

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, I guess it's time to get started. Next lecture, I'm going to try to summarize what we've done so that I want to try to finish what we're going to finish for the term today. And that means we have a lot of topics that all get slightly squeezed in together, including the martingale convergence theorem and the strengthening of the strong law of large numbers and the Kolmogorov's submartingale inequalities and stopped martingales. So all of these are fairly major topics.

One thing that it means is that we certainly aren't going to do much with the martingale convergence theorem. It's a major theorem and advanced work. What we're trying to do here is just give you some flavor of it. The other things, as we move up the chain there, we want to know more and more about it. And the things on top certainly explain what's going on in the things below.

OK, let's review what a martingale is. A sequence of random variables is a martingale if it satisfies this funny looking condition. When I write it this way, it's not completely obvious what it's saying.

But I think we know now the expected value of one thing, of one random variable, given a set of random variables, is really a random variable in its own right. And that's a random variable, which is a function of those conditioning random variables. So expected value of Z_n , given Z_{n-1} to Z_1 .

It's a random variable. It maps each sample point to the conditional expectation of Z_n , conditional in Z_1 to Z_{n-1} . For martingale, this expectation has to be this random variable, Z_{n-1} . It has to be the most recent random variable you've seen.

And then last time, we proved this lemma, a pretty major lemma, which says, for a martingale, expected value of Z_n is equal to Z_n , given not all of the past, but just a certain number of elements of the past Z_n conditional and Z_i , back to Z_1 . That's equal to Z_i . Expected value of Z_n is equal to the expected value of Z_i .

We didn't talk about this one last time. But this is obvious in terms of this. If you take the expected value of this over all of the conditioning random variables, then what you get is just the expected value of Z sub i . So that's what this says.

OK, question now. Why is it that if you condition on more random variables than just Z_n in the paths, if you condition on things into the future, including Z sub n its self, why is the expected value of Z_n given Z_m down to Z_n down to Z_1 ? Why is that equal to Z_n and not equal to Z_n minus 1?

If you understand this, you understand what conditional expectations are. If you don't understand it, you've got to spend some time really thinking about what conditional expectations are. OK, how many people can see why this, in fact, has to be Z_n and not something else?

AUDIENCE: We know Z_n already.

PROFESSOR: What?

AUDIENCE: Isn't it-- don't you know Z_n already?

PROFESSOR: That's right. Since you know the n already, then you know it. That's exactly what it says. And we'll talk more about this later because we actually use this a bunch of times when we're going on.

But what it says is that for any given value of each of these random variables, in other words, for a given value of Z sub n , the expected value of Z sub n -- I mean, if I just wrote it, it's what's the expected value of Z sub n , given Z sub n ?

What's the expected value of Z sub n -- I'll make it even easier. And you can see it more clearly. What's the expected value of Z sub n , given that Z sub n is equal to a particular value z sub n .

OK, now I hope you can see why it is that this is equal to $Z_{sub\ n}$. That's the only thing that the random variable $Z_{sub\ n}$ can be. For this sample point and the conditioning, we're assuming the sample value for which $Z_{sub\ n}$ is equal to $z_{sub\ n}$, and therefore that's what this random variable has to be.

That's what you said in one sentence. And I'm saying it in five sentences. It's hard enough that you want to say it in five sentences because this is not an obvious thing. It's not an obvious thing until you really have a sense of what these conditional expectations mean. And that's part of our function for the last couple of weeks, to sort out what those things mean because that's part of understanding what martingales are doing.

OK, when we go one step further on this, we've said that the expected value of $Z_{sub\ n}$, given $Z_{sub\ i}$ back to $Z_{sub\ n}$ is $Z_{sub\ i}$, so the expected value of $Z_{sub\ n}$, only given $Z_{sub\ 1}$, is equal to $Z_{sub\ 1}$. That's what this says, an expected value of $Z_{sub\ n}$ then is equal to the expected value of $Z_{sub\ 1}$. It says these marginal expectations are all the same. Yes?

AUDIENCE: Why don't you just say that the expectation of the Z s are constant. I mean, it seems like this is kind of a roundabout way of saying that.

PROFESSOR: Yeah, in a sense. But if you want to-- almost all the examples you can think of, it's hard to figure out what these $Z_{sub\ n}$'s are because the $Z_{sub\ n}$'s are given in terms of all of the previous random variables. If you want to sort out what a martingale is, and you can't even understand what $Z_{sub\ 1}$ is, then you're in trouble.

I mean, it makes the theorem more abstract to say the expected value of $Z_{sub\ n}$ is a constant random variable rather than saying which random variable it is. And it's obvious what random variable it is. I mean, Z_1 is one example of it. So you're right. You could say that way.

OK, so we talked about a number of simple examples of martingales last time. All of these examples assume that the expected value of the magnitude of $Z_{sub\ n}$ is less

than infinity for all n . Remember, this does not mean that the expected value of $Z_{\text{sub } n}$ is bounded over all n .

I mean, you can have expected value of Z to the n can be 2 to the n . It can be shooting off to infinity very, very fast. But it still is finite for every value of n . That's what this assumption is. Later on, when we talk about the martingale convergence theorem, we'll assume that these expectations are bounded, which is a much, much stronger constraint.

OK, we talked about these examples last time. All of them are pretty important because any time you're trying to do a problem with martingales, trying to prove something for martingales, what I like to do first is to look at all the simple examples I know and try to get some insight from them as to whether the result is true or whether it's not true. And if you can't see from the examples what's going on, then you're sort of stuck for the most part, unless you're lucky enough to construct the kind of proof you would find in a math book.

Now math book theorem proofs are beautiful because they're the-- I mean, mathematicians work to get the shortest possible proof they can, which has no holes in it at all. And that makes it very elegant.

But when you're trying to understand it, what is often done is somebody starts out understanding a theorem, and they write a proof which runs for three pages. And then they think about it for two weeks. They cut it down to one page. They think about it for another two weeks. They cut it down to half a page.

And then they publish it. And all they do is publish the half page. Everybody is stuck. Nobody knows where this came from. So what I'm trying to do here is, at least in some cases, to give you a little more than the half page to give you some idea of where these things come from, the extent I can.

OK so the zero mean random walk-- if $Z_{\text{sub } n}$ is equal to some of X_i , where X_i are IID and zero mean, then this zero mean random walk is, in fact, a martingale, just satisfies the conditions. This one is probably the most important of all of the simple

examples because it says if Z sub n is a sum of random variables-- don't know what the random variables so are-- the condition on the random variables is the expected value of X sub i , given all the previous ones is 0.

This is a general example because every martingale in the world you can look at the increments of that martingale, namely you can define X sub n to be Z sub n minus Z sub n minus 1. And as soon as you do that, X sub n satisfies this condition here. And what you've got for sure is a martingale, so that this condition here, that the X sub i are satisfying, is really the same as a martingale condition.

It's just that when people are talking about martingales, they're talking about the sums of random variables. Here we're just talking about the random variables themselves. It's like when we talk about sum of IID random variables, we prove the laws of large numbers and everything. What we're really talking about there is IID random variables.

What we're really talking about here is these random variables, which has the property that no matter what's happened in the past, the expected value of this new random variable is equal to 0. People call these fair games. And they call martingales examples of fair games.

And martingale, the expected value of Z sub n , you can think of it as you're expecting that worth of time n . And with these underlying random variables, the X sub i , the X sub i is, in a sense, your profit at time n .

And what this says is your profit at time n is independent of everything in the past, independent of every sample value of everything in the past. And this is why people call it fair game. It's really a very strong definition of a fair game. I mean, it's saying an awful lot.

I mean, everybody says life is not fair. Surely by this definition, life is not even close to fair because when you look at all your past-- I mean, you try to learn from your past. This is saying in these kinds of gambling games, you can't learn from the past. You can't do anything with it so long as you're interested only in the expectation.

The expectation of X_{i-1} is equal to 0, no matter what all the earlier random variables are. This, I think, gives you a sense of what a martingale is, probably better than the original definition.

OK, another one we talked about last time, this is very specific. Suppose that X_{i-1} is equal to the product of two random variables, U_{i-1} times Y_{i-1} . The U_{i-1} here are IID equiprobable, plus or minus 1. And the Y_{i-1} are independent of the U_{i-1} s. They can be anything at all.

And when you take these Y_{i-1} s, which are anything at all, but these quantities here, which are IID 1 and minus 1, when you look at this product here, any positive number this can be is equiprobable with the corresponding negative number. And that means when you take the expectation of X_{i-1} , given any old thing in the past, this U_{i-1} is enough to make the expectation equal to 0. So this is a fairly strong kind of example also, which gives you a sense of what these things are.

Product form martingales-- you use product form martingales primarily to find counter examples of theorems. If you stated a theorem and it isn't true-- almost all the examples I know of of reasonable martingale theorems which are not true, you look at a product martingale, and very often you look at this product martingale down here, and you find out either that the theorem is not true, which lets you stop looking at it. Or it says, well, it still isn't clear from that.

OK, so the product form martingale, there's a sequence of IID unit-mean random variables. And Z_n , which is the product, is then a martingale, if you assume this condition up here, of course. And this condition here, the probability that the n -th order product, namely the product of n of these sample values, if you get one 0, the product is 0. So you're done.

So the only question is, do you get all 1s? Or do you get something other than all 1s? If you get all 1s, then the product of these random variables is 2^n . You get 2^n with probability 2^{-n} . And you get zero with all the rest of the probability. The limit as n goes to infinity if Z_n is equal to zero with probability 1, namely eventually you go down to 0. And you stay there forever after.

And with this very small probability, you get to some humongous number. And you keep going up until eventually you lose, and you go down to 0. So the limit for the expected value of Z_n for every n the expected value of Z_n is equal to 1.

That's what makes this example interesting. The limit of the expected value of the Z_n s is equal to 1. But the Z_n s themselves go to 0 with probability 1. And the reason for that is that you had this enormous growth here with very small probability. But it's enough to keep the expectation equal to 1.

OK, then we started to talk about sub and super martingales. And I told you if you can't remember what the definition of a submartingale is in terms of is it less than or equal or greater than or equal, just remember that it's not what it should be. It's the opposite of what it should be. So the expected value of Z_n given the past is greater than or equal to Z_{n-1} . That means it's a submartingale.

In other words, submartingales grow in time. Supermartingales shrink in time. And that's strange. But that's the way it is.

If this quantity is a submartingale, then minus Z_n is a supermartingale and vice versa. So I'm going to only talk about submartingales from now on because supermartingales just do everything the same, but just look at minus signs instead of plus signs. So why bother yourself with one thing extra to think about?

So for submartingales, the expected value of Z_n given the past, given part of the past from i down to 1, is greater than or equal to Z_i . That's essentially the same as that theorem we've stated before, which said that for martingales, the expected values of Z_n given Z_i down to Z_1 was equal to Z_i . You remember we proved that in detail last time because that was a crucially important theorem.

You take that proof and you put this inequality in it instead of a quality, and it immediately gives you this. You just follow that proof step by step, putting an equality in in place of a quality. And same thing here, the expected value of Z_n is greater than or equal to the expected value of Z_i . That's true for all i .

So the expected value of Z_n is also greater than or equal to the expected value Z_1 . In other words, the expected values of these random variables, in fact, always grow. Or if they don't grow, they at least stay the same. They can't shrink.

OK, we started to talk about convex functions last time. I want to remind you what they are. A function which carries the real numbers into the real numbers is convex, if each tangent of the curve lies on or below the curve. Here's a picture of it. Here's a function h of x , a one-dimensional function. So you can draw x on the line. And h of x is something which goes up and down.

And here's another example. H of x is equal to the magnitude of x . You're usually used to thinking of convex functions in terms of functions that have a positive second derivative. Taking the geometric view, you get something considerably more general because it includes all of these cases, as well as all of these cases. And this idea of tangents lying on or below the line gives you the linkage here. You have something which comes down, goes to 0, and then goes up again.

Think of drawing tangents to this curve. Tangents have to have a slope starting here and going around to here. And all of those tangents hit at that point there. So this is a very pathological thing. But all the tangents indeed do lie below the curve. So you've satisfied the condition.

The lemma is Jensen's inequality. It says, if h is convex and Z is a random variable with finite expectation, then h of the expected value of Z is less than or equal to the expected value of h of Z . This seems sort of obvious, perhaps. It's one of those things which either seems obvious or it doesn't seem obvious. And if it doesn't seem obvious, it doesn't become obvious terribly easily.

But what I want to do here is to convince you of why it's true by looking at a little triangle here. I can think of the random variable Z as having three possible values-- one over here where we get this point comes into the curve, one here, and one here. Now if I take those three possible values of x and I think of assigning all possible probability assignments to those three possible values, what happens?

If I assign all the probability over here, I get that point. If I assign all the probability here, I get that point. If I assign all the probability here, I get that point. And for everything else, it lies inside that triangle. And you can convince yourselves of that pretty easily.

So for all of those probability measures where the expected value lies on this line, what you get is something between this and this as the expected value of h of Z . So you get something above the line for the expected value for h . Of expected values of Z , you just get this point right there, which is clearly smaller than anything you can generate out of that triangle or quadrilateral or any kind of straight line figure that you draw, which is the set of expected values you can get from probabilities using the points on that-- well, it's what I said it was.

OK, Jensen's inequality leads to the following theorem. And I'm not going to prove it here in class. It's one of those theorems which is sort of obvious and not quite. So if you want to see the proof, you can look at it. Or if you want to, you can just believe it.

If Z sub n is a martingale or it's a submartingale and if h is convex, then the expected value of the magnitude of h of Z_n is less than infinity for all n . Then h of Z_n is a submartingale. In other words, when you have that convex function, you go from something which is a martingale, which will be what you would get on the line to something which is bigger than that, so that what you get is the fact that expected value of h of Z is, in fact, growing with time, faster than the martingale itself is growing.

OK, so one example of this is if z of n is a martingale, then the absolute value of Z sub n is a submartingale. Submartingales are martingales. Martingales are submartingales also. So I don't have to keep saying that if it's a martingale or a submartingale. I can just say if it's a submartingale.

This theorem is usually stated as, if Z of n is a martingale, then h of Z_n is a submartingale, which is true. But just as obviously, if Z_n is a submartingale, which is more general, h of Z_n is also a submartingale. So you don't get out of the realm of

submartingales by taking convex functions.

And you also get that Z^2 is a martingale. And E to the rZ^n is a martingale because all of those are convex functions. So when you want to look at any of those, you just go from talking about a martingale to talking about a submartingale. And life is easy again.

OK, major topic-- stopped martingales. We've talked about stopping rules before. And a stopping rule, we were interested in stopping rules when we were mostly talking about renewal processes. But stopping rules can be applied to any sequence of random variables at all. What a stopping rule is, you remember, is you have a sequence of random variables, any kind of random variable. And a stopping rule is a rule where the time that you stop is determined by the things that you've seen up until the time that you stop.

You can't peak at future values. You have to look at these sample values one by one as they arrive. And a stopping rule is something which, when you get to the point you want to stop, you know that you want to stop there from the sample values you've already observed.

So when you're playing poker with somebody, which I think we talked about before, and you make a bet and you lose, you cannot withdraw your bet. You cannot say, I stopped! I stopped before! The time that you stop depends on what you've already seen up until the time that you stop.

We talked about possibly defective random variables before. I realized I never defined a possibly defective random variable. In fact, somebody asked me afterwards if it could be just any old thing at all. And I said, no.

And here's what it is. It's a mapping from the sample space to a set of real values, to the extended real values. And it has the property that for a defective random variable, the mapping can give you plus infinity. Or it can give you minus infinity. And it can give you either one of those with positive probability rather than just 0 probability.

So it applies to these cases where you have a threshold, a single threshold, for a random walk. And you might cross the threshold, or you might never cross the threshold. So it applies to conditions where sometimes you stop and sometimes you just keep on going forever. So it's nice for that kind of situation.

The other proviso a random variable back when we defined random variables still hold for a possibly defective random variable. So you have a distribution function. It's just the distribution function doesn't necessarily go to 1, and it doesn't necessarily start at 0. It could be somewhere in between.

OK, so a stop process for a possibly defective stopping time satisfies Z_{n^*} , which is Z_{n^*} is the stopping time. Let me start on that.

We have now defined stopping rules. We now want to define a stop process. A stop process is a process which runs along until you decide you're going to stop. But before when we stopped, the game was over and nothing else happened.

Here, the idea is to game continues forever. But you stop playing, OK? So the sequence of random variables continues forever. But the random variable of interest to you is this quantity Z_{n^*} , which at the point you stopped, then all subsequent Z_n 's are just equal to that stopped value.

So you're talking about some kind of gambling game now, perhaps, where you play for a while. And you're playing some game where the game continues forever. And you make your bets according to some strange algorithm.

And when you've made \$10, you say, that's all I want to make. I'm happy with that. And I'm not going to become the kind of gambler who can never stop. So I'm going to stop at that point.

So your capital remains \$10 forever after, although the game keeps on going. And if you start out with a capital of \$10 and you lose it all and you can't borrow anything, then you stop also when your capital becomes minus \$10.

So you can see that this is a useful thing when you're talking about threshold

crossings because when you have a random walk and you have a threshold crossing, you can stop at that point. And then, you just stay there forever after. But if you cross the other threshold, you stay there forever after.

And that makes it very convenient because you can look at what happens out at infinity as a way of saying what the value of a game was at the time you stopped. So you don't have to worry about what was the value of the game at the stopping point. You can keep on going forever. And you can talk about what the stopped process is.

And my guess-- or if you've read ahead a little bit, which I hope you have-- you will know that the main theorem here is that if you start out with a Martingale and you stop it someplace, you still have a Martingale. In other words, if you add stopping as one of your options in gambling and it's a fair game, if you can find a fair game any place, and you stop, then it's still a fair game. The stop process is still a fair game.

And that's as it should be because if it's a fair game, you should be able to stop. So for example, a given gambling strategy is Z_n is the net worth at time n . You can modify that to stop when Z_n reaches some given value.

So the stopped process remains at that value forever. And Z_n follows the original strategy. Here's the main theorem here.

If τ is a possibly defective stopping rule for a Martingale or a sub-Martingale and Z_n greater than or equal to 1, then the stop process, Z_n^* , is a Martingale if the original process is a Martingale and it's a sub-Martingale if the original process is a sub-Martingale.

And the proof is the following. You can almost say this looks obvious. If it looks obvious to you, you should admire your intuition. If it doesn't look obvious to you, you should admire your mathematical insight. And either way, the kind of intuition is that before stopping occurs, Z_n^* is equal to Z_n .

And after you stop, Z_n^* is constant. So it satisfies a Martingale condition because it's not going up and it's not going down. But in fact, when you try to think

through the whole thing, it's not quite enough. It's the kind of thing where after you look at it for a while, you say, yes, it has to be true. But why? And you can't explain why it's true.

I'm going to go through the proof here. And mostly the reason is that the proof I have in the notes, I can't understand it anymore. Well, I can sort of understand it when I correct a few errors in it. But I think this proof gives you a much better idea why it's true. And I think you can follow it in real time. Whereas that proof, I couldn't follow it in real time, or fairly extended time.

OK, so this stop process, I can express it in the following way. And let me try to explain why this is. If your stopping rule tells you that you stop at time m for a particular sample sequence, this indicator function, i of sub j equals n , this function here is 1 for all sample sequences for which you stop at time m . And it's 0 for all other sequences.

So the value of this stop process at time n is going to be the value at which it stopped, which is Z_m , when you have this indicator function, which is j equals m . And if it hasn't stopped yet, it's going to be Z sub n , which is what it really is. And it hasn't stopped. So the stop process hasn't yet stopped. So Z sub n star is equal to Z sub n .

OK, so as far as the magnitude is concerned, the magnitude of Z sub n is going to be less than or equal to the sum of those magnitudes. And the sum of those magnitudes, you can just ignore the indicator functions because they're either 0 or 1. So we can upper bound them by 1. So Z sub n star is less than or equal to the sum over m less than n of Z sub n plus Z sub n .

And this means that the expected value of Z sub n star has to be less than infinity because what it is in this bound here is a sum of finite numbers. When you take a finite sum-- this is a finite sum. There are only n plus 1 terms in it. When you take a finite sum of finite numbers, you get something finite. So expected value of Z sub n is--.

Excuse me. I was a little bit quick about that. The expected value of Z_{n^*} is now less than or equal to the expected value of each of the Z_{n^*} 's plus the expected value of Z_n .

Since the Z process is a Martingale, you know that all of those expected values are finite. And since all of those expected values are finite, the expected value of Z_{n^*} is finite as we said.

OK, so let's try to trace out what happens if we look at the expected value of Z_{n^*} conditional on the past history up until time $n-1$. We'll rewrite Z_{n^*} in terms of this expression here. So it's the sum over $m < n$ of the expected value of the stopping point if the stopping point was equal to m plus the expected value of Z_n if the stopping point was greater than n .

So we just want to analyze all of those terms. So we look at them. There's nothing complicated about it.

The expected value of this term here, expected value of Z_m times i of j equals m given Z_{n-1} . And now, we're going to be child-like about it. And we're going to assume a particular sample sequence, which is equal to little z_{n-1} . What is this expected value here?

It has to be Z_n if j is equal to m . Why is that? That's the argument I was just going through before. What's the expected value of a random variable conditional on the random variable, that same random variable, having a particular value?

The fact that you're given a large number of these quantities doesn't make any difference. The main thing that you're given here is the value of Z_n being little z_{n-1} , which says this quantity is equal to little z_m if j is equal to m . That's equal to 0 if j is unequal to m , which in fact is just equal to Z_m times the indicator function of j .

OK, so you-- and the same thing happens for the indicator function of j equals n . This should be j greater than or equal to n . This is Z_{n-1} .

And we add these things up. And what we get, finally, is this sum here. And now, you look at this last term here, which is a combination of here and here. And this is just the indicator function for j greater than or equal to n minus 1.

And if you look back at the definition of Z sub n star, this is just Z star of n minus 1. I see a lot of blank faces. But this is the kind of thing you almost have to look at twice. So we'll let it go with that.

So this shows that the expected value of Z sub n star given the past of the original process is equal to Z star n minus 1. That's not quite what you want.

You want the expected value of Z sub n star given Z star of n minus 1 to be equal to Z star n minus 1. In other words, you want to be able to replace this quantity in here with Z star n minus 1. And the question is, how do you do that? And that's what bothered me about the proof in the notes because it didn't even talk about that.

So the argument is Z star n minus 1 in the past is really a function of the past of the original process. If I give you the sample values of the original process, you can tell me where the process stops. And you can say what the stop process is both before and after that point.

So the stop values are a function of the unstopped values. So now what I can do is for every sample point of the original process leading to a given sequence of the stop process, we're going to have expected value Z sub n star given these values here as equal to Z star n minus 1.

And since that's true for all of the values for which this leads to that, this is true also. So that proves it. I'm doing this primarily because I think you ought to be tortured with at least one proof in every lecture. And the other thing is the proof in the notes was not quite sufficient. So I wanted to add to it here. So now, you have a proof of it.

Consequences of the theorem, this is for sub-Martingales, the marginal expected values of the stopped process lies in between the expected value of Z_1 and the expected value of Z sub n . In other words, when you take the stop process, it in some sense is intermediate between what happens at time 1 and what happens for

the original process at time n . It can't grow any faster than the original process.

This is also almost intuitively obvious. And it's proven in section 7.8. So you can find it there. It's quite a bit easier than the proof I just went through. The proof I just went through was a fairly difficult and fairly tricky proof.

Partly I went through that proof because everything we do from now on-- we're not going to do an awful lot of things. But the Martingale convergence theorem, the strong law of large numbers, and all of the other results we're talking about all depend critically on that theorem that we just went through. In other words, it's a really major theorem. It's not trivial little thing.

OK, this one is fairly major, too. But it follows very easily from the other one. Do I want to talk about this, this generating function product of the Martingale? No. Let's let that go. Let's not-- not that important.

So this is that, too. No, I guess I better go back to that. We need it.

OK, let's look at the generating function product Martingale that we had for a random walk. So X_n is a sequence of IID random variables. The partial sums form the variables of a random walk. S_n is a random walk where S_n is a sum.

For any r such that $\gamma(r)$ exists, we then define Z_n to be a Martingale, this Martingale here. That's called a generating function Martingale. Z_n is a Martingale. The expected value of Z_n is equal to 1.

You can see immediately from this that the expected value of Z_n is equal to 1. You don't need any of the theory we've gone through because the expected value of $e^{r S_n}$ is what? It's a moment generating function to the n -th power. That's just this term here. So this has to be equal to 1 for all n . The fact that this is a Martingale comes from that example of product form Martingale that we went through. So there's nothing very sophisticated here.

OK, so if we assume that $\gamma(r)$ exists and we let Z_n be this Martingale, well, this is just what we said before. So you see it here. Let j be the non-defective

stopping time that stops when either α greater than 0 or β less than 0. Since this is a stopping time, the expected value of e to the Z_n star is equal to 1 for all n greater than or equal to 1.

And the limit as n goes to infinity of Z_{n^*} is then going to be equal to the process at the time where you stopped. After you stop, you stay the same. So you never move.

And the expected value of Z_j is just this quantity here. What does that look like? That's the Wald identity coming up again. That's the Wald identity coming up for a random walk with two thresholds.

The nice thing about doing it this way is you can see that the proof applies to many other situations. You can have almost any stopping rule you want. And you still get the Wald identity. So it has a much more general form than we had before.

This business here, the limit of Z_{n^*} going to Z_j is a little fishy. The proof in the notes is fine. This limit does in fact equal this limit.

What's bizarre is that the expected value of this limit is not necessarily equal to the expected value of Z_j . So make a note to yourselves that if you ever want to use Wald's identity in this more general case, think carefully about what's going on as you go to the limit. Because it can be a little bit tricky there. It's not always what it looks like.

OK, so we're on to Mr. Kolmogorov again. Kolmogorov was the guy who did so many things in the subject. Most important, he said for a firm foundation to start with, he was the one that said you really need a model. And you really need some axioms. And then, he went on with all these other neat things that we've talked about from time to time.

His sub-Martingale inequality is a fairly simple result. But this follows after that stopping theorem that we've just talked about. And everything else depends on this.

So there's a sort of chain that runs through this whole development. And if you go

further in Martingales, you find that this is just an absolutely major theorem which comes up all the time. And it's wonderful because it's so simple.

OK, so let's let Z_n be a non-negative sub-Martingale. And then, for any positive integer m and any number a bigger than 0, the probability that the maximum of the first n terms of this Martingale is greater than or equal to the quantity a is the expected value of Z_n divided by a .

This looks like the Markov inequality. If instead of taking the maximum from 1 to m we just look at Z_n , the probability that Z_n is greater than or equal to a , then we get less than or equal to Z_n divided by a .

So what this is saying is it's really strengthening that Markov inequality and saying, you don't have to restrict yourself to Z_n . You can instead look at all of the terms up until m . And this bound here you get in the Markov inequality really covers the maximum of all of those terms to start out with, which says that for any m , you can look at the maximum over an enormous sum of terms if you want to. And it does this nice thing.

OK, I'm going to prove this also. But this proof is simple. So you can follow it in real time, I think.

So we want to start out with letting j be the stopping time, which is essentially the smallest term where you've crossed the threshold at a . And if you haven't crossed a threshold at a , then it's equal to the last term.

So here's the specific stopping rule that we're going to use. If Z_n is greater than or equal to a for any n , then j is the smallest n for which Z_n is greater than a . It's the first time at which we've crossed that threshold. If Z_n is less than a for all n up until m , then we make j equal to n . So we're insisting on stopping at some point.

This is not a defective stopping rule. It's a real stopping rule because you've set the limit on how far you want to look. OK, so the process has to stop by time m .

The value of the process at the time you stop-- remember this thing we've called Z

Z_j , which is the value at the stopping times. Z_j is greater than or equal to a if you stopped before time m . And Z_n is-- well, we're saying that Z_j is greater than or equal to a if and only if Z_n is greater than or equal to a for some n less than or equal to m .

If you haven't crossed a threshold by time m , then Z_m is equal to Z_n . But it's not above a . So the stopping time is this largest possible value that we've got until the process stops by time n . Z_j greater than or equal to a if and only if we've crossed a threshold for some n less than or equal to m . So the probability that we've crossed the threshold from 1 to n to n is equal to the probability that Z_j is greater than or equal to a , which is less than or equal to the expected value of Z_j divided by a . Since the process must be stopped by time m , we have Z_j is equal to Z_m^* . And the stop process, f time m and Z_n , is less than or equal-- expected value of the stop process is less than or equal to the expected value of the original process. Why is that? That's that theorem we just proved somewhere. Yeah, this one here, OK? That's submartingale consequence.

OK. So that completes the proof. And it's 10:30 now. So the Kolmogorov submartingale inequality is really a strengthening of the Markov inequality. So you get this extra soup to nuts form of it. Chebyshev inequality can be strengthened in the same way. That's called a Kolmogorov inequality also. Kolmogorov just got in here before anybody else and he took those axioms that he made up-- and he was a smart guy-- and he developed this whole school of probability in Russia. And along with that, since he had these original results, he just almost wiped up the field before anybody else knew what was going on. Partly a consequence of the fact that mathematicians in most other parts of the world didn't believe that there was any good probability theory going on in Russia. So they weren't really conscious of this until he cleaned up the whole field. So if you want to be famous mathematician, you should move away from the US and go to Upper Turkestan or something. And you then clean up the whole field the same way that Kolmogorov did. OK.

So the strengthening of the Kolmogorov inequality. What the result says is let Z_n be a submartingale with the expected value of Z_n^2 less than infinity. Then

the probability that the maximum of these terms, up to n , is greater than or equal to b , is less than or equal to the expected value of Z_n squared divided by b squared. You'll notice that that's almost the same thing as the submartingale inequality.

This one says the probability that the maximum of Z_i is greater than or equal to b is less than or equal to the expected value of Z_n squared over b squared. If you can't prove this, go back and look at the proof of the Chebyshev inequality. The proof of this given the submartingale inequality is exactly the same as the proof of the Chebyshev inequality given the Markov inequality. You just go through the same steps and it's fairly simple.

OK. So that is a nice result. What happens if you apply this to a random walk? If you apply it to a random walk, what you do is replace this Z_n with the sum of random variables, and the random walk, minus the mean of those random variables. We have seen that a zero-mean random walk is a martingale. So what we're going to do next is to use that zero-mean random walk of the martingale and then what this says is the probability that the maximum, from 1, less than or equal to n , of S_n minus $n\bar{X}$. That's Z_n because we're subtracting off the mean. The probability of that is greater than or equal to b , and we just give b another name-- m times ϵ -- and it's less than or equal to the expected value of Z_n squared. What's the expected value of this quantity squared?

It's n times σ^2 . Because S_n is just the sum of n IID random variables, which have variance σ^2 . So you take the expected value of this quantity squared, and this is n times the variance of X . So this then becomes σ^2 times n divided by m^2 times ϵ^2 , and the n cancels out. This gives you the Chebyshev inequality with the extra feature, but it deals with the whole sum from 1 up to n .

Now, you look at this and you say, gee. Wouldn't it be absolutely wonderful if instead of going from 1 to n , this went from n to infinity? Because then you'd be

saying the maximum of these terms-- the maximum is less than or equal to something. And you'd have the strong law of large numbers all sitting there for you. And life was not that good, but almost as good, because we can now do the strong law of large numbers assuming only IID random variables with the variance.

So we're going to use that expression that we just did-- we're going to plug it into what we need for the strong law of large numbers. Again, I'm going to give you the idea of the proof of that. I wasn't going to do that, but I looked at the proof in the notes, and I had trouble understanding that too. You understand, I have a problem here. I write things two years ago, I look at them now. I have a bad memory, so I have trouble understanding them. So I recreate a new proof, which looks obvious to me now because I've done it right now, and in two years, it might look just as difficult. So if you look at this half proof here and you can't understand it, let me know, and I'll go back to the drawing board and work on something else.

OK. So the theorem says let X_{i} be a sequence of IID random variables with mean \bar{x} and standard deviation $\sigma < \infty$. So I'm trying to do the strong law of large numbers before we assume the fourth moment. Here, I'm only assuming a second moment. If you work really hard, you can do it with the first absolute moment.

OK. Let the S_{n} be the sum of n random variables, then for any ϵ -- oh, I don't need an ϵ there. I don't know where that ϵ came from. Just cross that out. It doesn't belong. The probability that the limit, as n goes to infinity, of S_{n}/n is equal to \bar{x} . The probability of that event-- event happens for a whole bunch of sample sequences. It doesn't happen for others. And this says that the probability of the class of infinite length sequences for which that happens is equal to 1. That's the statement of the strong law of large numbers that we had before. It says that the probability of the set of sequences for which the sample average approaches the main probability of that set of sequences is equal to 1.

OK. So the idea of the proof is going to be the following thing. And what I'm going to use is this Chebyshev inequality we've already done. But since Chebyshev

inequality, in this new form-- namely the Kolmogorove inequality only goes up to n , what I'm going to do is look at successively larger values of n . So I'm going to try to crawl my way up on infinity by taking first a short length, then a longer length, then a longer length, and a the longer length. So I'm going to take this quantity here, which was the quantity in the Kolmogorove submartingale inequality. I'm going to ask what's the probability that the union of all of these things, from m equals sum k , which I'm going to let go to infinity later, what's the probability of this union? And the terms in the union, instead of going from 1 to n to n , I want to replace n by 2 to the m . The maximum of this. The probability that this is greater than or equal to 2 to the m times epsilon-- that's the biggest term times epsilon. I want to see that this is less than or equal to this quantity. Now, why is this less than or equal to that?

AUDIENCE: Union bound.

PROFESSOR: What?

AUDIENCE: Union bound.

PROFESSOR: Union bound, yes. That's all it is. I've just applied the union bound to this. This is less than or equal to the probability of this for m equals k , plus the probability of this for m equals k plus 1, and so forth. Each of these terms is sigma squared over 2 to the m times epsilon squared. That's what we had on the last page, I hope. Yes. Sigma squared over m times epsilon squared. Remember, we replaced m by 2 to the m , so this has changed in that way.

And now we can sum this. And when we sum it, we just get 2 sigma squared over 2 to the k times epsilon squared. What are we doing here? What's the whole of this? The Kolmogorove submartingale inequality, lets us, instead of looking at just one value of n , let's us look at a whole bunch of values altogether and maximize over them. So what I'm going do is use the Kolmogorove submartingale inequality over one big bunch of things and then over another much bigger bunch of things, then over another much, much bigger set of things. And because I'm hopping over these much larger sequences, I can now sum this quantity here, which I couldn't do if I only had an m here. If I replaced this 2 to the m by m and I tried to sum this, what

would happen? It's a harmonic series and it diverges. So what I've been able to do-- or, really, what Kolmogorov was able to do-- was instead of summing over all m , he was summing over bunches of things and using this maximum here.

So this probability is less than or equal to something finite. If I now let k go to infinity, this term goes to 0, which says that the tail end of this whole big thing goes to 0 as k gets larger, for any epsilon at all. But now, this doesn't quite look like what I want it to look like, so what I'm going to do is find something which is smaller than this that looks like what I would like it to look like. So I'm going to lower bound this quantity here by this quantity here. I still have the same union here. Instead of finding the max over 1 to n to 2 to the n , I will get a probability which is smaller because I'll only maximize over part of those terms. I'll only go from 2 to the m minus 1 , less than or equal to n , less than or equal to 2 to the n . So I'm maximizing over a smaller set of terms, which makes the probability of this smaller.

And then I'm replacing the 2 the m here by 2 times n , because now this bound is between 2 to the n minus 1 and 2 to the m . So I can replace it that way. And now, it's exactly the sum that I want. Yeah?

AUDIENCE: What does it mean to maximize the [INAUDIBLE]? Now n is [INAUDIBLE]. So it looks like you're maximizing over inequalities. Is it something like that?

PROFESSOR: Well, Yeah no. I probably want to take this quantity, subtract off-- no. What I really have to do to make this make any sense is write it as the maximum over the same set of things S_n over n minus x bar, greater than or equal to 2 epsilon. OK? Thank you.

AUDIENCE: OK. Another thing-- you're more likely to be bigger than a smaller quantity. Your n is smaller than 2 to the m , you're more likely to be bigger than a smaller quantity. So you're not bounding correctly, it would seem. Oh, no. You're doing 2 times m .

PROFESSOR: Well, there's no problem here. I think the question is here. Can I reduce this maximum down to a smaller sum and get a smaller probability? Oh, yes. There's a smaller probability that this smaller max will exceed a limit than if this will exceed a

limit. So I should do it in two steps. In fact, you pointed it out here. I should do it in three steps. So the first step replaces this maximum with this maximum. Then the second step is going to go through that step. And the third one is going to replace the m here with the n . Anyway.

I think it's OK, with a few minor twiddles. And before the term is over, I will get a new set of notes out on the web, and you can check them to see if you're actually satisfied with it, OK?

OK. So finally, the martingale convergence theorem. I'm not even going to try to prove this at all, but you might have some imagination for how this follows from dealing with stop processes, also. What it says is Z sub n is a martingale again. We're going to assume that there's something finite m , so the expected value of Z_n is less than or equal to n . So what I'm saying is the expected value of Z sub n is now bounded-- it's not finite, it's more than finite-- it's bounded. It can never exceed this quantity. I can have an expected value Z sub n , which is equal to 2 to the n , and that's fine for every n . But it's not bounded. This quantity, I'm assuming it's bounded.

And then, according to this super theorem, there's a random variable, Z . And don't ask what Z is. Z is usually a very complicated random variable. All this is doing is saying it exists. You don't know what it is. Such that the limit, as n goes to infinity, is Z sub n , is equal to this random variable. In other words, the limit of Z sub n minus Z is equal to 0 as n goes to infinity.

And the texts proves the theorem with the additional constraint that the expected value of Z sub n squared is bounded. Either one of those bounds is a very big constraint on these martingales. So the way you use these theorems is you take an original problem that you're dealing with and you twist it around, then you massage it and you do all sorts of things to it all in order to get another martingale, which satisfies this bound in the constraint. Then you apply this theorem, and then you go back to where you started. So that's the sort of general way of dealing with this.

And you see this theorem being used in all sorts of strange places where you would

never expect it to be used. For those of you in the communication field, about a couple of years ago, there was a very famous paper dealing with something called polar codes, which is a new kind of coding-- very careful. And they guy, in order to prove that these codes worked, used that. I don't know how-- I haven't checked it out yet-- but that was crucial in his proofs. So he had to turn these things into martingales somehow and then use that proof.

OK. We talked about branching processes, about the remarkable things about them. This theorem applies directly to these branching processes. A branching process, you remember-- the number of elements or organisms or whatever, at time n is the number of offspring of the set of elements at time $n - 1$. Each element at each time has a random number of offspring, which is independent of time that's independent of all the other elements. And it's a random variable Y . And the expected value of X_{n+1} is going to be X_n times the expected value of Y , because X_n is the number of elements in the n -th generation. \bar{Y} is the expected number of offspring of each one of them.

So you look at it and you say, ah. That theorem doesn't work. No good. You walk away. Then somebody else who is really interested in branching processes says, oh, this process is growing as-- I mean it's growing by \bar{Y} every unit of time. So I should be able to deal with that somehow. So I say, OK. Let's look at the number of elements in the n -th generation and divide it by \bar{Y}^n . When you do this, what happens?

The expected value of X_n / \bar{Y}^n is going to be equal to X_{n-1} / \bar{Y}^{n-1} .

AUDIENCE: You need to [INAUDIBLE].

PROFESSOR: Yes. That would help, wouldn't it. Thank you. Given X_{n-1} / \bar{Y}^{n-1} . And X_{n-2} / \bar{Y}^{n-2} . And so if we're given these things, we don't have to worry about this quantity if we're just given a number in each generation. If we're given a number in generation $n - 1$, the expected value of X_n / \bar{Y}^n is just X_{n-1} / \bar{Y}^{n-1} . Expected value, we pick up another

value of Y divided by \bar{Y} to the n , which is X^{n-1} / \bar{Y}^{n-1} .

OK. So the theorem applies here because that expected value is just 1, then. And this is a martingale. So the theorem says that this quantity approaches a random variable. And what does that mean? Well, if you observe this process for a long time, it might die out. If it dies out and it stays died out-- it never comes back again. And if it doesn't die out, it's going to start to grow. And if it starts to grow, it's going to start to grow in this way. After a very long time, if it's growing, X^{n-1} is humongous, and the law of large numbers says that the next generation should have very close to X^{n-1} times the \bar{Y} elements in it. So it says that after you get started, this thing wobbles around trying to decide whether it's going to go to zero or decide to get large. But if one starts to get large, then it becomes very stable from that time on. And it's going to increase by \bar{Y} with each unit of time. If it decides it's going to die out, it goes to 0 and it stays there.

So what the theorem is saying is just what I just said-- namely X^n / \bar{Y}^n in fact either grows in this very regular \bar{Y} or it goes to 0 and it stays there. So this random variable is 0 with a probability that the process dies out, and we evaluated that before. The other values of it are very hard to evaluate. The other values depend on how long this thing takes. If it's not going to go to 0, how long does it take before it really takes off? And sometimes it takes a long time before it really takes off, sometimes it takes a short time, and that's what the random variable Z is. But the random variable Z says that after a very long time, the value of this process is X^n is going to be approximately Z times this quantity here, which is growing exponentially.

OK. That gives us the martingale convergence theorem. Next time, I will try to review at least the whole course from Markov chains on.