a little more subtle than you might appreciate at this point, and I'm going to return to it at the beginning of the next lecture. For now, enjoy the first week of classes and have a good weekend. Thank you.