PROFESSOR: OK, we're all ready to go. This is Discrete Stochastic Processes as you are know. It is-- want to get it where I can read it, too.

We're going to try to deal with a bunch of different topics today, some of which are a little bit philosophical saying, what is probability really? You are supposed to have taken a course in probability, but unfortunately courses in probability are almost always courses in how to solve well-posed problems. The big problem in probability theory, and particularly stochastic processes is not so much how do you solve well-posed problems. Anybody can do that. Or anybody who has a little bit of background can do it.

The hard problem is finding the right models for a real-world problem. I will call these real-world problems. I hate people who call things real-world because they sound like they dislike theory or something. It's the only word I can think of though, because physical world is no longer adequate because so much of the applications of probability are to all sorts of different things, not having to do with the physical world very much, but having to do with things like the business world, or the economic world, or the biological world, or all of these things. So real-world is just a code word we'll use to distinguish theory from anything real that you might have to deal with.

Theory is very nice because theory-- everything is specified. There's a right answer to everything. There is a wrong answer usually. But there's at least one right answer. And most of us like those things because they're very specific. People go into engineering and science very often because they don't like the uncertainty of a lot of other fields.

The problem is as soon as you go into probability theory, you're moving away from that safe region where everything is specified, and you're moving into a region where things, in fact, are not very well-specified and you have to be careful about it.

OK, so first we're going to talk about probability in the real world and probability as a branch of mathematics. Then we're going to say what discrete stochastic processes are. Then we're going to talk just a very, very little bit about the processes we're going to study. If you want to see more of that, you have two chapters of the notes. You can look at them. You can look at the table of contents. And more than that, if you look at my website, you will see the notes for all the other chapters if you want to read ahead or if you want to really find out what kinds of things we're going to talk about and what kinds of things we're not going to talk about.

Then we're going to talk about when, where, and how is this useful? The short answer to that is it's useful everywhere. But we'll have to see why that is. Then we're going to talk about the axioms of probability theory. You cannot take any elementary course in probability, or even in statistics, without seeing the axioms of probability. And in almost all of those cases, and in almost all of the graduate courses I've seen, you see them, they disappear, and suddenly you're solving problems in whatever way you can, and the axioms have nothing to do with it anymore. So we're going to see that, in fact, the axioms do have something to do with this.

Those of you who want to be real engineers and not mathematicians, you'll find this a little uncomfortable at times. We are going to be proving things. And you will have to get used to that. And I'll try to convince you of why it's important to be able to prove things.

Then we're going on to a review of probability, independent events, experiments, and random variables. So that's what we'll do today.

Incidentally, this course started about-- must've been 25 years ago or so. I started it because we had a huge number of students at MIT who had been interested in communication and control, and who were suddenly starting to get interested in networks. And there were all sorts of queuing problems that they had to deal with every day. And they started to read about queuing theory. and it was the most disjointed, crazy theory in the world where there were 1,000 different kinds of queues and each one of them had to be treated in its own way. And we realized that stochastic processes was the right way to tie all of that together, so we started this course. And we made it mostly discrete so it would deal primarily with network type applications. As things have grown, it now deals with a whole lot more applications. And we'll see how that works later on.

OK, how did probability get started in the real world? Well, there were games of chance that everybody was interested in. People really like to gamble. I don't know why. I don't like to gamble that much. I would rather be certain about things. But most people love to gamble. And most people have an intuitive sense of what probability is about.

I mean, eight-year-old kids, when they start to learn to play games of chance-- and there are all sorts of board games that involve chance. These kids, if they're bright-- and I'm sure you people fall into that category-- they immediately start to figure out what the odds are. I mean, how many of you have never thought about what the odds are in some gambling game?

OK, that makes my point. So all of you understand this at an intuitive level. But what makes games of chance easier to deal with than all the other issues where we have uncertainty in life?

Well, games of chance are inherently repeatable. You play a game of chance and you play many, many hands, or many, many throws, or many, many trials. And after many, many trials of what's essentially the same experiment, you start to get a sense of what relative frequencies are. You start to get a sense of what the odds are because of doing this repeatedly. So games of chance are easy to use probability on because they are repeatable. You have essentially the same thing going on each time, but each time there's a different answer. You flip a coin and sometimes it comes up heads and sometimes it comes up tails. So in fact, we have to figure out how to deal with that fact that there is uncertainty there. I'll talk about that in just another minute.

But anyway, most of life's decisions involve uncertainty. I mean, for all of you, when you go into a PhD program, you have two problems. Am I going to enjoy this? And you don't know whether you're going to enjoy it because not until you really get into it do you have a sense of whether this set of problems you're dealing with is something that you like to deal with.

And the only way you can do that is to make guesses. You come up with likelihoods. There's some likelihood. There's a risk cost-benefit that you deal with. And in life, risk cost-benefits are always based on some sense of what the likelihood of something is.

Now, what is a likelihood? A likelihood is just a probability. It's a synonym for probability. When you get into the mathematics of probability, likelihood has a special meaning to it. But in the real world, likelihood is just a word you use when you don't want to let people know that you're really talking about probabilities.

OK, so that's where we are. But on the last slide, you saw the word "essentially, essentially, essentially." If you read the notes, and I hope you read the notes, because I spent the last three years doing virtually nothing but trying to make these notes clear. I would appreciate it if any of you, with whatever background you have, when you read these nodes, if you read them twice and you still don't understand something, tell me you don't understand it. If you know why you don't understand it, I'd appreciate it knowing that. But just saying "help" is enough to let me know that I still haven't made something as clear as it should be. At least as clear as it should be for some of the people who I think should be taking this course.

One of the problems we have at MIT now, and at every graduate school in the world I think, is that human knowledge has changed and grown so much in the last 50 years. So when you study something general, like probability, there's just an enormous mass of stuff you have to deal with. And because of that, when you try to write notes for a course, you don't know what anybody's background is anymore.

I mean, it used to be that when you saw a graduate course at MIT, people would know what limits were. People would know what basic mathematics is all about. They would know what continuity means. They would know some linear algebra. They would know all sorts of different things. Many people still do know those things. Many other people have studied all sorts of other fascinating and very interesting things. They're just as smart, but they have a very different background. So if your background is different, it's not your fault that you don't have the kind of background that makes probability easy. Just yell. Or yell in class. Please ask questions. The fact that we're videotaping this makes it far more interesting for anybody who's using OpenCourseWare to see some kinds of questions going on, so I very much encourage that.

I'm fairly old at this point, and my memory is getting shot. So if you ask a question and I don't remember what it's all about, just be patient with it. I will come back the next time, or I'll send you an email straightening it out. But I will often get confused doing something, and that's just because of my age. It's what we call "senior moments." It's not that I don't understand the subject. I think I understand it, I just don't remember it anymore.

Important point about probability. Think about flipping a coin. I'm going to talk about flipping coins a great deal this term. It's an absolutely trivial topic, but it's important because when you understand deep things about a large subject, the best way to understand them is to understand them in terms of the most trivial examples you can think of.

Now, when you flip a coin, the outcome-- heads or tails-- really depends on the initial velocity, the orientation of the person flipping it, or the machine flipping it, the coin surfaces, the ground surface. And after you put all of those things into a careful equation, you will know whether the coin is going to come up heads or tails.

I don't think quantum theory has anything to do with something as big as a coin. I might be wrong. I've never looked into it. And frankly, I don't care. Because the point that I'm trying to make is that flipping a coin, and many of the things that you view as random, when you look at them in a slightly different way, are not random.

There's a big field in communication called data compression. And data compression is based on random models for the data, which is going to be compressed.

Now, what I'm saying here today is by no means random. Or maybe it is partly random. Maybe it's coming out of a random mind, I don't know. But all of the data we try to compress to the people who have created that data, it's not random at all. If you study the data carefully enough-- I mean, code breakers and people like that are extremely good at sorting out what the meaning is in something, which cannot be done by data compression techniques at all.

So the point is, when you're doing data compression, you model the data as being random and having certain characteristics. But it really isn't. So the model is no good.

When you get to more important questions-- well, data compression is an important question. When you ask, what's the probability of another catastrophic oil spill in the next year? Or you ask the question, what's the probability that Google stock will double in five years? That's less important, but it's still important to many people. How do you model that? Understanding probability theory, understanding all the mathematics of this is not going to help you model this.

Now, why do I make such a big deal about this? Well, there have been a number of times in the last 10 or 15 years when the whole financial system of the world has almost been destroyed by very, very bright PhDs. Many of them coming from electrical engineering. Most of whom are really superb at understanding probability theory. And they have used their probability theory to analyze risk and other things in investments. And what has happened?

They do very well for a while. Suddenly they do so well that they think they can borrow all sorts of money and risk other people's money as well as their own. In fact, they try to do that right from the beginning. And then suddenly, the whole thing collapses. Because their models are no damn good. There's nothing wrong with their mathematics, it's that their models are no good.

So please, especially if you're going to do something important in your lives-- if you're just going to write papers in engineering journals, maybe it's all right. But if you're going to make decisions about things, please spend some time thinking about the probability models that you come up because this is vitally important.

OK, what's probability? It's a branch of mathematics. Now we're into something that's more familiar, something that's simpler, something we can deal with.

You might be uncomfortable with what probability really means. And all probability books, all stochastic process books are uncomfortable with this. Feller is the best book in probability there's ever been written. Any question you have, he probably has the answer to it.

When you look at what he says about real-world probability, the modeling issues, he's an extraordinarily bright guy. And he spent some time thinking about this. But you read it and you realize that it's pure nonsense. So please, take my word for it. Don't assume that real-world probability is something you're going to learn about from other people because you can't trust what any of them say. It's something you have to think through for yourselves, and we'll talk more about this as we go.

But now, when we get into mathematics, that's fine. We just create models. And once we have the model, we just use it.

We have standard models for all sorts of different standard problems. When you talk about coin tossing, what almost everyone means is not this crazy thing I was just talking about where you have an initial angular momentum when you flip a coin and all of that stuff. It's a purely mathematical model where a coin is flipped and with probability one half it comes up heads and with probability one half it comes up tails.

OK, students are given a well-specified model, and they calculate various things. This is in mathematical probability. Heads and tails are equiprobable in that system. Subsequent tosses are independent. Here's a little bit of cynicism. I apologize for insulting you people with it. I apologize to any faculty member who later reads this. And I particularly apologize to businessmen and government people who might read it.

Students compute, professors write papers, business and government leaders obtain questionable models and data on which they can blame failures.

Most cynical towards business leaders because business leaders often hire consultants. Not so much to learn what to do, but so they have excuses when what they do doesn't work out right.

When I say the students compute, what I mean is this in almost all the courses you've taken up until now-- and in this course also-- what you're going to be doing is solving well-posed problems. You solve well-posed exercises because that's a good way to understand what the mathematics of the subject is about.

Don't think that that's the only part of it. If that's the only thing you're doing, you might as well not waste your time. You might as well do something else. You might as well go out and shovel snow today instead of trying to learn about probability theory. It's more pleasant to learn about probability theory.

OK, the use of probability models has two major problems with it. The first problem is, how do you make a model for a real-world problem? And a partial answer is, learn about estimation and decisions in the context of standard models. In other words, decisions and estimation inside a completely mathematical framework.

Then you learn a great deal about the real-world problem itself. Not about the mathematics of it, but about how you actually understand what's going on. If you talk to somebody who is a superb architect in any field-- networks, computer systems, control systems, anything-- what are you going to find?

You're not going to find huge, involved sets of equations that they're going to use to explain something to you. They're going to pick at-- if there any good, they're going to take this big problem, and they're going to take your issue with this big problem. And they're going to find the one or two really important things that tell you something that you have to know. And that's what you want to get out of this course. You want to get the ability to take all of the chat, put it all together, and be able to say one or two important things which is really necessary. That's where you're going to.

Before you get there, you'll take low-level jobs in various companies and you'll compute a lot of things. You'll simulate a lot of things. You'll deal with a lot of detail. Eventually, you're going to get to the point where you've got to make major decisions. And you want to be ready for it. OK, that's enough philosophy. I will try to give no more philosophy today, except when I get pushed into it.

OK, one of the problems in this problem of finding a good model is that no model is perfect. Namely, what happens is you keep finding more and more complicated models, which deal with more and more of the issues. And as you deal with them, things get more complicated. You're more down in the level of details and you're finding out less. So you want to find some sort of match between a model that tells you something and a model which is complicated enough to deal with the issues.

There's a beautiful quote by Alfred North Whitehead. I don't know whether you've ever heard of Whitehead. You've probably heard of Bertrand Russell, who was both a great logician and a great philosopher, and had a lot to do with the origins of set theory. Whitehead and Russell together, wrote this massive book around the turn of the last century between the 1900s and the 2000s called *Principia Mathematica* where they try to resolve all of the paradoxes which were coming up in mathematics.

And Whitehead's general philosophical comment was, "Seek simplicity and distrust it."

Now, every time I look at that, I say, why in hell didn't he say, seek simplicity and question it? I mean, you all hear about questioning authority, of course, and that's important to do. Why when you find a simple model for something should you distrust it?

Well, the reason is psychological. If you find a simple model for something and you question it, you have an enormous psychological bias towards not giving up the simple model. You want to keep that simple model. And therefore, it takes an enormous amount of evidence before you're going to give something out.

Whitehead said something more than that. He said, "Seek simplicity and distrust it."

Now, why do I talk about the philosophy of science when we're trying to learn about probability theory? Well, probability theory is a mathematical theory. It's the basis for a great deal of science. And it's the place where modeling is most difficult. Scientific questions in most areas, if there's no probability or uncertainty involved, you just do an experiment that tells you the answer. You might not do it carefully enough and then 10 other people do it. And finally, everybody agrees, this is the answer to that problem. In probability, it ain't that simple. And that's why one has to focus on this a little more than usual.

The second problem is, how do you make a probability model that has no hidden paradoxes in it? In other words, when you make a mathematical model, how do you make sure that it really is well-posed? How do you make sure that when you solve a problem in that mathematical model that you don't come to something that doesn't make any sense?

Well, everyone's answer to that is you use Kolmogorov's axioms of probability. Because back in 1933, Kolmogorov published this little thin book. Those of you who are interested in the history of science probably ought to read it. You will find you only understand the first five pages the first time you look at it. But it's worthwhile doing that because here was one of the truly great minds of the early 20th century.

And he took everything he knew about probability, which was a whole lot more than I know certainly, and a whole lot more than anybody else at the time knew, and he collapsed it into these very simple axioms. And he said, if you obey these axioms in a model that you use in probability, those axioms will keep you out of any paradoxes at all. And then, he showed why that was and he showed how the axioms could be used and so forth. So we're going to spend a little bit of time talking about them today.

OK, quickly, what is a discrete stochastic process? Well, a stochastic process-- you've been talking about probability. And you might be getting the idea that I'm just using the name "stochastic processes" as a foil for talking about what I really love, which is the probability. And there's a certain amount of truth to that.

But stochastic processes are special types of probability models where the sample points represent functions in time. In other words, when we're dealing with a probability model, the basis of a probability model is a sample space. It's the set of possible things that might happen. And you can reduce that to the sample points, which are the indivisible, little, tiny crumbs of what happens when you do an experiment. It's the thing which specifies everything that can be specified in that model of that experiment.

OK, when you get to a stochastic process, what you're doing is you're looking at a situation in which these sample points, the solutions to what happens is, in fact, a whole sequence of random variables in time. And what that means is instead of looking at just a vector of random variables, you're going to be looking at a whole sequence of random variables.

Now, what is different about a vector of a very large number of random variables and an infinite sequence of random variables? Well, from an engineering standpoint, not very much. I mean, there's nothing you can do to actually look at an infinite sequence of random variables. If you start out at the Big Bang and you carry it on to what you might imagine is the time when the sun explodes or something, that's a finite amount of time. And if you imagine how fast you can observe things, there's a finite number of random variables you might observe.

All these models we're going to be dealing with get outside or that realm, and they deal with something that starts infinitely far in the past and goes infinitely far in the future. It doesn't make much sense, does it?

But then look at the alternative. You built a device which you're going to sell to people, and which they're going to use. And you know they're only going to use it three or four year until something better comes along. But do you want to build in to everything you're doing the idea that it's going to be obsolete in three years?

No. You want to design this thing so, in fact, it will work for an essentially arbitrary amount of time. And therefore, you make a mathematical model of it. You look at what happens over an infinite span of time. So whenever we get into mathematics, we always go to an infinite number of things rather than a finite number of things.

Now, discrete stochastic processes are those where the random variables are discrete in time. Namely, a finite number of possible outcomes from each of them. Or the set of possible sample values is discrete. What does that mean?

It doesn't mean a whole lot when you really start asking detailed questions about this. What it means is, I want to talk about a particular kind of stochastic processes. And it's a class of processes which will be more than we can deal with in one term. And I want to exclude certain processes, like noise processes, because we don't have time to do both of them. So don't worry too much about exactly what a discrete stochastic process is. It's whatever we want to call it when we deal with it.

Oops. Oh, where am I? Oh, I wanted to talk about the different processes we're going to study.

The first kind of process is what we call a counting process. The sample points in the process-- remember, a sample point specifies everything about an experiment. It tells you every little detail. And the sample points here in counting processes are sequences of arrivals. This is a very useful idea in dealing with queuing systems because queuing systems have arrivals. They have departures. They have rules for how the arrivals get processed before they get spit out. And a big part of that is studying first the arrival process, then we study the departure process. We study how to put them together. And when we get to chapter 2 of the notes, we're going to be studying Poisson processes, which are in a sense, the perfect discrete stochastic process. It's like coin tossing in probability. Everything that might be true with a Poisson process is true. The only things that aren't true are the things that obviously can't be true. And we'll find out why that is and how that works a little later.

We're then going to study renewal processes in chapter 4. We're going to put Markov chains in between. And you'll see why when we do it. And renewal processes are a more complicated kind of thing than Poisson processes. And there's no point confusing you at this point saying what the difference is, so I won't.

Markov processes are processes. In other words, the sequences in time of things where what happens in the future depends on what happens in the past, only through the state at the present. In other words, if you can specify the state in the present, you can forget about everything in the past.

If you have those kinds of processes around, you don't have to study history at all, which would be very nice. But unfortunately, not all processes behave that way. When you do the modeling to try to find out what the state is, which is what you have to know at the present, you find out there's a lot of history involved.

OK, finally, we're going to talk about random wa;ls and martingales. I'm not going to even say what a random walks or a martingale is. We will find out about that soon enough, but I want to tell you that's what's in chapter 7 of the notes. That's the last topic we will deal with.

We'll study all sorts of mixtures of these. Things which involve a little bit of each. We'll start out working on one thing and we'll find out another. One of these other topics is the right way to look at it.

If you want to know more about that, please go look at the notes, and you'll find out as much as you want. But it's not appropriate to talk about it right now.

OK, when, where, and how is this useful? You see, I'm almost at the point where we'll start actually talking about real stuff. And when I say real stuff, I mean mathematical stuff, which is not real stuff.

Broad answer-- probability in stochastic processes are an important adjunct to rational thought about all human and scientific endeavor. That's a very strong statement. I happen to believe it. You might not believe it. And you're welcome to not believe it. It won't be on a quiz or anything, believe me. But almost anything you have to deal with is dealing with something in the future. I mean, you have to plan for things which are going to happen in the future.

When you look at the future, there's a lot of uncertainty involved with it. One of the ways is dealing with uncertainty. And probably the only scientific way of dealing with uncertainty is through the mechanism of probability models. So anything you want to deal with, which is important, you're probably better off knowing something about probability than not.

A narrow answer is probability in stochastic processes are essential components of the following areas. Now, I must confess I made up this list in about 10 minutes without thinking about it very seriously. And these things are related to each other. Some of them are parts of others. Let me read them.

Communication systems and networks. That's where I got involved in this question, and very important there. Computer systems. I also got involved in it because of computer systems. Queuing in all areas. Well, I got involved in queuing because of being interested in networks. risk management in all areas. I got interested in that because I started to get disturbed about civilization destroying itself because people who have a great deal of power don't know anything about taking risks.

OK, catastrophe management. How do you prevent oil spills and things like that? How do you prevent nuclear plants from going off? How do you prevent nuclear weapons from falling in the hands of the wrong people? These again, are probability issues. These are important probability issues because most people don't regard them as probability issues.

If you say there is one chance in a billion that something will happen, 3/4 of the population will say, that's not acceptable. I don't want any risk. And these people are fools. But unfortunately, these fools outnumber those of us who have studied these issues. So we have to deal with it. We have to understand it if nothing else.

OK, failures in all types of systems-- operations research, biology, medicine, optical systems, and control system. Name your own favorite thing. You can put it all in. Probability gets used everywhere.

OK, let's go to the axioms. Probability models have three components to them. There's a sample space.

Now, here we're in mathematics again. The sample space is just a set of things. You don't have to be at all specific about what those things are. I mean, at this point we're right in to set theory, which is the most basic part of mathematics again. And a set contains elements. And that's what we're talking about here. So there's a sample space. There are the elements in that sample space. There's also a collection of things called events.

Now, the events are subsets of the sample space. And if you're dealing with a finite set of things, there's no reason why the events should not be all subsets of that countable collection of things.

If you have a deck of cards, there are 52 factorial ways of arranging the cards in that deck of cards. Very large number. But when you talk about subsets of that, you might as well talk about all combinations of those configurations of the deck.

You can talk about, what's the probability that the first five cards in that deck happen to contain 4 aces? That's an easy thing to compute. I'm sure you've all computed it at some point or other. Those who like to play poker, of course do this. It's fun. But it's a straightforward problem.

When you have these countable sets of things, there's no reason at all for not having the set of events consist of all possible subsets.

Well, people believed that for a long time. One of the things that forced Kolmogorov to start dealing with these axioms was the realization that when you had much more complicated sets, where in fact you had the set of real numbers as possible outcomes, or sequences of things which go from 0 to infinity, and all of these sets, which are uncountable, you really can't make sense out of probability models where all subsets of sample points are called events. So in terms of measure theory, you're forced to restrict the set of things you call events.

Now, we're not going to deal with measure theory in this subject. But every once in a while, we will have to mention it because the reason why a lot of things are the way they are is because of measure theory. So you'll have to be at least conscious of it.

If you really want to be serious, as far as your study of mathematical probability theory, you really have to take a course in measure theory at some point. But you don't have to do it now. In fact, I would almost urge most of you not to do it now. Because once you get all the way into measure theory, you're so far into measure theory that you can't come back and think about real problems anymore. You're suddenly stuck in the world of mathematics, which happens to lots of people.

So anyway, some of you should learn about all this mathematics. Some of you shouldn't. Some of you should learn about it later. So you can do whatever you want.

OK, the axioms about events is that if you have a set of events. In other words, a set of subsets, and it's a countable set, then the union of all of those-- the union from n equals 1 to infinity of a sub n is also an event.

I've gone for 50 minutes and nobody has asked a question yet. Who has a question? Who thinks that all of this is nonsense? How many of you? I do. OK, I'll come back in another 10 minutes. And if nobody has a question by then, I'm just going to stop and wait. OK, so anyway.

If you look at a union of events. Now, remember, that an event is a subset of points. We're just talking about set theory now. So the union of this union here-- excuse me. This union here is A1, all the points in A1, and all the points in A2, and all the points in A3, all the way up to infinity. That's what we're talking about here.

And one of the axioms of probability theory is that if each of these things are events, then this union is also an event. That's just an axiom. You can't define events if that's not true. And if you try to define events where this isn't true, you eventually come into the most god awful problems you might imagine. And suddenly, nothing make sense anymore.

Most of the time when we define a set of events in a probability model, each singleton event-- namely, each single point has a set, which contains only that element, is taken as an event. There's no real reason to not do that.

If you don't do that, you might as well just put those points together and not regard them as separate points. We will see an example in a little bit where, in fact, you might want to do that. But let's hold that off for a little bit.

OK, not all subsets need to be events. Usually, each sample point is taken to be a singleton event. And then non-events are truly weird things. I mean, as soon as you take all sample points to be events, all countable unions of sample points are events. And then intersections of events are events, and so forth, and so forth, and so forth. So most things are events. And just because of measure theory, you can't make all things events. And I'm not going to give you any example of that because examples are horrendous.

OK, the empty set has to be an event. Why does the empty set have to be an event. If we're going to believe these axioms-- I'm in a real bind here because every one of you people has seen these axioms before. And you've all gone on and said, I can get an A in any probability class in the world without having any idea of what these axioms are all about. And therefore, it's unimportant. So you see something that says, the empty set is an event. And you say, well, of course that has nothing to do with anything. Why should I worry about whether the empty set is an event or not? The empty set can't happen, so how can it be an event?

Well, because of these axioms, it has to be an event. The axioms say that if A is an event, and that's the whole sample space, then the complement has to be an event also. So that says that the empty set has to be an event. And that just follows from the axioms.

If all sample points are singleton events, then all finite and countable sets are events.

And finally, deMorgan's law. Is there anyone who isn't familiar with deMorgan's law? Anyone who hasn't seen even that small amount of set theory? If not, look it up on-- what's the name of the computer-- Wikipedia.

Most of you will think that things on Wikipedia are not reliable. Strangely enough, in terms of probability theory and a lot of mathematics, Wikipedia does things a whole lot better than most textbooks do. So any time you're unfamiliar with what a word means or something, you can look it up in your old probability textbook. If you've used [INAUDIBLE] and [INAUDIBLE], you will probably find the right answer there. Other textbooks, maybe the right answer. Wikipedia's more reliable than most of them. And it's also clearer than most of them. So I highly recommend using Wikipedia whenever you get totally confused by something.

OK, so probability measure and events satisfies the following axioms. We've said what things are events. The only things that have probabilities are events. So the entire set has a probability. When you do the experiment, something has to happen. So one of the sample points occurs. That's the whole idea of probability. And therefore, omega has probability 1. Capital Omega.

If A is an event, then the probability of A has to be greater than or equal to 0. You can probably see without too much trouble why it has to be less than or equal to 1 also. But that's not one of the axioms. You see, when you state a set of axioms for something, you'd like to use the minimum set of axioms you can, so that you don't have to verify too many things before you say, yes, this satisfies all the axioms. So the second one is the probability of A has to be greater than or equal to 0.

The third one says that if you have a sequence of disjoint events, incidentally when I say a sequence, I will almost always mean a countably infinite sequence-- A1, A2, A3, all the way up. If I'm talking about what most of you would call a finite sequence-- and I like the word "finite sequence," but I like to be able to talk about sequences. I'm talking about a finite sequence I will usually call it an n-tuple of random variables or an n-tuple of things. So sequence really means you go the whole way out.

OK, if A1, A2, all the way up are disjoint events-- disjoint. Disjoint means if omega is only in one, it can't be in any of the others. Then the probability of this countable union is going to be equal to the sum of the probabilities of the individual event.

Anyone who has ever done a probability problem knows all of these things. The only thing you don't know and you probably haven't thought about is why everything else follows from this. But this is the whole mathematical theory. Why should we study it anymore? We're done. We have the axioms. Everything else follows, it's just a matter of computation. Just sit down and do it. Not quite that simple.

Anyway, a few consequences of the probability of the empty set is 0, which says when you do an experiment something's going to happen. And therefore, the probability that nothing happens is 0 because that's what the model says. The probability of the complement of an event is 1 minus the probability of that event. Which, in fact, is what's says that all events have to have probabilities less than or equal to 1. And if the event A is contained in the event B-- remember when we talk about events, we're talking about two different things, both simultaneously.

One of them is this beautiful idea with measure theory worked into it and everything else. And the other is just a simple set theoretic idea. And all we need to be familiar with is a set theoretical idea. Within that set theoretical idea, A contained in B means that every sample point that's in A is also in B. It means that when you do an experiment, and the event A occurs, the event B has to occur because one of the things that compose A has to occur. And that thing has to be in B because A is contained in B. So the probability of A has to be less than or equal to the probability of B. That has to be less than or equal to 1. These are things you all know.

Another consequence is the union bound. Many of you have probably seen the union bound. We will use it probably almost every day in this course. So it's good to have that as one of the things you remember at the highest level.

If you have a set of events-- A1, A2, and so forth-- the probability of that union--

namely, the event that consists of all of them-- is less than or equal to the sum of the individual event probabilities.

I give a little proof here for just two events, A1 and A2. So you see why this is true. I hope you can extend this to 3 and 4. I can't draw a picture of it very easily for 3 and 4 and so forth. But here's the event A1. Here's the event A2. Visualize this as a set of sample points, which are just in the two-dimensional space here. So all these points here are in A1. All these points are in A2. This set of points here are the points that are in A1 and A2.

I will use just writing things next to each other to mean intersection. And sometimes I'll use a big cap to main intersection. So all of these things are both in A1 and A2. This is A2, but not A1. So the probability of this whole event, A1 union A2, is the probability of this thing and this thing together. So it's the probability of this plus the probability of this. The probability of this is less than the probability of A2 because this is contained in that whole rectangle. And therefore, the probability of the union of A1 and A2 is less than or equal to the probability of A1 plus probability of A2.

Now, the classy way to extend this to a countably infinite set is to use induction. And I leave that as something that you can all play with some time when it's not between 9:30 and 11:00 in the morning and you're struggling to stay awake. And if you don't want to do that on your own, you can look at it in the notes.

OK, these axioms look ho-hum to you. And you've always ignored them before, and you think you're going to be able to ignore them now. Partly you can, but partly you can't because every once in a while we'll start doing things where you really need to understand what the axioms say.

OK, one other thing which you might not have noticed. When you studied elementary probability, wherever you studied it, what do you spend most of your time doing? You spent most of your time talking about random variables and talking about expectations. The axioms don't have random variables in them. They don't have expectations in them. All they have in them is events and probabilities of events. So these axioms say that the really important things in probability are the events and the probabilities of events. And the random variables and the expectations are derived quantities, which we'll now start to talk about.

OK, so we're now down to independent events and experiments. Two events, A1 and A2, are independent if the probability of the two of them is equal to the product of their probabilities. You've all seen this. I'm sure you've all seen it. If you haven't at least seen it, you probably shouldn't be in this class because even though the text does everything in detail that has to be done, you need to have a little bit of insight from having dealt with these subjects before. If you don't have that, you're just going to get lost very, very quickly.

So the probability is the intersection of the event A1 and A2 is the product of the two. Now, in other words, you have a red die and a white die. You flip the dice, what's the probability that you get a 1 for the red die and a 1 for the white die? Well, the probability you get a 1 for the red die is 1/6. Just by symmetry, there are only 6 possible things that can happen. Probability of white die comes up as 1. Probability is 1/6 for that. And the probability of the two things, they're independent. There's a sense of real-world independence and probability theory independence. Real-world independence says the two things are isolated, they don't interfere with each other.

Probability theory says just by definition, well, the real-world idea of them not interfering with each other should say-- and I'm waving my hands here because this is so elementary, you all know it. And I would bore you if I talked about it more. But I probably should talk about it, but I'm not going to.

Anyway, this is the definition of independence. If you don't have any idea of how this corresponds to being unconnected in the real-world, then go to Wikipedia. Read the notes. Well, you should read the notes anyway. I hope you will read the notes because I'm not going to say everything in class that needs to be said. And you will get a better feeling for it.

Now, here's something important. Given two probability models, a combined model can be defined in which, first, the sample space, omega, is the Cartesian product of omega 1 and omega 2. Namely, it's the Cartesian product of the two sample spaces.

Think of rolling the red die and the white die. Rolling a red die is an experiment. There are 6 possible outcomes, a 1 to 6. Rolled a white die, there are 6 possible outcomes, a 1 to a 6. You roll the two dice together, and you really need to have some way of putting these two experiments together. How do you put them together? You talk about the outcome for the two dice, number for one and number for the other. The Cartesian product simply means you have the set made up of 1 to 6 Cartesian product with 1 to 6. So you have 36 possibilities.

It's an interesting thing, which comes from Kolmogorov's axioms. That, in fact, you can take in any two probability models for two different experiments. You can take this Cartesian product of sample points. You can assume that what happens here is independent of what happens here. And when you do this, you will, in fact, get something for the two experiments put together which satisfies Kolmogorov's axioms. That is neither trivial nor very hard to prove. I mean, for the case of two dice, you can see it almost immediately. I mean, you see what the sample space is. It's this Cartesian product. And you see what the probabilities have to be because the probability of, say, 1 and 2 for the red die and 1 and 2 for the white is 2, 6 times 2, 6. So with probability 1/9, you're going to get a 1 and a 2 combined with a 1 and a 2.

I'm going to talk a little bit later about something that you all know. What happens if you roll two white dice? This is something you all ought to think about a little bit because it really isn't as simple as it sounds. If you roll two dice, what's the probability that you'll get a 1 and a 2? And how can you justify that?

First, what's a sample space when you roll two white dice?

Well, if you look at the possible things that might happen, you can get 1, 1; 2, 2; 3, 3; 4, 4; 5, 5; 6, 6. You can also get 1, 2 or 2, 1. But you can't tell them apart, so there's one sample point, you might say, which is 1, 2, and 2, 1. Another sample point which is 2, 3; 3, 2, and so forth. If you count them up, there are 21 separate sample points that you might have. And when you look at what the probabilities ought to be, the probabilities of the pairs are 136 each. And the probabilities of the I J, where I is unequal to J is 1/18 each. That's awful. So what do you do?

When you're rolling two dice, you do the same thing that everybody else does. You say, well, even though it's two white dice, I'm going to think of it as if it's a white die and a red die. I'm going to think of it as if the two are indistinguishable. My sample space is going to be these 36 different things. I will never be able to distinguish a 1, 2 from a 2, 1. But I don't care because I now know the probability of each of them.

What I'm trying to say by this, is this a very, very trivial example of where you really have to think through the question of what kind of mathematical model do you want of the most simple situation you can think of almost. When you combine two different experiments together and you lose distinguishability, then what do you do?

Well, the sensible thing to do is assume that the distinguishability is still there, but it's not observable. But that makes it hard to make a correspondence between the real world and the probability world. So we'll come back to that later. But for the most part, you don't have to worry about it because this is something you've dealt with all of your lives.

I mean, you've done probability problems with dice. You've done probability problems with all sorts of other things where things are indistinguishable from each other. And after doing a few of these problems, you are used to being schizophrenic about it. And on one hand, thinking that these things are distinguishable to figure out what all the probabilities are. And then you go back to saying, well, they aren't really distinguishable, and you find the right answer. So you don't have to worry about it. All I'm trying to say here is that you should understand it. Because when you get the complicated situations, this is one of the main things which will cause confusion. It's one of the main things where people write papers and other people say that paper is wrong because they're both thinking of different models for it.

Important thing is if you satisfy Kolmogorov's axioms in each of a set of models, and most important thing is where each of these models are exactly the same. And then you make them each independent of each other, Kolmogorov's axioms are going to be satisfied for the combination, as well as the individual model.

Why do I care about that? Because we're studying stochastic processes. We're studying an infinite sequence of random variables. And I don't want to generate a complete probability model for an infinite set of random variables every time I talk about an infinite set of random variables.

If I'm talking about an infinite sequence of flipping a coin, I want to do what you do, which is say, for each coin, the coin is equiprobably a head or a tail. And the coin tosses are independent of each other. And I want to know that I can go from that to thinking about this sequence. Strange things will happen in these sequences when we go to the limit. But still, we don't want to have to worry about a model for the whole infinite sequence. So that's one of the things we should deal with.

Finally, random variables. Definition. Three years ago, I taught this course and I asked people to write down definition of what a random variable was. And almost no one really had any idea of what it was. They said it was something that had a probability density, or something that had a probability mass function, or something that had a distribution function, or something like that.

What it is, if you want to get a definition which fits in with the axioms, the only thing we know from the axioms is there's a sample space. There are events and there are probabilities. So a random variable, what it really is, is it's a function from the set of sample points to the set of real values. And as you get into this, you will realize that the set of real values does not include minus infinity or plus infinity. It says that every sample point gets mapped into a finite value. This happens, of course, when you flip a coin.

Well, flipping a coin, the outcome is not a random variable. But you'd like to make it a random variable, so you say, OK, I'm going to model tails as 0 and heads as 1, or vice versa. And then what happens?

Your model for coin tossing, a sequence of coin tosses becomes the same as your model for data. So that what you know about coin tossing, you can apply to data compression. You see, when you think about these things mathematically, then you can make all sorts of connections you couldn't make otherwise.

So random variables have to satisfy the constraint that-- they have to satisfy the constraint that the set of sample points, such that x, x of omega, which is a real number, is less than or equal to some given real number. That this set has to be an event. Because those are the only things that have probabilities. So if we want to be able to talk about the probabilities of these random variables lying in certain ranges, or things like this, or having PMFs, or anything that you like to do, you need this constraint on it. It's an event for all A in the set of real numbers.

Also, if this set of things here are each random variables. In other words, if each of them are functions from the sample space to the real line, then the set of omega such that x1 of omega is less than or equal to A1, up to An of omega is less than or equal to A n is an event also. You might recognize this as the distribution function, the joint distribution function for n random variables. You might recognize this as the distribution function evaluated at A for a single random variable. So you define a random variable.

And what we're doing here is-- it's kind of funny because we already have these axioms. But now when we want to define things in the context of these axioms, we need extra things in the definitions.

This is a distribution function, a distribution function of the random variable x is the probability that the random variable x is less than or equal to x, which means that x is a mapping from omega into real numbers. It says that with this mapping here, you're mapping this whole sample space into the real line. Some omegas get mapped into things less than or equal to a real number x. Some of them get mapped into things greater than the real number x. And the set that gets mapped into something less than or equal to x, according to the definition of a random variable, has to be an event. Therefore, it has to have a probability. And these probabilities increase as we go.

It is totally immaterial for all purposes whether we have a less than or equal to here or a less than here. And everyone follows the convention of using a less than or equal to here rather than a less than here. The importance of that is that when you look at a distribution function, the distribution function often has jumps. And the distribution will have a jump whenever there's a nonzero probability that the random variable takes on a particular value x here. It takes on this particular value with something more than probabilities here.

If you have a probability density for a random variable, this curve just moves up continuously. And the derivative of this curve is the probability density.

If you have a probability mass function, this is a staircase type of function. Because of the fact that we define the distribution function with a less than or equal to rather than a less than means that in every one of these jumps, the value here is the upper value of the jump. Value here is the upper value of the jump, and so forth.

Now, I'm going to-- I've already said half of this.

Affects maps only until finite or countable set of values. It's discrete. And it has a probability mass function-- this notation.

If the derivative exists, then you say that the random variable is continuous and it has a density. And most problems that you do in probability theory, you're dealing with random variables. And they either have a probability mass function if they're discrete or they have a density if they're continuous. And this is just saying some things are one way, some things are the other way. And some things are neither. And we'll see lots of things that are neither. And you need the distribution function to talk about things that are neither. We will find that the distribution function, which you've hardly ever used in the past, is extraordinarily important, both for theoretical purposes and for other purposes. You really need that as a way of solving problems, as well as keeping yourself out of trouble.

For every random variable, the distribution function exists. Why? Anybody know why this has to exist for every random variable? Yeah.

AUDIENCE: Because the [INAUDIBLE].

PROFESSOR: Yes. Because we insisted that it did. Namely, we insisted that this event actually was an event for all little x. That's part of the definition. So, in fact, when you do these things a little more carefully than you might be used to, the definition implies that the distribution function always exists. As a more real-world kind of argument, we now have a way of dealing with things that are discrete, and continuous, and mixed continuous and discrete, and anything else that you might think of because the definition restricts it.

Now, one other thing. How do I know that this starts at 0? That's a more complicated thing. And I'm not even going to do it in detail. But since every omega maps into a finite number, you can't have a jump down here at minus infinity. And you can't have a jump here at plus infinity. Because omegas don't map into plus infinity or minus infinity. So you have to start down here at 0. You have to climb up here to 1. You might never reach 1. You might reach it only as a limit, but you have to reach it as a limit. Yes?

AUDIENCE: In the first paragraph, [INAUDIBLE].

PROFESSOR: If we have a sequence of [INAUDIBLE], yeah.

AUDIENCE: It's [INAUDIBLE].

PROFESSOR: You are probably right. Yes. Well, I don't know I don't think about that one. I don't think you're right, but we can argue about it. But anyway, this has to start at 0. It has to go up to 1.

OK, we did this. Now, I'm going to go through a theoretical nitpick for the last five minutes of the class. Anyone who doesn't like theoretical nitpicks, you're welcome to either go to sleep for five minutes, or you're welcome to go out and get a cup of coffee, or whatever you want to do. I will do this to you occasionally. And I realize it's almost torture for some of you, because I want to get you used to thinking about how relatively obvious things actually get proven. I want to increase your ability to prove things.

The general statement about proving things, or at least the way I prove things, is not the way most mathematicians prove things. Most mathematicians prove things by starting out with axioms and going step by step until they get to what they're trying to prove.

Now, every time I talk to a good mathematician, I find out that's what they write down when they prove something, but that's not the way they think about it at all. All of us-- engineers, businesspeople, everyone-- thinks about problems in a different way. If we're trying to prove something, we first give a really half-assed proof of it. And after we do that, we look at it and we say, well, I don't see why this is true and I don't see why that's true. And then you go back and you patch these things up. And then after you patch things up, it starts to look ugly. So you go back and do it a nicer way. And you go back and forth and back and forth and back and forth, using both contradiction and implication. You use both of them.

Now, when you're proving things in this class, I don't care whether you make it look like you're a formal mathematician or not. I would just assume you didn't pretend you were a formal mathematician. I would like to see you prove things in such a way that it is at least difficult to poke a hole in your argument. In other words, I would like you to give an argument which you've thought about enough that there aren't obvious counter examples to it. And if you learn to do that, you're well on your way to learning to use this theory in a way where you can actually come up with correct answer.

And in fact, I'm not going to go through this proof at all. And I don't think I really wanted to. I just did it because-- well, I think it's something you ought to read. It is important to learn to prove things because when you get to complicated systems, you cannot see your way through them intuitively. And if you can't see your way through it intuitively, you need to understand something about how to prove things, and you need to put all the techniques of proving things that you learned together with all the techniques that you've learned for doing things intuitively. And you need to know how to put them together.

If you're stuck dealing only with things that are intuitive, or things that you learned in high school like calculus, then you really can't deal with complicated systems very well.

OK, I'm going to end at that point. You can read this theoretical nitpick if you want, and play with it. And we'll go on next time.