

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: I guess we should start. This is the last of these lectures. The final will be on next Wednesday, as I hope you all know by this time, in the ice rink, whatever that means.

And there was some question about how many sheets of paper you could bring in as crib sheets. And it seems like the reasonable thing is four sheets, which means you can bring in the two sheets you made up for the quiz plus two more. Or you can make up four new ones if you want or do whatever you want.

I don't think it's very important how many sheets you bring in, because I've never seen anybody referring to their sheets. I mean, it's a good way of organizing what you know to try to put it on four sheets of paper.

I want to mostly review what we've done throughout the term, with a few more general comments thrown in. I thought I'd start with martingales, because we then completely finish what we wanted to talk about last time.

And the Strong Law of Large Numbers was left slightly hanging. And I want to show you how to do that in a little better way. And also show you that it's a more general theorem than it appears to be at first sight.

So let's go with martingales. The basic definition is a sequence of random variables is a martingale, if, for all elements of the sequence, the expected value of Z_n , given all of the previous values, is equal to the random variable, Z_{n-1} .

Remember, and we've talked about this a number of times, when you're talking about the expected value of one random variable, given a bunch of other random variables, you're only taking the expectation over the first part.

You're only taking the expectation over Z_{n-1} . And the other quantities are still random variables. Namely, you have an expected value Z_{n-1} , for each sample value of Z_{n-1} , all the way down to Z_1 . And what the definition says is it's a martingale only if, for all sample values of those earlier values, the expected value is equal to the sample value of the most recent one.

Namely, the memory is all contained right in this last term, effectively. At least as far as expectation is concerned. Memory might be far broader than that for everything else.

And the first thing we did with martingales is we said the expected value was again, if you're only given part of the history, if you're only given the history from i back to 1 , where i is strictly less than $n-1$, that expected value is equal to Z_i .

So no matter where you start going back, the expected value of Z_n is the most recent value that is given. So if the most recent value given is Z_1 , then the expected value of Z_n , given Z_1 , is Z_1 . And also along with that, you have the relationship, the expected value of Z_n is equal to the expected value of Z_i , just by taking the expected value over Z_{i-1} . So all of that's sort of straightforward.

We talked a good deal about the increments of a martingale. The increments, $X_n = Z_n - Z_{n-1}$, are very much like the increments that we have when a renewal process, when a Poisson process, all of these processes we talked about we can define in various ways.

And here we can define a martingale in two ways also. One is by the actual martingale itself, which are, in a sense, the sums of the increments. And the other ways in terms of the increments. And the increments satisfy the property that the expected value of X_n , given all the earlier values, is equal to 0. Namely, no matter what are all the earlier values are, X_n has mean 0 in order to be a martingale.

A good special case of this is where X_n is equal to U_n times Y_n , where the U_n are IID, equiprobable 1 and minus 1. And the Y_n 's are anything you want them to be. It's just that the U_n 's have to be independent of

the $Y_{sub\ i}$.

So I think this shows that in fact martingales are really a pretty broad class of things. And they were invented to talk about fair gambling games, where they wanted to give the gambler the opportunity to do whatever he wanted to do.

But the game itself was defined in such a way that, no matter what you do, the game is fair. You establish bets in one whatever things you want to. And when you wind up with it, the expected value of $X_{sub\ n}$, given the past, is always 0. And that's equivalent to saying the expected value of $Z_{sub\ n}$, given the past, is equal to $Z_{sub\ i}$.

Examples we talked about are 0 mean random walks and products of unit-mean IID random variables. So they're both these product martingales, and there are these sum martingales. And those are just two simple examples, which come up all the time.

Then we talked about submartingales. A submartingale is like a martingale, except it grows with time. And we're not going to talk about supermartingales, because a supermartingale is just a negative submartingale. So we don't have to talk about that.

A martingale is a submartingale. So anything you know about submartingales applies to martingales also. So you can state theorems for submartingales and they apply to martingales just as well. You can say stronger things very often about martingales.

And then we have the same theorem for submartingales. Now that should say, and it did say, until my evil twin got a hold of it, if Z_n is a submartingales, then for n greater than i , greater than 0, this expected value is greater than or equal to Z_i . And the expected value Z_n is greater than equal to the expected value of Z_i .

In other words, this theorem, for submartingales, is the same as the corresponding theorem for martingales, except now you have inequalities there, just like you have

inequalities in the definition of the submartingales. So there's nothing strange there.

Then we found out that, if you have a convex function, from the reals into the reals, then Jensen's inequality says that the expected value of h of X is greater than or equal to h of the expected value of x . We showed a picture for that you remember. There's a convex curve. There's some straight line. And what Jensen's inequality says is you take an average over the expected value of X , and you're somewhere above the line. And you take the average first, and you're sitting on the line.

So if h of X is convex, that's what Jensen's inequality is. And it follows from that that, if Z_n is a submartingale-- and that includes martingales-- and h is convex and the expected value of h of X is finite, then h of Z_n is a martingale also.

In other words, if you have a martingale Z sub n , the expected value of Z sub n is a submartingale. The expected value of E to R Z_n is a martingale. Use whatever convex function you want to, and you wind up, martingales go into submartingales.

You can't get out of the range of submartingales that easily. We then talked about stopped martingales and stopped submartingales. We said a stopped process, for a possibly defective stopping time-- now you remember what a stopping time is? A stopping time is a random variable, which is a function of everything that takes place up until the time of stopping.

And you have to look at the definition carefully, because stopping time comes in too many places to just say it and understand what it means.

But it's clear what it means, if you view yourself as an observer watching a sequence of random variables, of sample values of random variables, one after another. And after you see a certain number of random variables, your rule says, stop. And then you don't observe anymore.

So you just observe this finite number. And then you stop at that point. And then you're all done. If it's a possibly defective stopping rule, then you might keep on going forever, or you might stop. You don't know what you're going to do.

The stopped process $Z_{\text{sub } n \text{ star}}$ is a little different from what we were doing before. Before what we were doing is we were sitting there observing this process. At a certain point, the stopping rule said stop. And before, we were very obedient. And when the stopping rule told us to stop, we stopped.

Now, since we know a little more, we question authority a little more. And when the stopping rule says stop, we break things into two processes. There's the original process, which keeps on going. And there this stopped process, which just stops.

And it's convenient to have a stopped process instead of just a stopping rule. Because with a stopped process, you can look at any time into the future, and if it's already stopped, you know what the stopped value is.

You know what it was when it stopped. You don't necessarily know when it stopped, by looking at in the future. But you know that it did stop.

So the stopped process, well, it says here what it is. It satisfies the stopped value at time n as equal to $Z_{\text{sub } n}$, if n is less than or equal to the stopping time J , and $Z_{\text{sub } n \text{ star}}$ is equal to $Z_{\text{sub } J}$, if n is greater than J .

So you get up to the stopping timing, and you stop. And then it just stays fixed forever after. And the nice theorem there is that the stopped process for a submartingale, with a possibly defective stopping rule, is a submartingale again.

What that means is it's just a concise way of writing, the stopped process for a martingale is a martingale in its own right. And the stopped process for a submartingale is a submartingale in its own right. So the convenient thing is, you can take a martingale, you can stop it, you still have a martingale. And everything you know about martingales applies to this stopping process.

So we're getting to the point where, starting out with a martingale, we can do lots of things with it. And that's the whole mathematical game. With a mathematical game, you build up theorems from nothing. As an experimentalist or an engineer, you sort of try to figure out those things from the reality around you. Here, we're just building it up.

And the other part of that theorem says that the expected value of Z_1 is less than or equal to the expected value of Z_n^* , is less than or equal to the expected value of Z_n for a submartingale. And they're all equal to a martingale.

In other words, the marginal expectations for a martingale, they start out at Z_1 . They stay at Z_1 . And for the stopped process, they stay at that same value.

And that's not too surprising. Because if you have a martingale, if you go until you reach the stopping point, from that stopping point on, the martingale has mean 0, from that point on. Not the martingale itself, but the increments of the martingale have mean 0, from that point on.

And the stopped process has mean 0. In other words, the stopped process, the increments are actually 0. Whereas for the original process, the increments wobble around. But they still have mean 0. So this is a very nice and useful thing to know.

If you look at this product martingale, Z_n is E to the r_n minus n gamma or r . Why is that a martingale? How do you know it's a martingale?

Well, you look at the expected value of this. And it's expected value of this. The expected value of E to the r_n is the moment generating function of Z_n , of s sub n . It's moment generating function of E to the r_n . And the moment generating function of E to the r_n is just E to the n gamma of r .

So this is clearly something which should be a martingale, because it just keeps at that level all along. If you have a stopping rule, such as a threshold crossing, then you've got a stopped martingale.

And subject to some little mathematical nitpicks, which the text talks about, this leads you to the much more general version of Wald's identity, which says that the expected value of Z , at the time of stopping, is equal to the expected value of E to the rJ minus J gamma of r equals 1.

This you remember is what Wald's identity was when we were just talking about

random walks. And this was a more general version, because it's talking about general stopping rules, instead of just to thresholds. But it does have these little mathematical nitpicks in it, which I'm not going to talk about here.

Then we have Kolmogorov's submartingale inequality. We talked about all of these things last time. So we're going pretty quickly through them. The submartingale inequality is really the Markov inequality souped up.

And what it says is, if you have a non-negative submartingale, that can include a non-negative martingale, for any positive integer m , the probability that the maximum is a $Z_{i, \text{ from } 1 \text{ to } m}$, is greater than or equal to a , is less than or equal to the expected value of $Z_{\text{sub } m}$ over a .

You see all that the Markov inequality says is the probability that $Z_{\text{sub } m}$ is greater than or equal to a , is less than or equal to this. This puts a lot more teeth into it, because it lets you talk about all of these random variables, up until time m . And it says the maximum of them satisfies this inequality.

I mean, we always knew that the Markov inequality was very, very weak. And this is also pretty weak. But it's not quite as weak, because it covers a lot more things. If you have a non-negative martingale-- this is submartingales, this is martingales.

You see here, with submartingales, the expected value of $Z_{\text{sub } m}$ keeps increasing with m . So there's a trade-off between making m large and not making m large.

If you're dealing with a martingale, then expected value $Z_{\text{sub } m}$ is constant over all time. It doesn't change. And therefore, you can take this inequality here. You can go to the limit, as m goes to infinity. And you wind up with a probability, the sup of Z_m , greater than or equal to a , is less than or equal to the expected value of the first of those random variables, the expected value of Z_1 divided by a .

So this looks like a very powerful inequality. It turns out that I don't know many applications of that. And I don't know why. It seems like it ought to be very useful. But I know one reason, which is what I'm going to show you next, which is how you can really use the submartingale inequality to make it do an awful lot of things that

you wouldn't imagine that it could do otherwise.

First, you go to the Kolmogorov version of the Chebyshev inequality. This has the same relationship to the Kolmogorov submartingale inequality as Chebyshev has to Markov. Namely, what you do is, instead of looking at the random variables $Z_{\text{sub } n}$, you look at the random variable $Z_{\text{sub } n}^2$.

And what do we know now? If $Z_{\text{sub } n}$ is a martingale or a submartingale, $Z_{\text{sub } n}^2$ is a martingale or submartingale also. Namely, well, the only thing we can be sure of is that $Z_{\text{sub } n}^2$ is a submartingale.

But if it's a submartingale, then we can apply this inequality again. And what it tells us, in this case, is the probability at the maximum of the magnitudes of these random variables. Probably the maximum is greater than or equal to b , is less than or equal to the expected value of $Z_{\text{sub } m}^2$ over b^2 .

So before, just like the Markov inequality, the Markov inequality only works for non-negative random variables. You go to the Chebyshev inequality, because that works for negative or positive random variables. So that makes it kind of neat. And then what you have is this thing, which goes down as $1/b^2$, which looks a little stronger. But that's not the real reason that you want to use it.

Now, this inequality here only works for the first m values of this random variable. What we're usually interested in here is what happens as m gets very large. As m gets very large, this thing, very often, blows up.

So this [INAUDIBLE] does not really do what you would like an inequality to do. So what we're going to do is, first, we're going to say, if you had this inequality here, then you can lower bound this by taking just a maximum, not over 1 up to m , but only over 2 up to m .

Now why do we want to do that? Well, hold on and you'll see. But anyway this is bigger than, greater than or equal to, this. So what we're going to do now is we're going to take this inequality. We're going to use it for m equals 2 to the m , for m equals 2 to the k plus 1 , m equals 2 to the k plus 2 , all the way up to infinity.

And so we're going to find the probability of the union over j greater than or equal to k of this quantity here, but now just maximized over 2 to the j minus 1 , less than n , less than or equal to 2 to the j . And then the maximum of Z sub n , greater than or equal to. And now, for each one of these j 's here, we'll put in whatever b sub j we want.

So the general form of this inequality then becomes. We have this term on the left. We use the union bound. And we get this term on the right.

So at this point, we have an inequality, which works for all n , instead of just for values smaller than some given amount. So this is sort of a general technique for taking an inequality, which only works up to a certain value, and extending it so it works over all values. You have to be pretty careful about how you choose b sub j .

Now what we're going to do is say, OK. And remember, what is happening here is we started out with a submartingale or a martingale. When we take Z squared, we still have a submartingale. So we can use a submartingale inequality, which is what we're doing here. We're using the submartingale inequality on Z squared rather than on Z . And Z squared is non-negative, so that works there.

Then we go down to this point. We take a union over all of these terms. And note what happens. Every n is included in one of these terms, every n beyond 2 to the k . So if we want to prove something about the limiting values of Z sub n , we have everything included there, everything beyond 2 to the k .

But as far as the limit is concerned, you don't care about any initial finite set. You care what happens after that initial finite set.

So what we have then [INAUDIBLE] of these terms, less than or equal to this term. When I apply this to a random walk S sub n , S sub n is a submartingale, at this point. The expected value of x squared will assume a σ squared.

The expected value now S sub n , or Z sub n is we'll call it, is the sum of these n IID random variables. So the expected value--

AUDIENCE: 10 o'clock.

PROFESSOR: The expected value of Z to the $2J$ is just 2 to the J times the expected value of x squared, in other words, σ squared. [INAUDIBLE] just doing this for a 0 mean [INAUDIBLE] variable, because [INAUDIBLE] given an arbitrary non- 0 [INAUDIBLE] random variable. You can look at it as to mean plus a random variable, which is 0 mean. So that's the same ideas we're using here.

So we take this inequality now, and I'm going to use for b sub J , $3/2$ to the J . Why $3/2$ to J ? Well you'll see in just a second. But when I use $3/2$ to the J here I get the maximum over S sub n , greater than or equal to $3/2$ to the J .

And over here I get b sub J squared is $9/4$ to the J . And here I have 2 the J also. So when I sum this, it winds up with $8/9$ to the k times 9 sigma squared.

So what I have now is something where, when k gets larger, this term is going to 0 . And I have something over here, well that doesn't look quite so attractive, but just wait a minute.

What I'm really interested in is not S sub n . But I'm interested in S sub n over n . For the strong law of large numbers, I'd like to show that S sub n over n approaches a limit. And n in this case runs between 2 to the J minus 1 and 2 to the J .

So when I put that in here-- we'll see what that amounts to in the next slide. For the strong law of large numbers, what our theorem says is that the probability of the set of sample points, for which S sub n over n equals 0 , that set of sample points has probability y . So the proof of that, if I pick up this equation from the previous slide, and when I lower bound the left side of this, what I'm going to get, I'm going to divide by n here. And I'm going to divide by something a little bit smaller, which is 2 to the J minus 1 here. So I get the maximum of S sub n over n , greater than or equal to 2 times $3/4$ to the J .

Now you see why I picked-- I think you see at this point why I picked the sub j the way I did. I wanted to pick it t to be smaller than 2 to the J . And I wanted to pick it to

be big enough that it drove the right hand term to 0.

So now we're done really. Because, if I look at this expression here, a sample sequence S_n of ω , that's not contained in this union, has to approach 0. Because these terms from 2 to the J minus 1 to 2 to the J , in order to be in this set, they have to be greater than or equal to $2 \times 3/4$ to the J .

As j gets larger and larger, this term goes to 0. So the only terms that exceed that are terms that are arbitrarily small. So the complement of this set is the set of terms for which S_n over n does not approach 0.

But the probability of that is $8/9$ to the k times time some garbage over here. So now it's true for all k . The terms which approach 0, namely the sampled values for which S_n over n approaches 0 are all complimentary to this set.

So the probability that S_n over ω over n approaches 0 is greater than 1 minus this quantity here. That's true for all k . And since it's true for all k , this term goes to 0. And the theorem is proven.

Now why did I want to go through this. There are perhaps easier ways to prove the strong law of large numbers, just assuming that the variance is finite. Why this particular y ?

Well, if you look at this, it applies to much more than just sums of IID random variables. It applies to arbitrary martingales, so long as these conditions are satisfied. It applies to these cases, like where you have a random variable, which is plus or minus 1 times some arbitrary random variable.

So this gives you sort of a general way of proving strong laws of large numbers for strange sequences of random variables. So that's the reason for going through this. We now have a way of proving strong laws of large numbers for lots of different kinds of martingales, rather than just for this set of things here.

So let's move on back to Markov chains, countable or finite state. I'm moving back to chapter three and five in the text, mostly chapter five, and trying to finish some

sort of review of what we've done.

When I look back at what we've done, it seems like we've proven an awful lot theorems. So all I can do is talk about the theorems. I should say something, again, this last day, on this last lecture, about why we spend so much time proving theorems.

In other words, we've just proven a theorem here. I promised you I would prove a theorem every lecture, along with talking about why they're important and so on. And most of you are engineers or you're scientists in various fields. You're not mathematicians. Why should you be interested in all these theorems.

Why should you take abstract courses, which look like math courses? And the reason is this kind of stuff is more important for you than it is for mathematicians.

And it's more important for you, because when you're dealing with a real engineering or real scientific problem, how do you deal with it? I mean, you have a real mess facing you. You spend a lot of time trying to understand what that mess is all about.

And you don't form a model of it, and then apply theorems. What you do is to try to understand it. You look at multiple models. When we were looking at hypothesis testing, we said we're going to assume a priori probabilities. I lied about that a little bit.

We we're not assuming a priori probabilities. We we're assuming a class of probability models, each of which had a priori probabilities in them. And then we said something about that class of probability models. And saying something about that class of probability models, we were able to say a great deal more than you can say if you refuse to even think about a model, which doesn't have a priori probabilities in it.

So by looking at lots of different models, you can understand an enormous number of things without really having any one model which describes the whole situation for you. And that's why we try to prove theorems for models, because then, when

you understand lots of simple models, you have these complicated physical situations, and you play with them.

You play with them by applying various simple models that you understand to them. And as you do this, you gradually understand the physical process better. And that's the way we discover things.

OK, end of lecture. Not end of lecture, but end of partial lecture about why you want to learn some mathematics.

The first passage time from state i to j , remember, is the smallest n , when you start off in state i , at which you get to state j . You start off in state i . You jump from one lily pad to another. You eventually wind up at lily pad number j .

And we want to know how long it takes you to get to j . That's a random variable, obviously. And this T_{ij} , as possibly the effective random variable, that has the probability mass function. It's the definition of what this probability mass function is. And it has a distribution function.

And the probability mass function-- you probably remember how we derived this. We derived it by sort of crawling up on it, by looking at it, first, for n equals 1, in which case it's just a transition probability of n equals 2.

In which case, it's the probability that you first go to k , and then in n minus 1 steps, you go to j . But you have to leave j out, because if you go to j in the first step, you've already had your first passage. so

We define a state to be recurrent, if T_{jj} is non-defective. And we define it to be transient otherwise. In other words, if it's not certain that you ever get to state j , then you define it to be transient, if it's recurrent and it's positive recurrent, if the expected value of T_{jj} is less than infinity. And it's null recurrent otherwise.

How do we know how to analyze this? Well we study renewal processes. And if you look at the renewal process where you've got a renewal every time you hit state j , you start out on stage j . The first time you hit state j , that's a renewal. The next time

you hit state j , that's another renewal. You have a renewal process where the interrenewal time is a random variable, which has the PMF F_{ij} .

Excuse me, if you have a renewal process, if you start in state j , where T_{jj} is the amount of time before renewal occurs, from that time on, you get another renewal with another random variable with the same distribution as T_{jj} . And F_{ij} is the PMF of that renewal time. And F_{ij} is the distribution function of it.

So then when we define the state j as being recurrent, what we're really doing is going back to what we know about renewal processes and saying a Markov chain is recurrent if the renewal process that we define for that countable state Markov chain has these various properties for this renewal random variable.

For each recurrent j , there's an integer renewal counting process $N_{jj}(t)$. You start in state j at time t , which is after t steps of the Markov process. What you're interested in is how many times have you hit state j , up until time t . That's the counting process we talk about in renewal theory.

So $N_{jj}(t)$ is the number of visits to j starting in j . And it has the interrenewal distribution F_{jj} , which is that quantity up there. We have a delayed renewal counting process $N_{ij}(t)$, if we count visits to j , starting in i .

We didn't talk much about delayed renewal processes, except for pointing out that when you have a delayed renewal process, it really is the same as a renewal process. It just has some arbitrary amount of time that's required to get to state j for the first time and then keep recurrent on.

Even if the expected time to get to j for first time is infinite, and the expected time for renewals from j to j is finite, you still have this same renewal process. You can even lose an infinite amount of time at the beginning, and you amortize it over time. Don't ask me why you can amortize an infinite amount of time over time. But you can.

And actually if you read about delayed renewal processes, you see why you actually get that.

So all states in a class are positive recurrent, or all are null recurrent, or all are transient. We've proved that theorem. It wasn't really a very hard theorem to prove. And you can sort of see that it ought to be.

Then we define the chain as being irreducible, if all state pairs communicate. In other words, if for every pair of states, there's a path that goes from one state to the other state. This is intuitively a simple idea, if you have a finite state and Markov chain. If you have a countably infinite state Markov chain, it seems to be a little more peculiar. But it really isn't.

For a countably infinite state Markov chain every state has a finite number. And you can take every pair of states. You can identify them. And you can see whether there's a path going from one to the other.

For all of these birth-death processes we've talked about, I mean, it's obvious whether the states all communicate or not. You just see if there's any break in the chain at any point. And it really looks like a chain. It's a node, two transitions, another node, two transitions, another node. And that's just the way chains are supposed to work.

An irreducible class might be positive recurrent. It might be null recurrent. Or it might be transient. And we already have seen what makes a state null recurrent or transient. And it's the same thing for the class. We started out by saying a state is either null recurrent, positive recurrent, or transient depending on this renewal process associated with it. And now there's this theorem, which says that if one node in a class of states is positive recurrent, they all are.

And you ought to be able to sort of see the reason for that. If I have one state which is positive recurrent, it means that the expected time to go from this state to this state is finite. Now if I had some other state, I have to go from here to there. I can go through here and then off to there.

So the amount of time it takes to get to there, and then from there to there, is also finite, expected amount, and the same backwards. So that was the way we proved

this.

If we have an irreducible Markov chain-- now this is the theorem you really use all the time. This is sort of says how you operate with these things. It says the steady state equations-- they're the equations you've used in half the problems you've done with Markov chains-- if these equations have a solution for the π_j 's, remember the Markov chain is defined in terms of the transition probabilities P_{ij} .

We solve these equations to find out what the steady state probabilities π_j are. And the theorem says, if you can find the solution to those equations-- π_j 's have to add up to 1-- then the solution is unique. The π_j 's are equal to 1 over the mean time to go from that state back to that state again.

And what does that mean? What that really gives you is not a way to find π_j . It gives you a way to find a T_{jj} . Because these equations are more often the way that you solve for the steady state probabilities. And then that gives you a way to find the mean recurrence time between visits to this given state.

And what else does this theorem say? It says if the states are positive recurrent, then the steady state equations have a solution. So this is an if and only if kind of statement.

It relates these equations, these steady state equations, to solutions and says, if these equations have a solution, then in fact you have the steady state equations. They satisfy all these relationships about mean recurrence time. And if the states are positive recurrent, then those equations have a solution. And in the solutions, the π_j 's are all possible.

So it's an infinite set of equations, so you can't necessarily solve it. But you sort of know everything there is to know about it, at this point. Well, there's one other thing, when you have a birth-death chain, these equations simplify a great deal.

The counting processes under positive recurrence have to satisfy this equation. And

my evil twin brother got a hold of this and left out the n in the copy that you have. And I spotted it when I looked at it just a little bit. He was still sleeping, so I've managed to find it. So it's corrected here.

And what does that say? It says, when you have positive recurrence, if you look from 0 out to t , and you count the number of times that you hit state j , that's a random variable. If you take that and divide by n , you look from time 0 out to time N , N sub ij of N , it's the number of times you visit state j .

You divide that by N , and you go to the limit. And there's a strong law of large numbers there, which was a strong law of large numbers for renewal processes, which says that it has a limit with probability 1. And this says that limit is π sub j .

And that's sort of obvious, again. I mean, visualize what happens. You start out in state j . For one unit of time, you're in state j . Then you go away from state j , and for a long time you're out in the wilderness. And then you finally get back to state j again. Think of a renewal reward process, where you get 1 unit of reward every time you're in state j and 0 reward every time you're not in state j .

That means every interrenewal period, you pick up one unit of reward. Well, this is what that says. It says that the fraction of those visits to state j -- that out of the total visits in the Markov chain, the ones that go to state j have probability π sub j .

So again this is another relationship with these steady state probabilities. The steady state probabilities tell you what these mean recurrence times are. And that tells you what this is. This, in a sense, is the same as this. Those are just sort of the same results. So there's nothing special about it.

We talked a little bit about the Markov model of age of renewal process for any integer valued renewal process, you can find a Markov chain which gives you the age of that process. You visualize being in state j . And you visualize being in state 0, of this Markov model, at the point where you have a renewal. One step later, if you have another renewal, that happens with probability P sub 00 , you go back to state 0 again.

If you don't have a renewal in the next time, you go to state 1. From state 1, you might go to state 2. When you're in state 2, it means you're two time units away from state 0. If you go back to state 0, it means you have a renewal in three time units. Otherwise you go to state 3. Then you might have a renewal and so forth.

So for this very simple kind of Markov chain, this tells you everything there is to know, in the sense, about integer value renewal processes. So there's this nice connection between the two. And it lets you see pretty easily about when you have no recurrence.

Now we spend a lot of time talking about these birth-death Markov chains. And the easy way to solve for birth-death Markov chains is to say intuitively that between any two adjacent states, the number of times you go up has to equal the number of times you go down, plus or minus 1. If you start out here and you end up here, you're going this way one more time than you've gone that way and vice versa.

And combining that with the steady state equations that we now have been talking about, it must be that the steady state probability of π_i is $\pi_i P_i$ is the probability of going from state 2 to state 3. It's the probability of being in state 2 and making a transition to state 3.

This probability here is the probability of being in state 3 and going to state 2. And we're saying that asymptotically, as you look over an infinite number of transitions, those two have to be the same.

The other way to do it, if you like algebra, is to start out with a steady state equation. And you can derive this right away. I think it's nicer to see intuitively why it has to be true. And what that says is if ρ_i is equal to $P_i / (Q_i + 1)$, P_i is the up transition probability. Q_i is the down transition probability. ρ_i is the ratio of the two state probabilities. And that's equal to this equation here.

That's just how to calculate these things. And you've done that. Let's go on to Markov processes. I have no idea where I'm going to finish up. I had a lot to do. I better not waste too much time.

Remember what a Markov process is now. At least the way we started out thinking about, it's a Markov chain along with a holding time. And each state is a Markov chain. And the holding times are exponential, to be a countable state Markov process.

So we can visualize it as a sequence of states, X_0, X_1, X_2, X_3 . And a sequence of holding times, U_1, U_2, U_3, U_4 . these are all random variables. And this kind of dependence diagram says what random variables depend on what random variables.

U_1 , given X_0 , is independent of the rest of the world. U_2 , given X_1 , is independent of the rest of the world, and so forth. And if you look at this graph here and you visualize the fact that because of Bayes' rule, you could go both ways on this.

In other words, if this, given this, is independent of everything else, we can go through the same kind of argument. And we can make these arrows go the opposite way. And we can say, if we just consider these states here, we can say that, given X_3 , U_4 is independent of X_2 and also independent of U_3 and X_1 and U_2 and so forth.

So if you look at the dependence graph of a Markov chain, which is which states depend on which other states, those arrows there that we have, which make it easier to see what's going on, you can take them off. You can redraw them in any way you want to and look at the dependencies in the opposite way.

Now to understand what the state is at any time t , there's an equation to do that. It's an equation that isn't much help. I think it's more help to look at this and to see from this what's going on.

You start in some state that's 0. And starting in state 0, there's a holding time in U_0 . The holding time is U_1 . And you stay in. And the time in U_1 is an exponential random variable with rate U_i . That's what this says. So at the end of that holding time, you go from state i to some other state.

This is the state you go to. The state you go to is according to the Markov chain probabilities. And it's state j in this case. You stay in state j until the holding time U_2 , which is a function of j , finishes you up at this time and so forth. So if you want to look at what state you're in at a given time, namely pick a time here and say what's the state at this time, as a random variable.

So what you have to do then is you have to climb your way up from here to there. And you have to talk about the value of S_1 , S_2 , and S_3 . And those are exponential random variables. But they're exponential random variables that depend on the state that you're in.

So as you're climbing your way up and looking at this sample function of the process, you have to look at U_1 and X_0 . X_0 defines what U_1 is, as a random variable. It says that U_1 is an exponential random variable, with rate U_i .

So you get to here, then you have some holding time here, which is a function of j and so forth, the whole way up. Which is why I said that an equation for $X(t)$, in terms of these S 's is not going to help you a great deal. Understanding how the process is working I think helps you a lot more.

We said that there were three ways to represent a Markov process, which I'm giving here in terms just of Markov chains. The first one-- and the fact that these are all for M/M/1 doesn't make any difference. It's just these three general [INAUDIBLE].

One of them is, you look at it in terms of the embedded Markov chain. For this embedded Markov chain, the transition probabilities, when you're in state 0 in an M/M/1 queue, what's the next state you go to?

Well the only state you can go to is state 1. Because we don't have any self transitions. So you go up to state 1 eventually. From state 1, you can go that way, with probability $\mu / (\lambda + \mu)$. Or you can go this way, with probability $\lambda / (\lambda + \mu)$, and so forth the whole way out.

The next way of describing it, which is almost the same, is instead of using the transition probabilities and the embedded chain, you look directly at the transition

rates for the Poisson process.

Meaning the transition rates are the new sub λ 's associated with the different states. When you get in state i , the amount of time you spend in state i is an exponential random variable. And when you make a transition, you're either going to go to one state or another state, in this case. In general, you might go to any one of a number of states.

Now if I tell that we start out in state one and the next state we go is state 2, now I ask you what's the expected amount of time that that transition took? What's the answer? Is it λ_1 , λ_2 , or is it μ_1 ?

Anybody awake out there?

AUDIENCE: Sir, could you repeat the question?

PROFESSOR: Yes. The question is, we started out in state 1. Given that we started out in state 1 and given that the next state is state 2, what's the amount of time that it takes to go from 1 to 2? It's an exponential random variable. What's the rate of that random variable?

AUDIENCE: $\lambda + \mu$.

PROFESSOR: What?

AUDIENCE: $\lambda + \mu$.

PROFESSOR: $\lambda + \mu$? Yes. $\lambda + \mu$ in the case of M/M/1 queue. If you have an arbitrary change, why the amount of time that it takes is μ^{-1} . This is just back to this old thing about splitting and combining of Poisson processes.

When you have a combined Poisson process, which is what you have here, when you're in state i , there's a combined Poisson process, which is running, which says you go right with probability λ , you go left with probability μ for an M/M/1 queue.

And you can look at it in terms of, first, you see what the next state is. And then you ask how long did it take to get there? Or you look at in terms of how long does it take to make a transition and then which state did you go to?

And with these combined Poisson processes, those two questions are independent of each other. And if there's one thing you remember from all of this, please remember that. Because it's something that you use in almost every problem that you do with Markov chains and Markov processes. It just comes up all the time.

This final version here is looking at the same Markov process, but looking at it in sample time instead of looking at the embedded queue.

Now the important thing here is, when you look at it in sample time, you might not be able to do this. Because with this entire cannibal state Markov chain, you might not be able to define these self-loop transition probabilities. Because these numbers might get too large. But for the $M/M/1$ queue, you can do it.

The important thing is that the steady state probabilities you find for these states are not the same as the steady state probabilities you find for the embedded Markov chain. They are in fact the same as the steady state probabilities for the Markov process itself.

That's these steady state probabilities are the fraction of time that you spend in state j . And this is a sample time Markov process. It is the same fraction of time you spend in state j .

Here you have this embedded chain. And for example, in the embedded chain, the only place you go from state 0 is state 1. Here from state 0, you can stay in state 0 for a long time. Because here the increments of time are constant.

We can look at delayed renewal reward theorems for the renewal process to see what's going on here, for the fraction of time we spend in state j . We look at that picture up there. We start out in state j , for example. Same as the renewal reward process that we had for a Markov chain.

We got a reward of 1 for the amount of time that we stay in state j . After that, we're wandering around in the wilderness. We finally come back to state j again. We get 1 unit of reward times the amount of time we spend here.

In other words, we're accumulating reward at a rate of 1 unit per unit time, up to there. So the average reward we get per unit time is the expected value of U of j divided by the expected interrenewal time, which is 1 over μ_j times the expected time, from one renewal to the next.

Which tells us that the fraction of time we spend in state j is equal to the fraction of transitions that go to state j , divided by the rate at which we leave state j , times the expected number of overall transitions per unit time.

This is an important result. Because depending on what M sub i is, depending on what the number of transitions per unit time is, it really tells you what's going on. Because all of these bizarre Markov processes that we've looked at are bizarre because of the way that this behaves. This can infinite or can be 0.

At this point, we've been talking about the expected number of transitions per unit time as a random variable, as a limit in probability 1, given that we start in state i . And suddenly, we see that it doesn't depend on i at all.

So there is some number, M bar, which is the expected number of transitions per unit time, which is independent of what state we started in. We call that M bar instead M sub i . And that's this quantity here.

And what we get from that it is the fraction of time we spend in state j is proportional to π_j over μ_j . But since it has to add up to 1, we have to divide it by this quantity here. And this quantity here is one over-- this is the expected number of transitions per unit time.

And if we try to get the π_j 's from P_j 's, the corresponding thing, as we find out, the expected number transitions per unit time as a sum over i , P_i , times μ_i . You can play all sorts of games with these equations. And when you do so, all of those things become evident.

I would advise you to just cross this equation out. I don't know what it came from. But it doesn't mean anything.

We spent a lot of time talking about what happens when the expected number of transitions per unit time is either 0 or infinity. We had this case we looked at of an M/M/1 type queue, where the server got rattled as time went on.

And the server got rattled with more and more customers waiting. The customer's got discouraged and didn't come in. So we had a process where the longer the queue got, the longer time it took for anything to happen.

So that as far as the embedded Markov chain went, everything was fine. But then we looked at the process itself, the time that it took in each of these higher order states was so large, that, as a process, it didn't make any sense. So the P sub i 's were all 0. The π sub i 's all looked fine.

And the other kind of cases, where the expected number of transitions per unit time becomes infinite. And that's just the opposite kind of case, where, when you get to the higher ordered states, things start happening very, very fast.

The higher ordered state you go to, the faster the transitions occur. It's like a small child. I mean, the more excited the small child gets, the faster things happen. And the faster things happen, the more excited the child gets. So pretty soon things are happening so fast, the child just collapses. And if you're lucky, the child sleeps. So you can think of it that way.

We talked about reversibility. And reversibility for Markov processes I think is somewhat easier to see than reversibility for Markov chains. If you're dealing with a Markov process, we're sitting in state i for a while. At some time we make a transition. We go to state j . We sit there for a long time. Then we go to state k and so forth.

If we try to look at this process coming back the other way, we see that we're in state k . At a certain point, we had a transition. We had a transition into state j . And

how long does it take before that transition is over?

We're in state j , so the amount of time that it takes is an exponentially distributed random variable. And it's exponentially distributed with the same amount of time, whether we're coming in this way or whether we're coming in this way.

And that's the notion of reversibility. It doesn't make any difference whether you look at it from right to left or from left to right. And in this kind of situation, if you find the steady state probabilities for these transitions or you find the steady state fraction of time you spend in each state.

I mean, we just showed that if you look at this process going backwards, if you define all the probabilities coming backwards, the expected amount of time that you spend in state i or the rate for leaving state i is independent of right to left.

And a slightly more complicated argument says the P sub i 's are the same going right to left. And the fraction of time you spend in each state is obviously the same going from right to left as these limits occur.

So that gives you all these bizarre conditions for queuing, which are very useful.

I'm not going to say any more about that except the guessing theorem. The guessing theorem says suppose a Markov process is irreducible. You can check pretty easily whether it's irreducible or not. You can't necessarily check very easily whether it's recurrent.

And suppose P sub i is a set of probabilities that satisfies P sub i times Q sub ij equals P sub j times Q sub ji . In other words, this is the probability of being in state i , and the next transition is to state j . This is the probability of being in state j , and the next transition to state i .

This says that if you can find a set of probabilities which satisfy these equations, and if they also satisfy this condition, P sub i , μ sub i , less than infinity, then P sub i is greater than 0 for all i . P sub i is a steady state time averaged probability state i . The processes is reversible. And the embedded chain is positive recurring.

So all you have to do is solve those equations. And if you can solve those equations, you're done. Everything is fine. You don't have to know anything about reversibility or renewal theory or anything else. If you have that theorem, you just solve for those equations. Solve these equations by guessing what the solution is, and then you in fact have a reversible process.

So the useful application of this is that all birth-death processes are reversible if this equation is satisfied. And you can immediately find the steady state probabilities of them.

I'm not going to have much time for random walks. But random walks are what we've been talking about all term. We just didn't call them random walks until we got to the seventh chapter. But a random walk is a sequence of random variables, where each S_n in the sequence is a sum of some number of underlying IID random variables, X_1 up to X_n .

Well we're interested in exponential bounds on S_n for large n . These are known as Chernoff bounds. We talked about them back in chapter one. I'm not going to mention them again now.

We're interested in threshold crossings. If you have two thresholds, one positive threshold, one negative threshold, you would like to know what's the stopping time when S_n first crosses α ? Or what's the stopping time when it first crosses β ? What's the probability of crossing α before you cross β or vice versa? And what's the distribution of the overshoot, when you pass one of them?

So there all those questions. We pretty much talked about the first two. The question of overshoot, I think I mentioned this. The text doesn't say much about it. Overshoot is just a nasty, nasty problem.

If you ever have to find the overshoot of something, go look for a computer program to simulate it or something. You're not going to solve the problem very easily.

Fowler is the only book I know which does a reasonable job of trying to solve this.

And you have to be extraordinarily patient. I mean Fowler does everything in the nicest possible way. Or at least he always seem to do everything in the nicest possible way.

Most textbooks you look at, after you understand the subject, you look at and you say, oh, he should have done it this way. I've never had that experience with Fowler at all. Always, I look at it. I say, oh, there's an easier way to do it. I try to do it the easier way. And then I find something's wrong with it. And then I go back and say, ah, I got to do it the way Fowler did it.

So if you're serious about this field and you don't have a copy of this very old book, get it, because it's solid gold.

Suppose a random variable has a moment generating function, expected value of E to the zr over some positive region of r . And suppose it has a mean which is negative. The Chernoff bound says that for any α greater than 0 and any r in 0 to r plus, the probability that Z is greater than or equal to α is less than or equal to this quantity here.

You remember, we derived this. The derivation is very simple. It's a an obvious result. It's a little strange. Because this says that for this random variable it's complimentary distribution function has to go down as e to the minus r alpha.

Now all random variables can't go down exponentially as e to the minus r alpha. The reason for this is that these moment generating functions down exist for all alpha. So what it's really saying is where it exists, it goes down with alpha as e to the minus r alpha.

We then define the semi-invariant moment generating function. And then a more convenient way of stating the Chernoff bound was in this way. You look here. And you say, for a fixed value of n here, this probability of S sub n is greater than or equal to n a, is something which is going down exponentially with n .

And if you optimize over r , this bound is exponentially tight. In other words, if you try to replace this with anything smaller, namely which goes down faster, than for large

enough n , the bound will be false. So this is the tightest bound you can get when you optimize it over r .

So its exponential in n . Mostly we wanted to use it for threshold crossings. And for threshold crossings, we would like to look at it in another way. And we dealt with this graphically. Probability of S_n greater than or equal to α .

Now what we want to do is hold α constant. α is some threshold up there. We want to ask, what's the probability that after n trials, we're sitting above α ? And we'd like to try to solve that for different values of n .

The Chernoff bound, in this case, this quantity here is this intercept here. You take the semi-invariant moment generating function as convex. You draw this curve. You take a tangent of slope α over n . And you see where it hits here. And this is the exponent that you have. This is a negative exponent.

As you vary n , this tilts around on this curve. And it comes in to this point. It goes back out again. That's what happens to it. And that smallest exponent, as you vary n , is the most likely time at which you're going to cross that threshold.

And what we found, from looking at Wald's equality is that-- let me go on, because we're running out of time. Wald's identity for two thresholds says this. And the corollary says, if the underlying random variable is less than 0, and if the r at which the-- the second solution of γ of r equals 0.

You have this convex curve. γ is always equal to 0. There's some other value of r , for which γ is equal to 0. And that's r^* . And this says that the probability that we have crossed α at time j , where j is the time of first crossing, is less than or equal to e to the minus αr^* .

This bound is tight also. And that's a very nice result. Because that just says that all you got to do is find r^* . And that tells you what the probability of crossing a threshold is. And it's a very tight bound if α is very large. It doesn't make any difference what the negative threshold is, or whether it's there or not. This tells you the thing you want to know.

I think I'm going to stop at that point, because I have been sort of rushing to get to this point. And it doesn't do any good to keep rushing. So thank you all for being around all term. I appreciate it. Thank you.