The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

PROFESSOR: OK, so let's get started. We're going to essentially finish up on Poisson processes today. And today, we have the part of it that's really the fun part. What we had until now was the dry stuff, defining everything and so forth.

As I said before, Poisson processes are these perfect processes where everything that could be true is true. And you have so many different ways of looking at problems that you can solve problems in a wide variety of ways.

What you want try to do in the problem set this week is to get out of the mode of starting to write equations before you think about it. I mean, write equations, yes, but while you're doing it, think about what the right way of approaching the problem is. As far as I know, every part of every problem there can be solved in one or two lines. There's nothing long. There's nothing complicated. But you have to find exactly the right way of doing it.

And that's what you're supposed to learn because you find all sorts of processes in the world, which are poorly modelled as Poisson processes. You get a lot of insight about the real process by looking at Poisson processes, but you don't get the whole story. And if you don't understand how these relationships are connected to each other then you have no hope of getting some sense of when Poisson process theory is telling you something and when it isn't.

That's why engineers are different than mathematicians. Mathematicians live in this beautiful world. And I love it. I love to live there. Love to go there for vacations and so on because everything is perfectly clean. Everything has a right solution. And if it's not a right solution, it's a wrong solution.

In engineering, everything is kind of hazy. And you get insights about things when you put a lot of insights together. You finally make judgments about things. You use all sorts of models to do this. And what you use a course like this for is to understand what all these models are saying. And then be able to use them.

So, the stuff that we'll be talking about today is stuff you will use all the rest of the term. Because everything we do, surprisingly enough, from now on, has some relationships with Poisson processes. It doesn't sound like they do, but, in fact, they do. And this has a lot of the things that look like tricks, but which are more than tricks. They're really what comes from a basic understanding of Poisson processes.

So first, I'm going to review a little bit what we did last time. Poisson process is an arrival process. Remember what an arrival process is. It's just a bunch of arrival epochs, which have some statistics associated with them. And it has IID exponentially-distributed interarrival times. So the time between successive arrivals is independent from one arrival to the next. And it has this exponential distribution, which is what gives the Poisson process it's very special characteristic.

It can be represented by its arrival epochs. These are the things you see here, S1, S2, S3, and so forth, are the arrival epochs. If you can specify what the probability relationship is for this set of joint random variables then you know everything there is to know about a Poisson process. If you specify the joint distribution of these interarrival times, and that's trivial because they're IID exponentially-distributed random variables, then you know everything there is to know about the process. And if you can specify N of t, which is the number of arrivals up until time t for every t then that specifies the process completely also.

So we take the viewpoint here, I mean, usually we view a stochastic process as either a sequence of random variables or a continuum of random variables. Here, we're viewing as this three ways of looking at the same thing. So a Poisson process is then either the sequence of interarrival times, the sequence of arrival epochs, or, what we call the counting process, N of t at each t greater than zero. The arrival epochs in N of t are related either this way or this way. And we talked about that.

The interarrival times are memoryless. In other words, they satisfy this relationship here. The probability that an interarrival time Xi is greater than t plus x, for any t and any x, which are positive, given that it's greater than t. In other words, given that you've already wasted time t waiting, the probability that the time from here until the actual occurrence occurs is again exponential. This is stated in conditional form. We stated it before in just joint probability form.

OK, which says, that if you wait for a while and nothing has happened, you just keep on waiting. You're right where you started. You haven't lost anything. You haven't gained anything. And we said that other renewal processes, which are IID interarrival random variables, you can have these heavy tailed distributions where if nothing happens after while then you start to really feel badly because you know nothing's going to happen for an awful lot longer.

Heavy tailed distribution's best example is when you're trying to catch an airplane and they say, it's going to be 10 minutes late. That's the worst heavy tailed distribution there is. And it drives you crazy. Because I've never caught a plane that was supposed to be 10 minutes late that wasn't at least an hour late. And often, it got canceled, which makes it not a random variable at all.

One of the things we're interested in now, and we talked about it a lot last time, is you pick some arbitrary time t, that can be any time at all. And you ask, how long is it from time t-- t might be when you arrive to wait for a bus or something-- how long is it until the next bus comes? So Z is the random variable that goes. And you really should put some indices on this. But what it is is the random variable from t until this slow arrival here that's poking along finally comes in.

Now, what we found is that the interval Z, which is the time from this arrival back to t, is exponential. And the way we showed that is to say, let's condition Z on anything we want to condition it on. And the things that's important too condition it on is the value of N of t here, which is N. And once we condition it on the fact that N of t is n, we then condition on S sub n, which is the time that this last arrival came.

So if we condition Z on N of t at time t and the time that this last arrival came in,

which is S sub two, in this case, then Z turns out to be, again, just the time left to wait after we've already waited for this given amount of time. We then find out that Z, conditional on these two things that we don't understand at all, it's just exponential, no matter what they are. And since it's exponential no matter what they are, we don't go to the trouble of trying to figure out what they are. We just say, whatever distribution they have, Z itself, unconditionally, is just the same exponential random variable. So that was one of the main things we did last time.

Next thing we did was we started to look at any set of time, say t1, t2, up to t sub k, and then we looked at the increments. How many arrivals occurred between zero and t1? How many arrivals occurred between t1 and t2? How many between t2 and t3? So we looked at these Poisson process increments, a whole bunch of random variables, and we said, these are stationary and independent Poisson counting processes, over their given intervals.

In other words, if you stop the first process, you stop looking at this at time t1. Then you look at this from time t1 to t2. You look at this one from tk minus 1 to tk. And you look at the last one, the next last one, all the way up to tk. And there should be one where you look at it from tk on to t. No, you're only looking at k of them. So these are the things that you're looking at. The statement is that these are independent random variables.

The other statement is they're stationary, which means, if you look at the number of arrivals in this interval, here, it's a function of t2 minus t1. But it's not a function of t1 alone. It's only a function of the length of the interval. The number of arrivals in any interval of length tk minus tk minus 1 is a function only of the length of that interval and not of where the interval is. That's a reasonable way to look at stationarity, I think. How many arrivals come in in a given area is independent of where the area is. It depends only on how long the interval is.

OK, then we found that the probability mass function for N of t, and now we're just looking from zero to t because we know we get the same thing over any interval, this probability mass function is this nice function, which depends only and the

product lambda t. It never depends on lambda alone or t alone. Lambda t to n, e to the minus lambda t over n factorial.

What the n factorial is doing there. Well, it came out in the derivation. By the time we finish today you should have more ideas of where that factorial came from. And we'll try to understand that.

By the stationary and independent increment property, we know that these two things are independent, N of t1 and the number of arrivals in t1 to t. This is a Poisson random variable. This is a Poisson random variable. We know that the number of arrivals between zero and t is also a Poisson random variable. And what does that tell you?

It tells you, you don't have to go through all of this discrete convolution stuff. You probably should go through it once just for your own edification to see that this all works. But for a very lazy person like me, who likes using arguments like this, I say, well, these two things are independent. They are Poisson random variables. Their sum is Poisson. And therefore, whenever you have two independent Poisson random variables and you add them together, what you get is a Poisson random variable whose mean is the sum of the means of the two individual random variables. In general, sums of independent Poisson random variables are Poisson with the means adding.

Them we went through a couple of alternate definitions of a Poisson process. And at this point, just from what I've said so far, and from reading the notes and understanding what I've said so far, it ought to be almost clear that these alternate definitions, which we talked about last time have to be valid. If an arrival process has the stationary and independent increment properties and if N of t has the Poisson PMF for given lambda, and all t, then the process is Poisson.

Now what is that saying? I mean, we've said that if we can specify all of the random variables N of t for all t, then we've specified the process. What does it mean to specify a whole bunch of random variables? It is not sufficient to find the distribution function of all those random variables.

And one of the problems in the homeworks at this time is to explicitly show that for a simpler process, the Bernoulli process. And to actually construct an example of where N of t is specified everywhere. But you don't have the independence between different intervals, and therefore, you don't have a Bernoulli process. You just have this nice binomial formula everywhere. But it doesn't really tell you much.

OK, so, but here we're adding on the independent increment properties, which says over any set of intervals the joint distribution of how many arrivals there are here, how many here, how many here, how many here, those joint random variables are independent of each other, which is what the independent increment property says. So in fact, this tells you everything you want to know because you now know the relationship between each one of these intervals. So we see why the process is Poisson.

This one's a little trickier. If an arrival process has the stationary and independent increment properties and it satisfies this incremental condition, then the processes Poisson. And the incremental condition says that if you're looking at the number of arrivals in some interval of size delta, the probability that this is equal to N has the form 1 minus lambda delta plus o of delta, where n equals zero. Lambda delta plus o of delta, for n equals 1. o of delta for n greater than or equal to 2.

This intuitively is only supposed to talk about very, very small delta. So if you take the Poisson distribution lambda t to the N, e to the minus lambda t over n factorial, and you look at what happens when t is very small, this is what it turns into. When t is very small the probability that there are no arrivals in this interval of size delta is very large. It's 1 minus lambda delta plus this extra term that says-- first point of view whenever you see an o of delta is to say, oh, that's not important. And for N equals 1, there's going to be one arrival with some fudge factor, which is not important. And there's going to be two or more arrivals with some fudge factor, which is not important.

The next thing we talked about is that o of delta really is defined as any function which is a function of delta where the limit of o of delta divided by delta is equal to 0.

In other words, it's some function that goes to zero faster than delta does. So it's something which is insignificant with relationship to this as delta gets very small. Now, how do you use this kind of statement to make some statement about larger intervals?

Well, you're clearly stuck looking at differential equations. And the text does that. I refuse to talk about differential equations in lecture or anyplace else. When I retired I said, I will no longer talk about differential equations anymore. And you know, you don't need to because you can see what's happening here. And what you see is happening is, in fact, what's happening. OK, so, that's where we finished up last time.

And now, we come to the really fun stuff where we want to combine independent Poisson processes. And then, we want to split Poisson processes. And we want to play all sorts of games with multiple Poisson processes, which looks very hard, and because of this, it's very easy. Yes?

AUDIENCE: The previous definitions, they are if and only statements, right?

PROFESSOR: They are what?

AUDIENCE: If and only, all Poisson processes satisfy tho--

PROFESSOR: Yes. OK, in other words, what you're saying is, if you can satisfy those properties, this is a Poisson process. I mean, we've already shown that a Poisson process satisfies those properties. As a matter of fact, the way I put it in the notes is these are three alternate definitions where you could start out with any one of them and derive the whole thing. Many people like to start out with this incremental definition because it's very physical. But it makes all the mathematics much, much harder.

And so, it's just a question of what you prefer. I like to start out with something clean, then derive things, and then say, does it make any sense for the physical situation? And that's what we usually do. We don't usually start out with a physical situation and analyze the hell out of it and say, aha, this is a Poisson process because it satisfies all these properties. It never satisfies all those properties. I

mean, you say it's a Poisson process because a Poisson process is simple and you can get some insight from it, not because it really is a Poisson process.

Let's talk about taking two independent Poisson processes. Just to be a little more precise, two counting processes, N1 of t and N2 of t are independent if for all t1 up to t sub n the random variables N1 of t1 to N1 of tn are independent of N1 of t1 up to N2 of tn. Why don't I just say that for all t they're independent? Because I don't even know what that means. I mean, we've never defined independence for an infinite number of things. So all we can do is say, for all finite sets, we have this independence.

Now, give you a short pop quiz. Suppose that instead of doing it this way, I say, for all t1 to tn the random variables N1 of t1 to N1 of tn are independent of for all tao1, tao2, tao3, tao4, N2 of tao1 up to N2 of tao sub n? That sounds much more general, doesn't it? Because it's saying that I can count one process at one set of times and the other process at another set of times. Now, why isn't it any more general to do it that way?

Well, it's an unfair pop quiz because if you can answer that question in the short one sentence answer that you'd be willing to give in a class like this, I would just give you A plus and tell you go away. And you already know all the things you should know.

The argument is the following, if you order these different times, first tao1 of tao1 is less than t1 then t1, if that's the next one, and you order them all along. And then you apply this definition to that ordered set. t1, tao1, t2, t3, t4, tao2, t5, tao3, and so forth, and you apply this definition, and then you get the other definition. So one is not more general than the other.

The theorem then, is that if N1 of t and N2 of t are independent Poisson processes, one of them has a rate lambda 1. One of them has a rate lambda 2. And if N of t is equal to N1 of t plus N2 of t, that just means for every t this random variable N of t is the sum of the random variable N1 of t plus the random variable N2 of 2. This is true for all. This is, definition, for all t greater than 0, then the sum N of t is a Poisson process of rate lambda equals lambda 1 plus lambda 2.

Looks almost obvious, doesn't it? I said that today there was lots of fun stuff. There's also a little bit of ugly stuff and this is one of those half obvious things that's ugly. And I'm not going to waste a lot of time on it. You can read all about the details in the notes. But I will spend a little bit of time on it because it's important.

The idea is that if you look at any small increment, t to t plus delta, the number of arrivals in the interval t to t plus delta is equal to the number of arrivals in the interval t to t plus delta from the first process plus that in the second process. So the probability that there's one arrival in this combined process is the probability that there's one arrival in the first process and no arrivals in the second process, or that there's no arrivals in the first process and one arrival in the second process. That's just a very simple case of convolution. Those are the only ways you can get one in the combined process.

This term here is delta times lambda 1-- too early in the morning. I'm confusing my deltas and lambdas. There are too many of each of them. --plus o of delta. And this term here, probability that there's zero for the second process is 1 minus delta lambda 2 plus o of delta. And then this term is just the opposite term corresponding to this.

Now I multiply these terms out. And what do I get? Well, this 1 here combines with a delta lambda 1. Then there's a delta lambda 2 times delta lambda 1, which is delta squared. It's a delta squared term, so that's really an o of delta term. It's negligible as delta goes to zero. So we forget about that.

There's an o of delta times 1, that's still an o of delta. There's an o of delta times a delta lambda. And that's sort of an o of delta squared if you wish. But it's still an o of delta. It goes to zero as delta gets large. And goes to zero faster than delta.

What we're trying to do is to find the terms that are significant in terms of delta. Namely, when delta gets very small, I want to find things that are at least proportional to delta and not of lower order than delta. So when I get done with all of that, this is delta times lambda 1 plus lambda 2 plus o of delta. That's the

incremental property that we want a Poisson process to have. So it has that incremental property.

And those are the same sort of argument if you want to for the number of arrivals in t, that t plus delta. Maybe a picture of this would help. Pictures always help in the morning.

Here we have two processes. We're looking at some interval of time, t to t plus delta. t to plus delta. And we might have an arrival here. We might have an arrival here, an arrival here, an arrival here. Well, the probability of an arrival here and an arrival here is something of order delta squared. So that's something we ignore.

So it says, we might have an arrival here and no arrival here. This is a lambda 1. This is lambda 2 here. We might have an arrival here and none here, or we might have an arrival here and none there. Two arrivals is just too unlikely to worry about, so we forget about it at least for the time being.

Now, after going through this incremental argument, if you go back and say, let's forget about all these o of deltas because they're very confusing, let's just do the convolution knowing what N of t is in both of these intervals. If you do that, it's much easier to find out that the number of arrivals in this sum of intervals. Number arrivals here, plus the number of arrivals here, this is Poisson. This is Poisson. You add two Poisson, you get Poisson. The rate of the sum of the two Poisson is Poisson. So it's Poisson with lambda 1 plus lambda 2.

Now how many of you saw that and says, why is this idiot going through this incremental argument? Anyone? I won't be embarrassed. I knew it anyway. And I did it for a reason. But I must confess, when I wrote the first edition of this book I didn't recognize that. So I went through this terribly tedious argument.

Anyway, the more important issue is if you have the sum of many, many small independent arrival processes-- OK, in other words, you have the internet. And a node in the internet is getting jobs, or messages, from all sorts of people, and all sorts of processes, and all sorts of nonsense going to all sorts of people, all sorts of

processes, and all sorts of nonsense and those are independent of each other, not really independent of each other. But relative to the data rate that's travelling over this internet, each of those processes are very, very small.

And what happens is the sum of many, many small independent arrival processes tend to be Poisson even if the small processes are not. In a sense, the independence between the processes overcomes the dependence between successive arrivals in each process. Now, I look at that and I say, well, it's sort of a plausibility argument. You look at the argument in the text, and you say, ah, it's sort of a plausibility argument.

I mean, proving this statement, you need to put a lot of conditions on it. And you need to really go through an awful lot of work. It's like proving the central limit theorem, but it's probably harder than that. So, if you read the text, and say you don't really understand the argument there, I don't understand it either because I don't think it's exactly right. And I was just trying to say something to give some idea of why this is plausible. It should probably be changed.

Next we want to talk about splitting a Poisson process. So we start out with a Poisson process here and a t of rate lambda. And what happens? These arrivals come to some point. And some character is standing there.

It's like when you're having your passport checked when you come back to Boston after being away. In some places you press a button, and if the button come up one way, you're sent off to one line to be interrogated and all sorts of junk. And it's supposed to be random. I have no idea whether it's random or not, but it's supposed to be random. And otherwise, you go through and you get through very quickly.

So it's the same sort of thing here. You have a bunch of arrivals. And each arrival is effectively randomly shoveled this way or shoveled this way. With probability p, it's shoveled this way. With probability 1 minus p, it's shoveled this way. So you can characterize this switch as a Bernoulli process. It's a Bernoulli process because it's random and independent from, not time to time now, but arrival to arrival.

When we first looked at a Poisson process, we said it's a sequence of random variables. Sometimes we look at in terms of time. Time doesn't really make any difference there. It's just a sequence of IID random variables.

So you have a sequence of IID random variables doing this switching here. You have a Poisson process coming in. And when you look at the combination of the Poisson process and the Bernoulli process, you get some kind of process of things coming out here. And another kind of process of things coming out here. And the theorem says, that when you combine this Poisson process with this independent Bernoulli process, what you get is a Poisson process here and an independent Poisson process here. And, of course, you need to know what the probability is of being switched this way and being switched that way.

Each new process clearly has a stationary and independent increment property. Why is that? Well, you look at some increment of time and this process here is independent from one period of time to another period of time. The Bernoulli process is just switching the terms within that interval of time, which is independent of all other intervals of time. So that when you look at the combination of the Bernoulli process and the Poisson process, you have the stationary and independent increment property. And each satisfies the small increment property. If you look at very small delta here, unless each is Poisson.

There's a more careful argument in the notes. What I'm trying to do in the lecture is not to give you careful proofs of things, but to give you some insight into why they're true. So that you can read the proof. And instead of going through each line and saying, yes, I agree with that, yes, I agree with that and you finally come to the end of the proof and you think what? Which I've done all the time because I don't really know what's going on. And you don't learn anything from that kind of proof. If you're surprised when it's done it means you ought to go back and look at it more carefully because there's something there that you didn't understand while it was wandering past you.

The small increment property really doesn't make it clear that the split processes

are independent. And for independence both processes must sometimes have arrivals in the same small increment. And this independence is hidden in those o of delta terms. And if you want to resolve that for yourself, you really have to look at the text. And you have to do some work too.

The nice thing about combining and splitting is that you typically do them together. Most of the places where you use combining and splitting you use both of them repeatedly. The typical thing you do is first, you look at separate independent Poisson processes. And you take those separate independent Poisson processes, and you say, I want to look at those as a combined process. And after you look at them as a combined process, you then split them again.

And what you're doing, when you then split them again, is you're saying these two independent Poisson processes that I started with, I can view them as one Poisson process plus a Bernoulli process. And you'd be amazed at how many problems you can solve by doing that. You will be amazed when you do the homework this time. If you don't use that property at least 10 times, you haven't really understood what's being asked for in those problems.

Let's look at a simple example. Let's look at a last come, first served queue. Last come, first served queues are rather peculiar. They are queues where arrivals come in, and for some reason or other, when a new arrival comes in that arrival goes into the server immediately. And the queue gets backed up. Whatever was being served gets backed up. And the new arrival gets served, which is fine for the new arrival, unless another arrival comes in before it finishes. And then it gets backed up too. So things get backed up in the queue. And those that are lucky enough to get through, get through. And those that aren't, don't.

Anybody have any idea why that-- I mean, it sounds like a very unfair thing to do. But, in fact, it's not unfair because you're not singling out any particular group or anything. You're just that's the rule you use. And it applies to everyone equally, so it's not unfair. But why might it make sense sometimes? Yeah?

AUDIENCE: If you're doing a first come, first served you're going to have a certain distribution for

your serve time. Whereas, if you do last come, first served-- I don't know. I'm just trying to think that there might some situations where the distribution of service times in that is favorable overall, even though some people, every once in a while, are feeling screwed.

PROFESSOR: It's a good try. But it's not quite right. And, in fact, it's not right for a Poisson process because for a Poisson process the server is just chunking things out at rate mu, no matter what. So it doesn't really help anyone.

If you have a heavy tail distribution, if somebody who comes in and requires an enormous amount of service, then you get everybody else done first because that customer with a huge service requirement keeps getting pushed back every time the queue is empty he gets some service again. And then other people come in. And people with small service requirements get served quickly.

And what that means is it's not quite as crazy as it sounds. But the reason we want to look at it here is because it's a nice example of this combining and splitting of Poisson processes. So how does that happen?

Well, you view services as a Poisson process. Namely, we have an exponential server here, where the time for each service is exponentially distributed. Now, if you're awake, you point out, well, what happens when the server has nothing to do? Well, just suppose that there's some very low priority set of jobs that the server is doing, which also are exponentially distributed that the server goes to when it has nothing to do. So the server is still doing things exponentially at rate lambda, but if there's nothing real to do then the output is wasted.

So that's a Poisson process. The income is Poisson also. So the arrival process, plus the service process, they're independent. So they're Poisson. And the rate is lambda plus mu.

Now, interesting question, what's the probability that an arrival completes service before being interrupted? I'm lucky, I get into the server. I start getting served. What's the probability that I finish before I get interrupted?

Well, I get into service at a particular time t. I look at this combined Poisson process of arrivals and services and if the first arrival in this combined process gets switched to service, I'm done. If gets switched to arrival, I'm interrupted. So the question is, what's the probability that I get switched to-- I want to find the probability that I'm interrupted, so what's the probability that a new arrival in the combined process gets switched to arrivals because that's the case where I get interrupted.

And it's just lambda. The probability that I get interrupted is lambda divided by lambda plus mu. And the probability that I complete my service is mu over lambda plus mu. When mu is very large, when the server is going very, very fast, I have a good chance of finishing before being interrupted. When it's the other way, I have a much smaller chance of finishing.

More interesting, and more difficult case, given that you're interrupted, what is the probability that you have no further interruptions? In other words, I'm being served, I get interrupted, so now I'm sitting at the front of the queue. Everybody else that came in before me is in back of me. I'm sitting there at the front of the queue. This interrupting customer is being served. What's the probability that I'm going to finish my service before any further interruption occurs? Have to be careful about how to state this.

The probability that there is no further interruption is that two services occur before the next arrival. And the probability of that is mu over lambda plus mu the quantity squared. Now, whether you agree with that number or not is immaterial. The thing I want you to understand is that this is a method which you can use in quite complicated circumstances. And it's something that applies in so many places that it's amazing.

OK, let's talk a little bit about non-homogeneous Poisson processes. Maybe the most important application of this is optical transmission. There's an optical stream of photons that's modulated by variable power. Photon stream is reasonably modeled as a Poisson process, not perfectly modeled that way. But again, what we're doing here is saying, let's look at this model. And then, see whether the

consequences of the model apply to the physical situation.

The modulation converts the steady photon rate into a variable rate. So the arrivals are being served, namely, they're being transmitted at some rate, lambda of t, where t is varying. I mean, the photons have nothing to do the information at all. They're just random photons. And the information is represented by lambda t. The question is can you actually send anything this way?

We model the number of photons in any interval t and t plus delta as a Poisson random variable whose rate parameter over t and t plus delta, or very small delta, is the average photon rate over t and t plus delta times delta. So we go through this small increment model again. But in the small increment model, we have a lambda t instead of a lambda and everything else is the same.

And when you carry this out, what you find-- and I'm not going to show this or anything. I'm just going to tell it to you because it's in the notes if you want to read more about it. The probability that you have a given number of arrivals in some given interval is a Poisson random variable again. Namely, whether the process is homogeneous or not homogeneous, if the original photons are Poisson then what you wind up with is this parameter of the Poisson process. m hat to the N e the minus m hat divided by n factorial. And this parameter here is what? It's the overall arrival rate in that interval. So it's exactly what you'd expect it to be.

So what this is saying is combining and splitting non-homogeneous processes still works like in the homogeneous case. The independent exponential in arrivals don't work. OK, let me say that another way.

When you're looking at non-homogeneous Poisson processes, looking at them in terms of a Poisson counting process is really a neat thing to do. Because all of the Poisson counting process formulas still work. And all you have to do when you want to look at the distribution of the number arrivals in an interval is look at the average arrival rate over that interval. And that tells you everything. When you look at the interarrival times and things like that they aren't exponential anymore. And none of that works. So you get half the pie, but not the whole thing.

Now, the other thing, which is really fun, and which will take us a bit of time, is to study, now it's not a small increment but a bigger increment, of a Poisson process when you condition on N of t. In other words, if somebody tells you how many arrivals have occurred by time t, how do you analyze where those are arrivals occur, where the arrival epochs are, and things like that?

Well, since we're conditioning on N of t, let's start out with N of t equals 23 because that's obviously the simplest number to start out with. Obviously not. We start out with N of t equals 1 because that clearly is a simpler thing to start with.

So here's the situation we have. We're assuming that there's a one arrival up until this time. That one arrival has to occur somewhere in this interval here. We don't know where it is. So we say, well, let's suppose that it's in the interval s1 to s1 plus delta and try to find the probability that it actually is in that interval.

So we start out with something that looks ugly. But we'll see it isn't as ugly as it looks like. We'll look at the conditional density of s1 given that N of t is equal to 1. That's what this is. This is a conditional density here. And we say, well, how do we find the conditional density? This a one dimensional density. So what we're going to do is we're going to look at the probability that there's one arrival in this tiny interval. And then we're going to divide by delta and that gives us the density, if everything is well defined.

So the density is going to be the limit as delta goes to zero, which is the way you always define densities, of the probability that there is no arrivals in here. That there's one arrival here. And that there's no arrivals here. So probability zero arrivals here. Probability of one arrival here. And probability of zero arrivals here. We want to find that probability. We want to divide for the conditioning by the probability that there was one arrival in the whole interval. This delta here is because we want to go to the limit and get a density.

OK, when we look at this and we calculate this, e to the minus lambda s1 is what this is. Lambda delta times e to the minus lambda delta is what this is. And e to the

minus lambda t minus s1 minus delta is what this is. Because this last interval is t minus s1 minus delta.

Now, the remarkable thing is when I add, I multiply e to the minus lambda s1 times e to the minus lambda delta times e to the minus lambda t minus s1 minus delta. There's an s1 here. An s1 one here. There's a delta here. And a delta here. This whole thing is e to the minus lambda delta. And it cancels out with this e to the minus lambda delta. That's going to happen no matter how many arrivals I put in here.

So that's an interesting thing to observe. So what I wind up with is lambda delta up here, lambda delta t here, and my probability density is 1 over t. And now we look at that and we say, my God, how strange. And then we look at it and we say, oh, of course, it's obvious. Yes?

AUDIENCE: Why do you use N s1 and then the N tilde for the other two in the first part in the limit?

PROFESSOR: Here?

AUDIENCE: You have p and s1 and it's e N tilde on the top. Why do you use the N and the N tilde?

PROFESSOR: N of t is not shown in the figure. N of t just says there is one arrival up until time t, which is what lets me draw the figure.

AUDIENCE: I meant in the limit as delta approaches zero, Pn, Professor. You used Pn and then Pn tilde for the other two. Does that signify anything?

PROFESSOR: OK, this term is the probability that we have no arrivals between zero and s1. This N tilde means number of arrivals in an interval, which starts at s1 and goes to s1 plus delta. So this is the probability we have one arrival somewhere in this interval. And this term here is the probability that we have no arrivals in s1 plus delta up until t.

And when we write all of those things out everything cancels out except the 1 over t. And then we look at that and we say, of course. If you look at the small increment

definition of a Poisson process, it says that arrivals are equally likely at any point along here. If I condition on the fact that there's been one arrival in this whole interval, and they're equally likely to occur anywhere along here, the only sensible thing that can happen is that the probability density for where that arrival happens is uniform over this whole interval.

Now, the important point is that this doesn't depend on s1. And it doesn't depend on lambda. That's a little surprising also. I mean, we have this Poisson process where arrivals are dependent on lambda, but you see, as soon as you say N of t equals 1, you sort of wash that out. Because you're already conditioning on the fact that there was only one arrival by this time. So you have this nice result here. But a result which looks a little suspicious.

Well, we were successful with that. Let's go on to N of t equals 2. We'll get the general picture with this. We want to look at the probability density of arrivals at s1. This s1 and an arrival here, given that there were only two arrivals overall.

And again, we take the limit as delta goes to zero. And we take a little delta interval here, a little delta interval here. So we're finding the probability that there are two arrivals, one between s1 and s1 plus delta and the other one between s2 and s2 plus delta. So that probability should be proportional to delta squared. And we're dividing by delta squared to find the joint probability density.

You don't believe that, go back and look at how joint probability densities are the defined. I mean, you have to define them somehow. So you have to look at the probability in a small area. And then shrink the area to zero. And the area of the area is delta squared. And you have to divide by delta squared as you go to the limit.

OK, so, we do the same thing that we did before. The probability of no arrivals in here is e to the minus lambda s1. In other words, we are taking the unconditional probability and then we're dividing by the conditioning probability.

So the probability that there's nothing arriving in here is that. The probability there's

one arrival in here is that. Probability that there's nothing in this interval is this. Probability there's one in this interval is lambda delta times e to the minus lambda delta. And finally, this last term here, we have to divide by delta squared to go to a density. We have to divide by the probability that there were actually just two arrivals here.

Well, again, the same thing happens. If we take all the exponentials, lambda s1, lambda delta, lambda 2 in here, this term cancels out this term. This term here cancels out this term. And all we're left with is e to the minus lambda t up here, e to the minus lambda t down here. And because the PMF function for N of t is e to the minus lambda t over n factorial, n factorial for n equals 2 is 2. We wind up with 2 over t squared.

Now that's a little bit surprising. I'm going to show you a picture in just a little bit, which makes it clear what that 2 is doing there. But let's not worry about it too much for the time being. The important thing here is, again, it's independent lambda. And it's a independent of s1 and s2. Namely, given that there are two arrivals in here doesn't make any difference where they are. This is, in some sense, uniform over s1 and s2. And I have to be more careful about what uniform means here. But we'll do that in just a little bit.

Now we can do the same thing for N of t equals N, for arbitrary N. And when I put in an arbitrary number of arrivals here I still get this property that when I take e to the minus lambda of this region, e to the minus lambda times this region, this, this, this, this, this, this and so forth, when I get all done, it's e to the minus lambda t. And it cancels out with this term down here. We've done that down here.

And what you come up with is this joint density is equal to N factorial over t to the N. So again, it's the same story. It's uniform. It doesn't depend on where the s's are, except for the fact that s1 is less than s2 is less than s3. We have to sort that out. It has this peculiar N factorial here that doesn't depend on lambda, as I said. It does depend N in this way.

That's a uniform N dimensional probability density over the volume t to the N over N

factorial corresponding to the constraint region zero less than s1 less than blah, blah, blah, less than s sub N. Now that region is a little peculiar. This N factorial is a little peculiar. The t to the N is pretty obvious. Because when you have a joint density over t things, each bounded between zero and t, you expect to see a t to the N there.

So if you ask, what's going on? In a very simple minded way, the question is how did this derivation know that the volume of s1 to sn over zero less than s1, less than sn, less than s2, is N factorial over t to the N? I have a uniform probability density over some strange volume. And it's a strange volume which satisfies this constraint. How do I know that this is N factorial here? Very confusing at first.

We saw the same question when we derived the airline density, remember? There was an N factorial there. And that N factorial when we derived the Poisson distribution led to an N factorial there. And, in fact, the fact that we were using the Poisson distribution there is why that N factorial appears here. It was because we stuck that in when we were doing the derivation of the Poisson PMF. But it still seems a little mystifying. So I want to try to resolve that mystery.

And to resolve it let's look at a supposedly unrelated problem. And the unrelated problem is the following, suppose U1 up to U sub n are n IID random variables, independent and identically distributed random variables, each uniform over zero to t. Well, now, the probability density for those n random variables is very, very simple. Because each one of them is uniform over one to t. The joint density is 1 over t to the n.

OK, no problem there. They're each independent. So the probability densities multiply. Each one has a probability density 1 over t. No matter where it is because it's uniform. So you multiply them together. And no matter where you are in this n dimensional cube, the probability density is 1 over t to the n.

The next thing I want to do is define-- now these are supposedly not the same random variables as these arrival epochs we had before. These are just defined in terms of these uniform random variables. You're seeing a good example here of

why in probability theory we start out with axioms, we derive properties. and we don't lock ourselves in completely to some physical problem that we're trying to solve. Because if we had locked ourselves completely into a Poisson process, we wouldn't even be able to look at this. We'd say, ah, that's nothing to do with our problem. But it does. So wait.

We're going to define s1 as the minimum of U1 up to U sub n. In other words, these are the order statistics of U1 up to U sub n. I choose U1 to be anything between zero and t, U2 to be anything between zero and t. And then I choose s1 to be the smaller of those two and s2 to be the larger of those two.

This is zero to t. And here is say, U1. And here is U2. And then I transfer and call this s1 and this s2. And if I have a U3 out here then this becomes s2 and this becomes s3.

So the order statistics are just, you take these arrivals and you-- well, they don't have to be arrivals. They're whatever they are. They're just uniform random variables. And I just order them.

So now my question is, the region of volume t to the n, name I have a volume t to the n, for the possible values of U1 up to U sub n. That volume where the density of Un is non-zero, I can think of it as being partitioned into n factorial regions. First region is U1 less than U2, and so forth up to U sub n. The second one is U2 is less than U1, less than U3 and so forth up.

And for every permutation of U1 to U sub n, I get a different ordering here. I don't even care about the ordering anymore. All I care about is that the number of permutations of U1 to U sub n-- or the integers one to n, really is what I'm talking about. Number of permutations of 1 to n is n factorial. And for each one of those permutations, I can talk about the region in which the first permutation is less than the second permutation less than the third and so forth.

Those are disjoint-- they have to be the same volume by symmetry. Now how does that symmetry argument work there? I have n integers. There's no way to tell the

difference between them, except by giving them names. And each of them corresponds to a IID random variable. And now, what I'm saying is that the region which U1 is less than U2 has the same area as the region where U2 is less than U1.

And the argument is whatever you want to do with U1 and U2 to find the area of that, I can take your argument and every time you've written a U1, I'll substitute U2 for it. And every time you've written a U2 I'll substitute U1 for it. And it's the same argument. So that if you find the area using your permutation, I get the same area just using your argument again. And I can do this for as many terms as I want to.

OK, so this says that from symmetry each volume is the same and thus, each volume is t to the N divided by N factorial. Now the region where s sub N is non-zero, s sub N are these ordering statistics. These are the terms in which you wanted less than U2 less than U sub n after I reordered them. The volume of the region where S1 is less than S2 less than Sn less than or equal to t, the region of that volume is exactly t to the N divided by N factorial. Because it's one over N factorial of the entire region.

Now, you just have to think about that symmetry argument for a while. The other thing you can do is just integrate, which is what we did when we were deriving the airline distribution. I mean, you just integrate the thing out. And you find out that that N factorial magically appears there. I think you get a better idea of what's going on here.

Let me go on to this slide, which will explain a little bit what's going on here. Suppose I'm looking at S1 and S2. The area where S1 is less than S2 is this cross hatched area here. All the points in here are points where S2 is bigger than S1. As advertised, the region where S2 is bigger than S1 has the same area as the region where S2 is less than S1. And all I'm saying is this same property happens for N equals 3, N equals 4, and so forth.

And the argument there really is a symmetry argument. OK, I don't want to worry about what the x's are at this point. I'm just showing you that a little tiny area volume there has to be part of this triangle. And I get that factor of 2 because as S2 is

bigger than S1.

Let me give you an example of using the ordering statistics, which we just talked about. Namely, looking at N IID random variables, and then choosing S1 to be the smallest, S2 to be the next smallest and so forth. What we've seen is that this S1 to S sub N is identically distributed to the problem of the epochs of arrivals in a Poisson process conditional on N arrivals up until time t. So we can use either the uniform distribution and ordering or we can use Poisson results. Either one can give us results about either process.

Here what I'm going to do is use order statistics to find the distribution function of S1 conditional on N arrivals until time N. Conditional in the fact that I found 1,000 arrivals up until time 1,000. I want to find the distribution function of when the first arrival occurs.

Now, if I know that 1,000 arrivals have occurred by time 1,000, the first arrival is probably going to be pretty close to a Poisson random variable. But I'd like to see that.

OK, so the probability that the minimum of these U sub i's, these uniform random variables, is bigger than S1 is the product of the probabilities that each one is bigger than S1. The only way that the minimum can be bigger than S1 is that each of them are bigger than S1. So this is the product from I equals 1 to N of 1 minus S1 over t.

Then we take the product of things that are all the same. It's 1 minus S1 over t to the N. And then I go into the domain of these arrival epochs. And the probability that the first arrival epoch, the complimentary distribution function of that, is then 1 minus S1 over t to the N.

You can also find the expected value of S1 just by integrating this. Namely, integrate the complimentary distribution function to get the expected value. And what do you get? You get t over N plus 1. A little surprising, but not as surprising as you would think.

We're looking at an interval from zero to t. We have three arrivals there. And then

we're asking where does the first one occur? And the argument is this interval, this interval, this interval, and this interval ought to have the same distribution. I mean, we haven't talked about this interval yet, but it's there. I mean, the last of these arrivals is not a t, it's somewhere before t.

So I have n plus 1 intervals on the side of these n arrivals. And what this is saying is the nice symmetric result at the expected place where S1 winds up is at t over n plus one. These intervals are in some sense, at least as far as expectation is concerned, are uniform. And we wind up with t over N plus 1 as the expected value of where the first one is. That's nice and clean.

If you look at this problem and you think in terms of little tiny arrivals coming in any place, and, in fact, you look at this U1 up to U sub N, one arrival. I can think of it as one arrival from each of N processes. And these are all Poisson processes. They all add up to give a joint Poisson process, a combined Poisson process, with N arrivals. And then, I order the arrivals to correspond to the overall sum process. And this is the answer I get. So the uniform idea is related to adding up lots of little tiny processes.

OK, now the next thing I want to do, and I can see I'm not really going to get time to do it, but I'd like to give you the results of it. If I look at a box, a little cube, of area delta squared in the S1, S2 space. And I look at what that maps in to, in terms of the interarrivals x1 and x2, if you just map these points into this, you see that this square here is going into a parallelepiped.

And if know a little bit of linear algebra-- well, you don't even need to know any linear algebra to see that no matter where you put this little square anywhere around here it's always going to map into the same kind of a parallelepiped. So if I take this whole area and I break it up into a grid, this area is going to be broken up into a grid where it's things are going down this way. And this way, we're going to have a lot of these little tiny parallelepipeds. If I have uniform probability density there, I'll have uniform probability density there.

The place where you need some linear algebra, if you're dealing with n dimensions

instead of 2, is to see that the volume of this delta cube here, for this problem here, is going to be identical to the volume of the delta cube here. OK, so when you get all done that, the area in the x1, x2 space where zero is less than x1 plus x2 is less than t is t squared over 2 again. And the probability density of these two interarrival intervals is again 2 over t squared. It's the same as the arrivals.

I'm going to skip this slide. I want to get the one other thing I wanted to talk about a little bit. There's a paradox. And the main interarrival time for Poisson process is one over lambda. If I come at time t and I start waiting for the next arrival, the mean time I have to wait is 1 over lambda. If I come in and I start looking back, at the last arrival, well, this says it's 1 over lambda times 1 minus e to the minus lambda t, which is what it is. We haven't derived that. But we know it's something.

So, what's going on here? I come in at time t, the interval between the last arrival and the next arrival, the mean interval there is bigger than the mean interval between arrivals in a Poisson process. That is a real paradox. And how do I explain that?

It means that if I'm waiting for buses, I'm always unlucky. And you're unlucky too and we're all unlucky. Because whenever you arrive, the amount of time between the last bus and the next bus is bigger than it ought to be. And how do we explain that?

Well, this sort of gives you a really half-assed explanation of it. It's not very good. When we study renewals, we'll find a good explanation of it.

First choose a sample path of a Poisson process. I mean, start out with a sample path. And that has arrivals going along. And we want to get rid of zero because funny things happen around zero. So we'll start at one eon and stop at n eons. And we'll look over that interval.

And then we'll choose t to be a random variable in this interval here. OK, so, we choose some t after we've looked at the sample path. And then we say for this random value of t, how big is the interval between the most recent arrival and the

next arrival.

And what you see is that since I have these arrivals of a Poisson process laid out here, the large intervals take up proportionally more area than the little intervals. If I have a bunch of little tiny intervals and some big intervals and then I choose a random t along here, I'm much more likely to choose a t in here, proportionally, than I am to choose a t in here.

This mean time between the arrivals is the average of this, this, this, this, and this. And what I'm looking at here is I picked some arbitrary point in time. These arbitrary points in time are likely to lie in very large intervals. Yes?

AUDIENCE: Can you please explain again what is exactly is the paradox?

PROFESSOR: The paradox is that the mean time between arrival is one over lambda. But if I arrive to look for a bus and the buses are Poisson, I arrive in an interval whose mean length is larger than 1 over lambda. And that seems strange to me.

Well, I think I'll stop here. And maybe we will spend just a little time sorting this out next time.