STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

Department of Computer Science and Automation

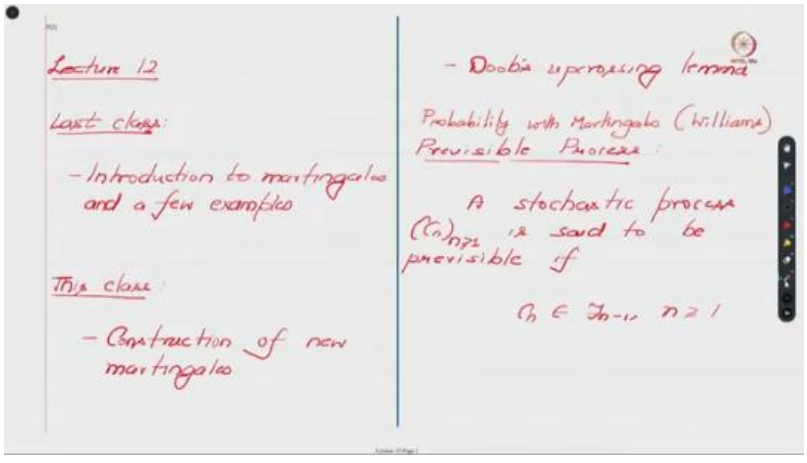
Indian Institute of Science

Lecture 12

Doob's Uncrossing Lemma

Hello and Namaste, everyone. Welcome to Lecture 12 of this NPTEL course on stochastic approximation. In the previous class, we looked at the definition of martingale sequences, martingale difference sequences, and also at some examples of martingale sequences. In this class, we will go a step beyond and look at some nice properties of martingale sequences and then make some initial or prove some initial results that will help us understand the conditions under which a martingale sequence can converge. All right.

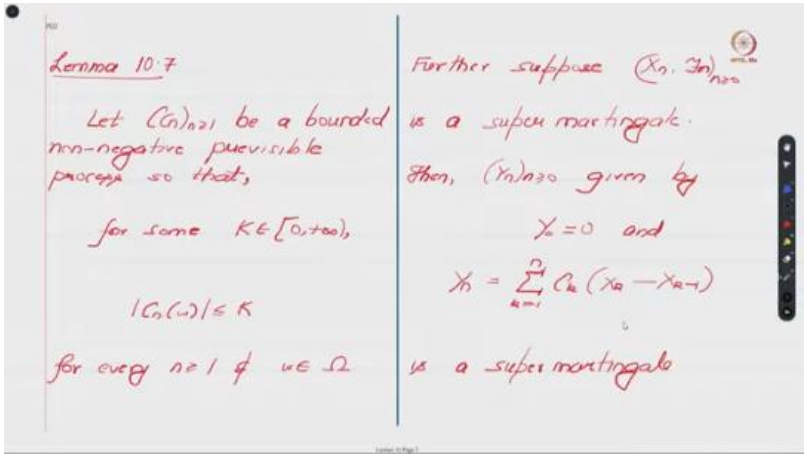
So, in particular, in this class, we will look at the construction of new martingale sequences from existing martingale sequences. And also, we will focus on what is called as the Dube's upcrossing lemma. For this, I mean, for this week's discussion, I will be mostly relying on this textbook called Probability with Martingales by David Williams. And whenever I refer to a particular result, I will actually refer to it from this textbook. So, before we sort of look at this construction of new martingales, we need to understand this special type of process which is called the previsible process.



So, stochastic process CN, where n is greater than or equal to 1. So, this stochastic process is said to be previsible with respect to a filtration Fn if the following property holds: that is, Cn is measurable with respect to Fn minus 1 for n greater than or equal to 1.

$$C_n \in \mathcal{F}_{n-1}, n \geq 1$$

So, recall that when we spoke about adapted processes, we required that this nth random variable be measurable with respect to Fn. Now, in a previsible stochastic process, we have that Cn is actually measurable with respect to Fn minus 1. So, now we will use such a previsible sequence to construct a martingale.



In fact, in this example, we will see how to construct a supermartingale, and because a martingale is a special case of a supermartingale, the result will also apply to supermartingales. So, in this result, we require that Cn be a bounded non-negative previsible process, right? Furthermore, we require that the absolute value of Cn(ω) be less than or equal to k. So, we require this bound to hold for every n and every ω in Ω, okay?

$$|C_n(\omega)| \leq K$$

Furthermore, suppose that the sequence (Xn, Fn) is a supermartingale. So again, this notation here means that Xn is adapted with respect to Fn and is a supermartingale in that, on average, this Xn sequence actually decreases.

Now, let us construct a new sequence using Xn and Cn with the following rule. The new sequence will be denoted by Yn, where Y0 equals 0, and for every n ≥ 1, Yn equals the

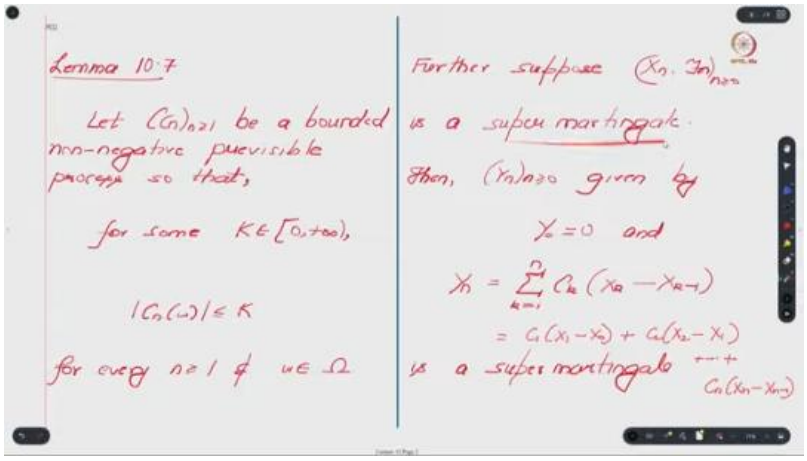
sum given here. So, you can see that here we look at the sum from k = 1 to n, and we have Ck times the difference Xk minus Xk−1, right? If you expand this, we would have this equals C1 times (X1 − X0) plus C2 times (X2 − X1) and so on up to Cn times (Xn − Xn−1).

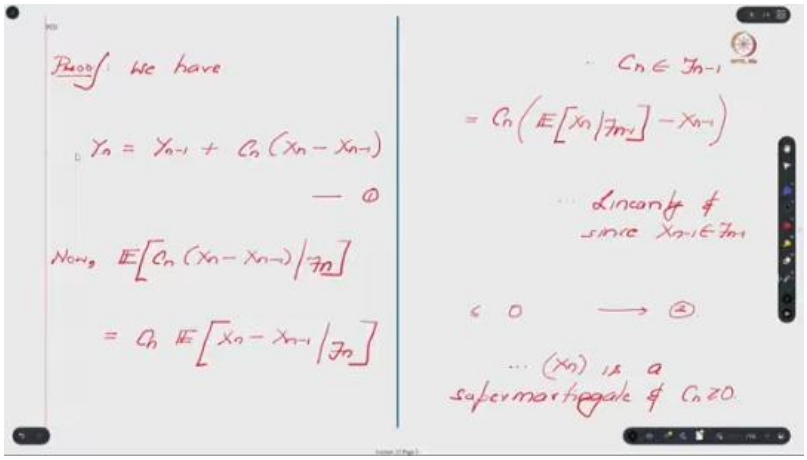
$$Y_0 = 0$$

$$Y_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1})$$

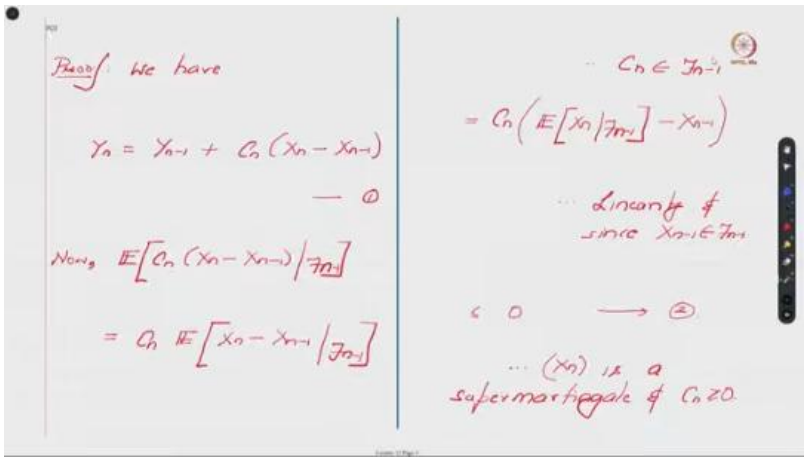
$$= C_1(X_1 - X_0) + C_2(X_2 - X_1) + \cdots + C_n(X_n - X_{n-1})$$

The claim is that this new sequence Yn, generated in the following way, is a supermartingale if Xn is a supermartingale. Alternatively, if Xn is a martingale, then the Yn sequence will also be a martingale. So, let us go over the proof of this result. From the definition of Yn, it is easy to see that Yn equals Yn−1 plus Cn times (Xn − Xn−1). That is, we have just separated out the last term in the sum, right?



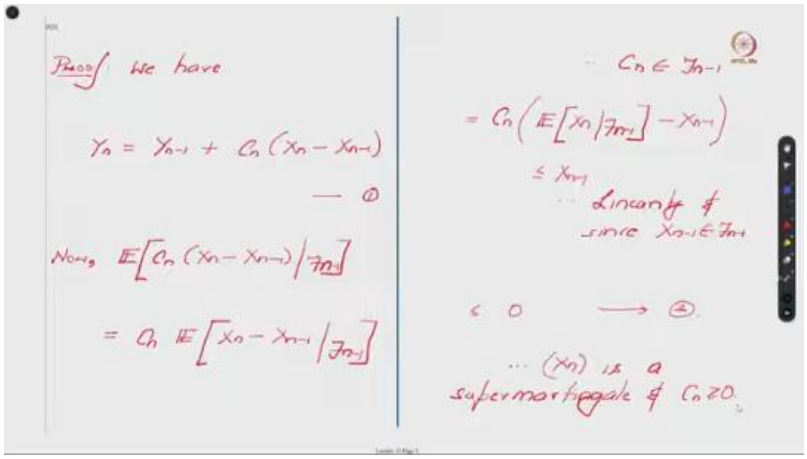


So, now, if you look at the conditional expectation of this additional term, right, we can say that the conditional expectation of this expression given Fn is where Fn is the given filtration, that is the filtration with respect to which your Xn sequence is a supermartingale. So, if you take the conditional expectation of this expression with respect to Fn, we can say that this conditional expectation equals this, and this holds because cn is pre-visible. Okay, so I think I made a mistake here; this should be Fn minus 1 here. Sorry about that. Okay, so because your cn is pre-visible, that implies cn is measurable with respect to Fn minus 1, and hence we can pull it outside this conditional expectation and write it as cn times the conditional expectation of the difference between Xn and Xn minus 1, right?



Now, by invoking the linearity of conditional expectation, we can write this as the conditional expectation of Xn given Fn minus 1 and the conditional expectation of Xn minus 1 given Fn minus 1. Now, since Xn is a martingale sequence, it is adapted. In particular, your Xn minus 1 is measurable with respect to Fn minus 1, and hence this

conditional expectation—that is, the conditional expectation of Xn minus 1 given Fn minus 1—becomes Xn minus 1 itself, right? And hence we would have this expression. Now, because your Xn is a supermartingale, this expression is upper bounded. by Xn minus 1, and if you take the difference between Xn minus 1 and Xn minus 1, we would end up with 0, and this inequality is the same as this inequality because this Cn is a non-negative process. So, this inequality holds because Xn is a supermartingale and Cn is greater than or equal to 0.



$$Y_n = Y_{n-1} + C_n(X_n - X_{n-1})$$

$$\mathbb{E}\left[C_n(X_n - X_{n-1})|_{\mathcal{F}_{n-1}}\right] = C_n\mathbb{E}\left[X_n - X_{n-1}|_{\mathcal{F}_{n-1}}\right]$$

$$C_n \in \mathcal{F}_{n-1}$$

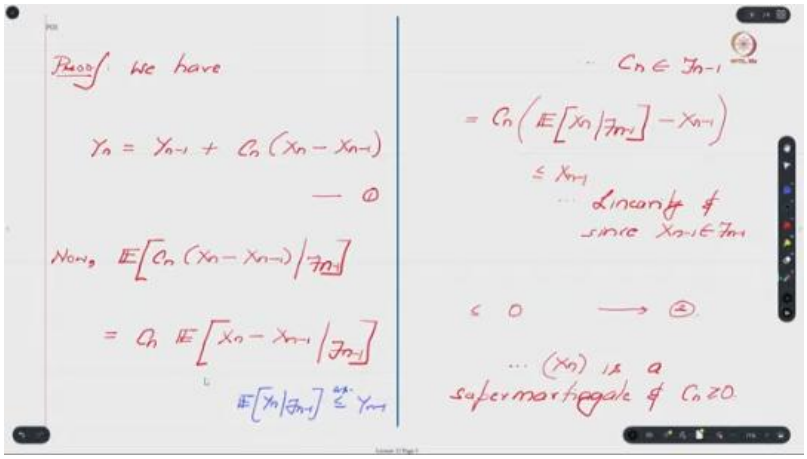
$$= C_n\left(\mathbb{E}\left[X_n|_{\mathcal{F}_{n-1}}\right] - X_{n-1}\right)$$

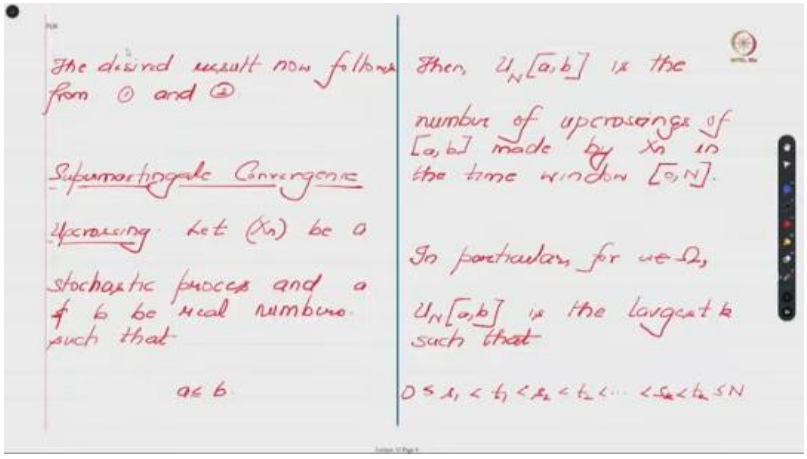
$$\leq X_{n-1}$$

$$\leq 0$$

Now, if you combine these two equations, one can conclude that the conditional expectation of Yn given Fn minus 1 is less than or equal to Yn minus 1. So, let me just officially write that down: the conditional expectation of Yn given Fn minus 1 is almost surely less than Yn minus 1. And since this relation holds for all n greater than or equal to 1, one can conclude that your Yn is actually a supermartingale, as desired. So, this was the

first result that we wanted to discuss, and in particular, you can see that by using a given martingale or a supermartingale and a pre-visible process, we can actually construct a new martingale or a corresponding supermartingale.



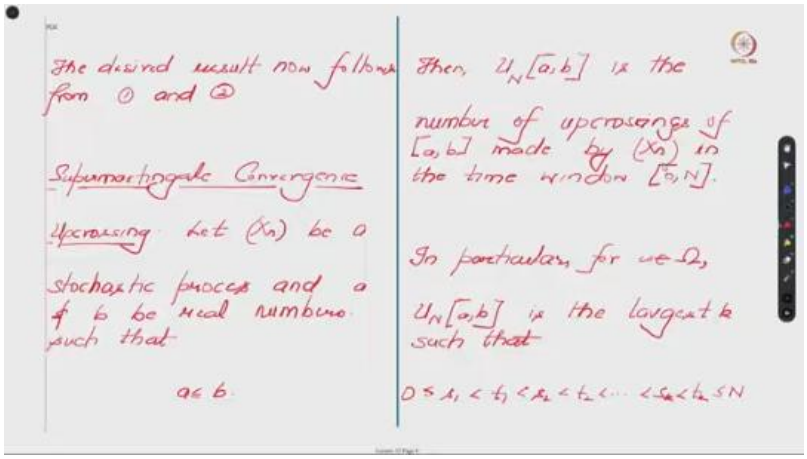


So, with that result in place, we can now talk about the supermartingale convergence result. And in particular, we need to first understand this concept called an upcrossing. This will be needed to make sense of the supermartingale convergence. So, what is an upcrossing? So, for an upcrossing, we first need a stochastic process.

To define upcrossings, we do not need martingale sequences or, you know, a stochastic process that has some special property. We can talk about upcrossings with respect to any stochastic process. So, let Xn be some arbitrary stochastic process, and let A and B be two real numbers with the property that A is less than or equal to B. And now we want to talk

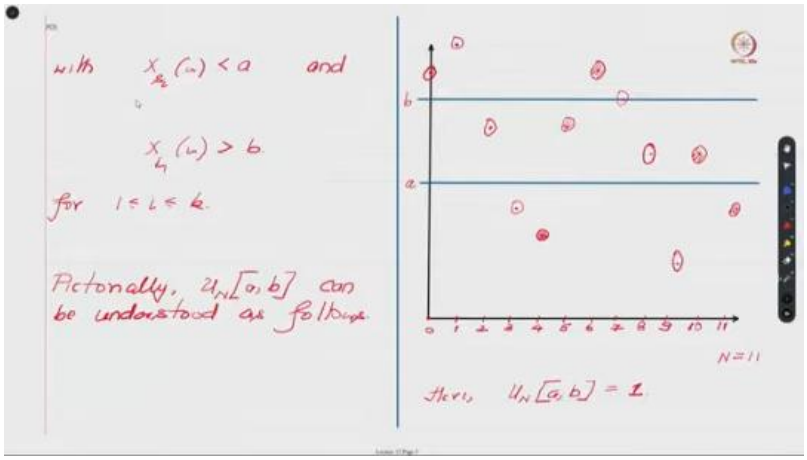
about the number of upcrossings of this interval [A, B]. So, the closed interval [A, B] made by your Xn process. So, let me again repeat this.

We have been given this stochastic process Xn. And we want to look at the number of upcrossings made by this process Xn. of the interval [A, B], right, in the time window 0 to n, and we will denote the number of upcrossings by Un(A, B). So, because this Xn is a stochastic process, it is not difficult to see that this Un(A, B) will actually be a random number. So, now let us look at multiple interpretations of what this Un(A, B) is.



In particular, we will try to get a good understanding of what this Un(A, B) is, and I also have a picture later on to illustrate what this Un(A, B) is. right so let's look at an algebraic description first for every sample point little Omega in your sample space capital Omega UN AB of Omega okay this is the largest K such that the following two properties hold the first property is that We should have s1, t1, s2, t2 all the way up till sk, tk satisfying the following inequality. That is your s1 should be greater than equal to 0. It should be strictly be less than t1.

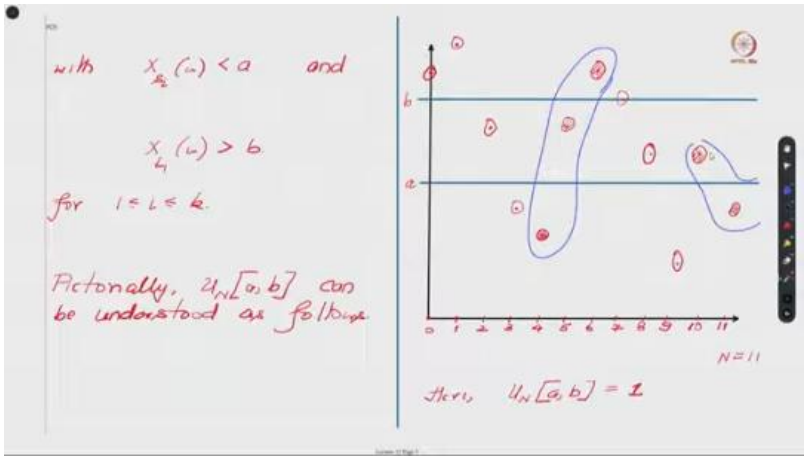
t1 should be strictly less than s2. s2 should be strictly be less than t2 and all the way up till sk being strictly less than tk and tk being less than equal to n where this n is the end point of your interval. So, UNAB is the largest K such that this sequence of inequalities hold. Furthermore, at every i, your XSI of omega, that is SI is some integer, right?



So, XSI is the value of the sequence at time SI. We require that this XSI be strictly less than a and XTI of omega be strictly bigger than b and we require this to hold for every i between 1 to k. So, in words the idea is that at time SI your process should be below A and at time instance TI the process should be above B right and by up crossing we mean that you know you start from below A and go all the way above B. So, this starting from below A and crossing B is what is referred to as an up crossing. So, now we will pictorially try to understand what is an up crossing. So, let us look at a pictorial description.

So, here is the picture that is given. So, on the X axis, we have time and the Y axis, we have values and you see these two horizontal lines where I have marked A and B, right? So, we only require that AB less than equal to B and I have identified some arbitrary places for this A and B, right? And you see these dots with circles, these are the values of your XN process, right? And you can see that you know at some places you have this circle which is I mean at certain time instances these circles are unshaded and at certain time instances these circles are shaded and I will tell you when it is shaded and when it is not shaded.

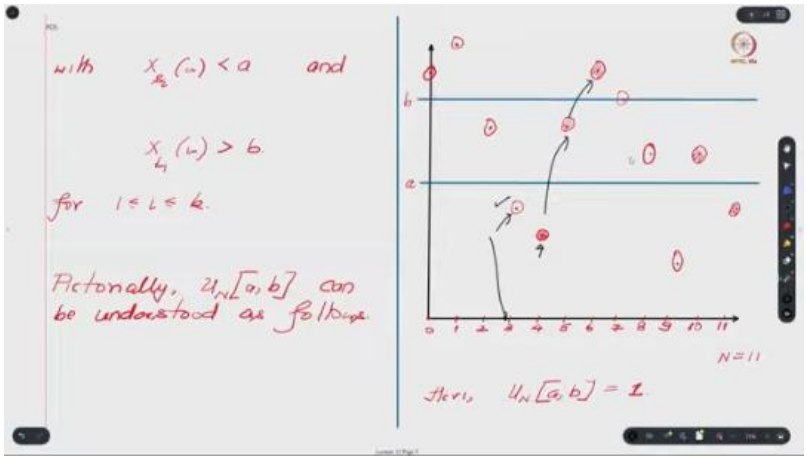
At a very high level the shading denotes an up crossing so you can see that from here to here You have one up crossing and at this point you started an up crossing but it did not finish. So the up crossing started and it did not finish. So this example over here is to illustrate that when I mean an up crossing or when I say up crossing I do not mean that something has to monotonically increase. We only require that it starts below A and then goes above B.

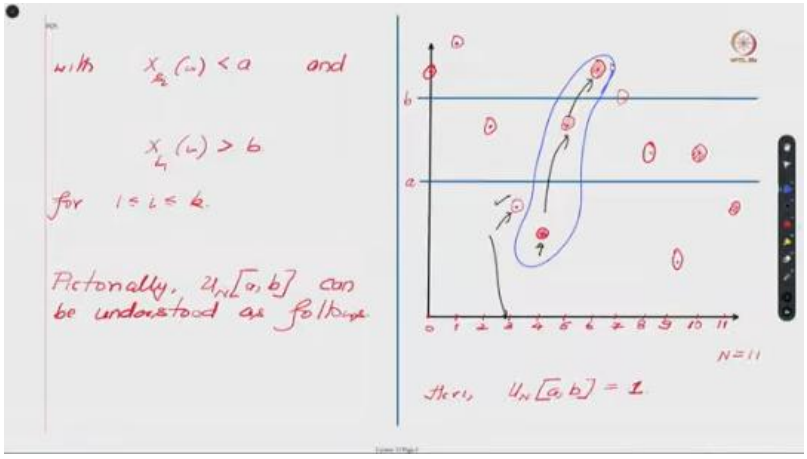


So, let us now try to understand this notion of up crossing. So, you can see that at this point, which is 3 over here, we went below A for the first time. We went below A for the first time and at time instance 4, we have sort of shaded this circle. This circle being shaded implies that an up crossing is in process. So, we go from here to here, the up crossing continues.

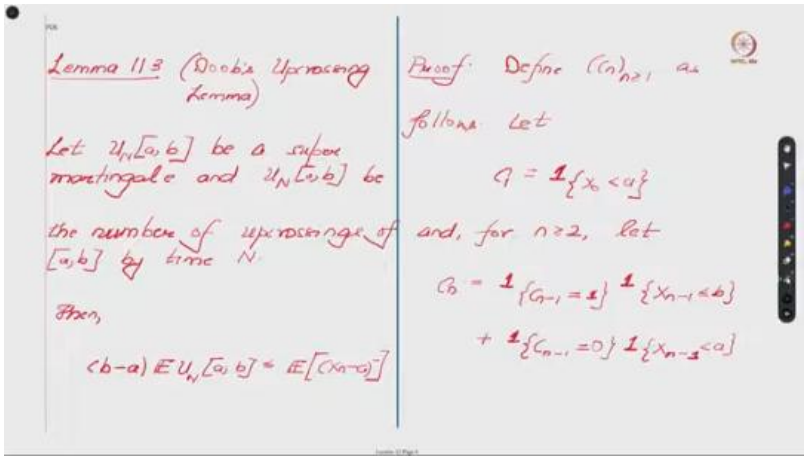
We go from here to here. Now, notice that at this point, which is time instance 7 or maybe 6, right? Yeah, time instance 6, right? We go above this value of b, right? And as soon as we go above the value of b, the upcrossing ends, and you can see that the next circle is no longer shaded.

So, this is not shaded again because we are not below A. The upcrossing means you have to start below A and go all the way to B. Since we have not gone below A yet, the second upcrossing has not started. So, at this point in time, you go below A. Again, the upcrossing starts. So, that is why the next circle is shaded. Now, this circle goes below because Xn can take a value below also, and this upcrossing has not ended. Hence, when you consider n equals 11, the number of upcrossings of AB is actually only one, which is determined by these three values over here. Is this okay?

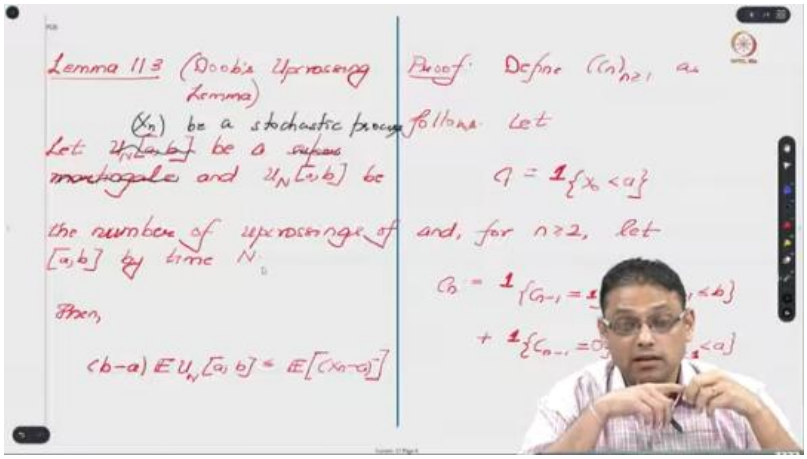




So, I hope the pictorial description of an upcrossing is clear. So now what we will do is we will try to formally denote a process which counts these upcrossings and somehow relate it to your supermartingale process. Is this okay? In fact, this result applies to any stochastic process, and we will use this inequality, the one that is given over here, later on for supermartingales. So, this result applies for any stochastic process, and this result is denoted as the Dube's upcrossing lemma, one of the fundamental results in martingale theory. This number over here corresponds to the number in this textbook called Probability with Martingales by David Williams.

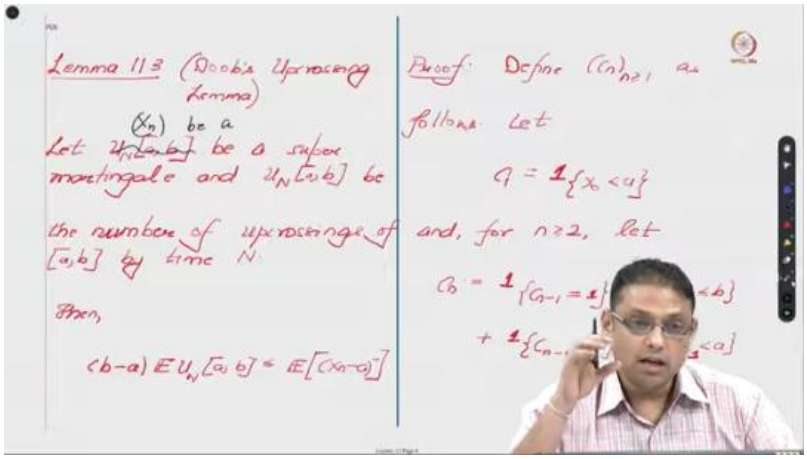


So, what does this lemma say? This lemma says that let u_n be a—there is a mistake over here—it should say let x_n be a stochastic process. Okay, and u_n be the number of upcrossings of the interval [a, b] by time n of this stochastic process x_n, right? So, this result requires that you have some stochastic process, and u_n(a, b) looks at the number of times this process starts below a and goes above b until time n. And the claim is that for this stochastic process, if you look at the expected value of upcrossings and multiply it by (b − a), then this expression will be upper bounded by the expected value of (x_n − a)^−. So, actually, I realize that for this result, we also require things to be a supermartingale.



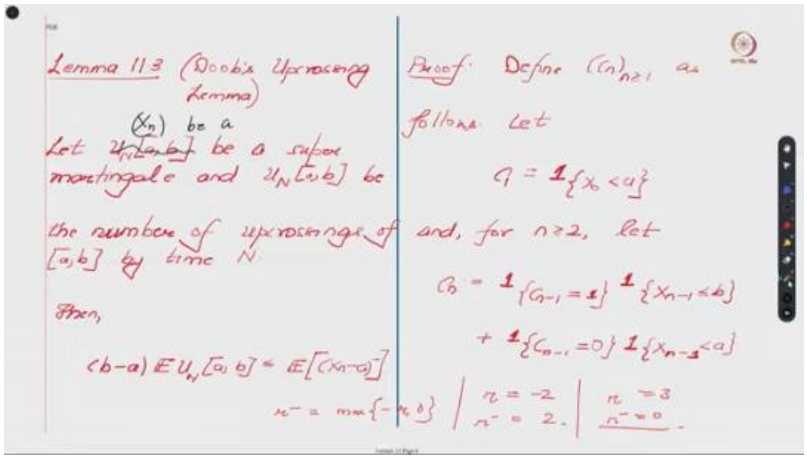
So, I take back what I said. You know, this is actually incorrect. We require that your x_n be a supermartingale. So, let me again state the statement of this result because I made a mistake. So, we require that x_n be a supermartingale and u_n be the number of upcrossings of this interval [a, b]. Then the claim is that (b − a) times the expected value

of u_n(a, b) is less than or equal to the expected value of (x_n − a)^−. So, let me first tell you what this notation means for any real number.



Let us say the real number is r. Then r^− is the max of (−r, 0), right? So, in other words, if you have r = −2, then r^− = 2. On the other hand, if r = 3, then r^− = 0. Is this okay?

So, in other words, whenever the argument to this superscript minus is negative, then the output is some positive number. On the other hand, if the argument to this superscript minus is positive, then the output is actually 0. Now, we extrapolate this definition to the random variable itself. So, Xn minus a right is some real number for different little omegas. So, you take that number now—either that number can be positive or negative—and depending on whether it is positive or negative, you define this minus superscript minus operation.



So, that will be another random variable, and you can take its expectation.

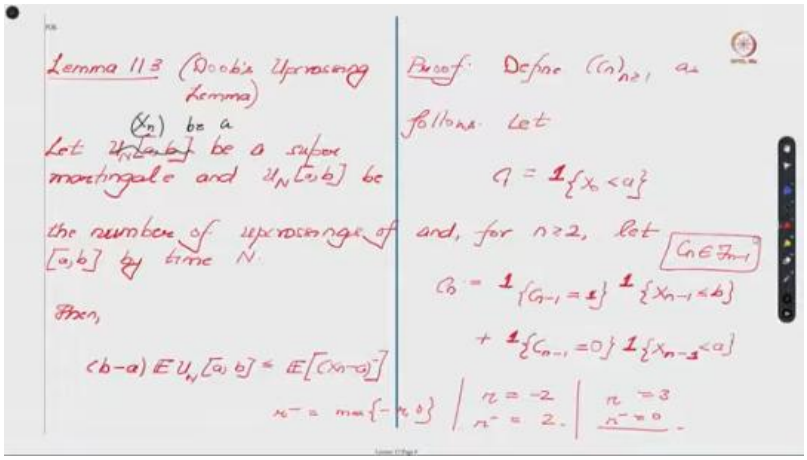
$$(b-a)\mathbb{E}\mathcal{U}_N[a,b] \leq \mathbb{E}[(X_n-0)^-]$$

So, this is the claim, and we will now prove this claim. Right, so towards that, what we will do is we will first define a previsible process. Right? So, what is this previsible process? The previsible process we will denote by CN, and the first element in this process will be C1, which is set to be the indicator that X0 is less than A. So, notice that C1, because it is an indicator function, can either take the value plus 1 or 0, and it will take the value plus 1 if X0—that is, you know, notice the index is actually 1 less than this index—right. So, if X0 is strictly less than A,

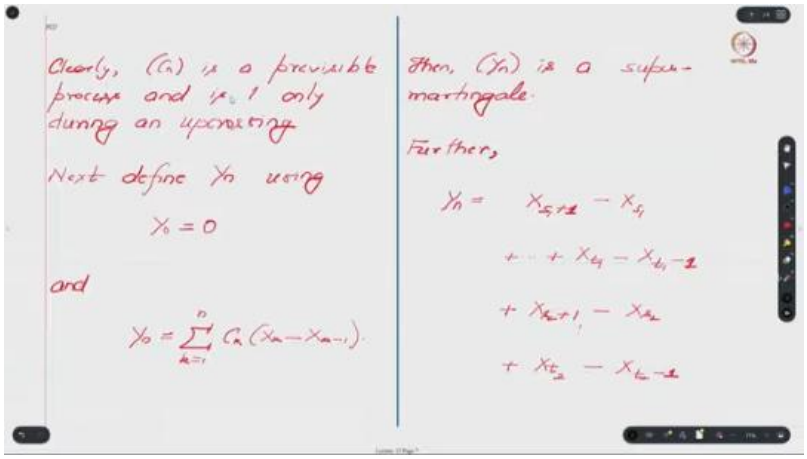
then C1 will be 1, indicating that your up-crossing has begun at time 0 itself. On the other hand, if X0 is above A, then, you know, your up-crossing has not begun, right? And then, this Cn is actually defined in an iterative fashion. So, Cn is either can be 1 if these two conditions simultaneously hold or it will be 1 if these two conditions simultaneously hold.

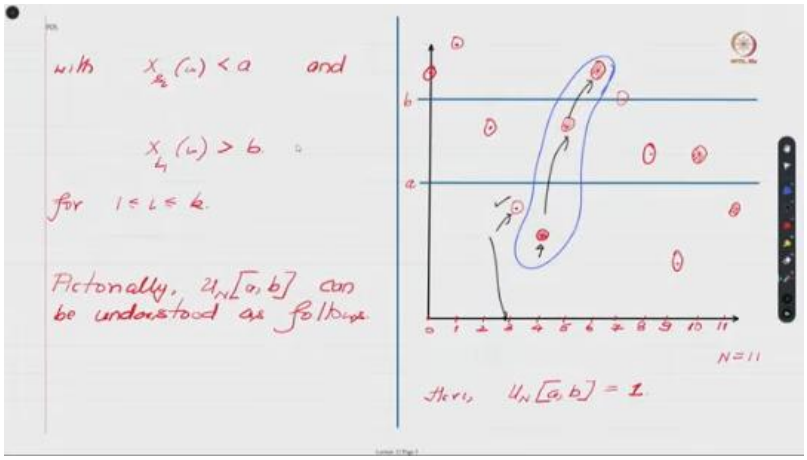
So, let us look at these two conditions. It says that an up-crossing was already going on, which is what Cn-1 equals 1 indicates, and the previous value was still below B. So, because this was 1 and the previous value was still below B, you know the up-crossing is still going on, and hence Cn should be 1. On the other hand, Cn minus 1 equals 0 implies that there was no ongoing up-crossing. And now, because your Xn minus 1 is strictly less than A, a new up-crossing has begun.

Is this okay? So, in either of these two scenarios, Cn will take the value 1, right? And one can easily check that because of, you know, these two things being complementary events. Either this product will be 1 or this product will be 1, and also notice that Cn is a function of only Cn minus 1 and Xn minus 1, and consequently, your Cn will be measurable with respect to Fn minus 1. In other words, your Cn process will be pre-visible.

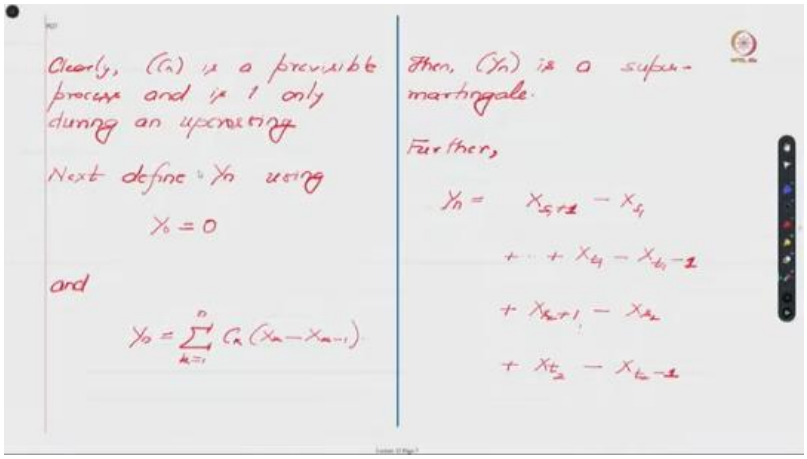


Right, and it is easy to see that your Cn will actually take the value 1 only during an up-crossing. Right? So now, if you go back to this pictorial representation that I had, you can see that the reason I had shaded some circles and left some circles unshaded was primarily those time instances where Cn took the value 1. So, you can look at the definition of Cn and conclude for yourself that C4 is 1 over here, C5 is 1 over here, C6 is 1 over here. However, C7 is 0, and so on and so forth. Right.





And the value of C4, for instance, only depends on what happens at this point. Similarly, the value of C5 only depends on what happens here, and so on and so forth. So that sort of illustrates what a previsible process is. And now, our goal is to prove this inequality. So, towards that, what we will do is we will first define this stochastic process Yn in the following way.

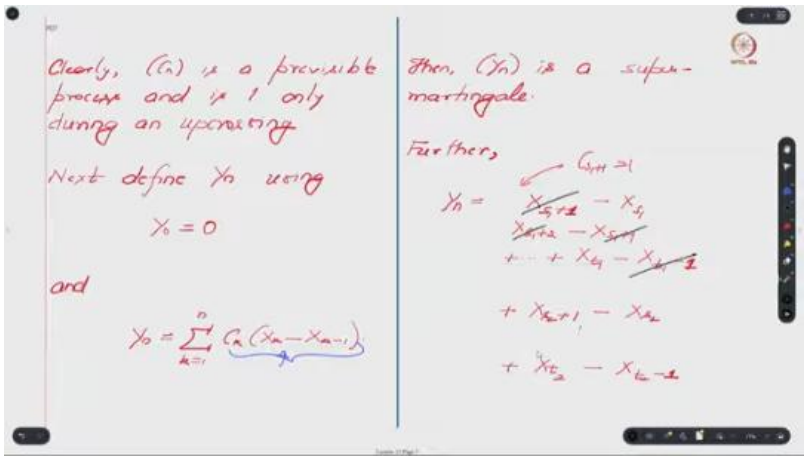


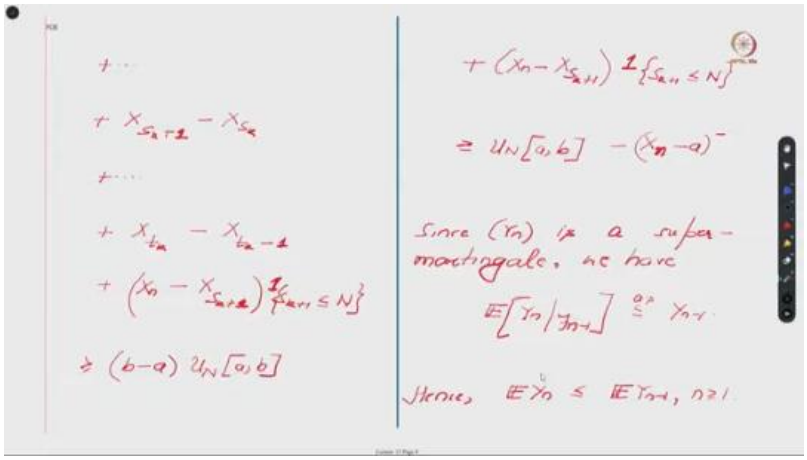
So, we will first set Y0 equals 0 and Yn equals this sum of the first n elements, as we had done in the previous lemma. And because we have defined it this way, and your CKs are bounded non-negative random variables and they are also previsible, it is easy to conclude that this sequence Yn is a supermartingale based on the previous lemma, right? And now, we will look at an alternative description of Yn, right? Which is very, very interesting. So, Yn, one can see, is equal to this sum over here.

So, let me illustrate how you got this sum. So here, if you notice, each term here is multiplied by ck. So if ck is 0, this whole expression will be 0. On the other hand, if ck is 1, then this expression will be xk minus xk minus 1. So if you look at S1, before S1, your CK term would be 0 because no upcrossing has begun.

But as soon as you go to S1, which is the first time, you know went to below a then your c value will become 1 and the first term in this sum where in the c corresponding c value will be 1 would be this expression that is c s1 plus 1 is 1 and hence we have this expression right and until your you know up crossing continues one can see that you would have some expression like this is this okay and one can see that in this dotted sum we actually have a telescopic sum by telescopic sum meaning you know the next term in this expression will be s 1 plus 2 minus s 1 plus 1 and so on And because of this telescopic sum, we would be able to cancel off these terms, cancel off these terms and eventually cancel off these terms. And we will only be left with Xt1 minus Xs1.

And similarly, we will have another telescopic sum. for the next up crossing which will start at S2 and go all the way up till T2 right. So, again we have a bunch of dots here indicating that this you know sequential difference will start at S2 and go all the way up till T2 right and this will continue until the kth up crossing right and then Finally, right, so this is the kth up crossing, right? And finally, we will have some additional terms over here, okay?

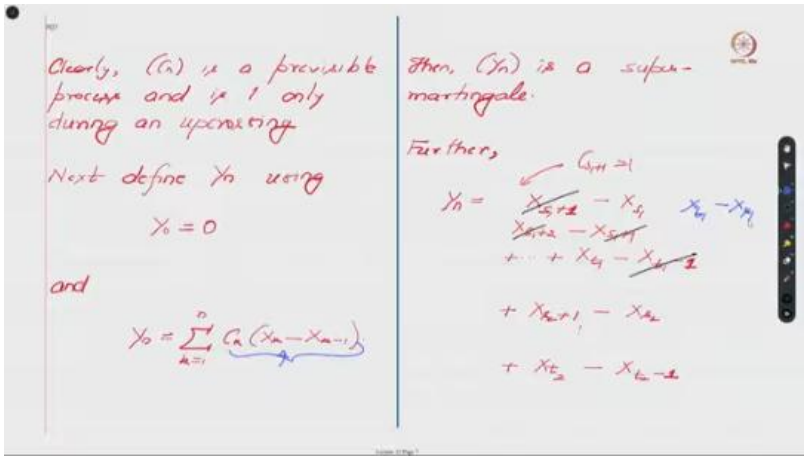


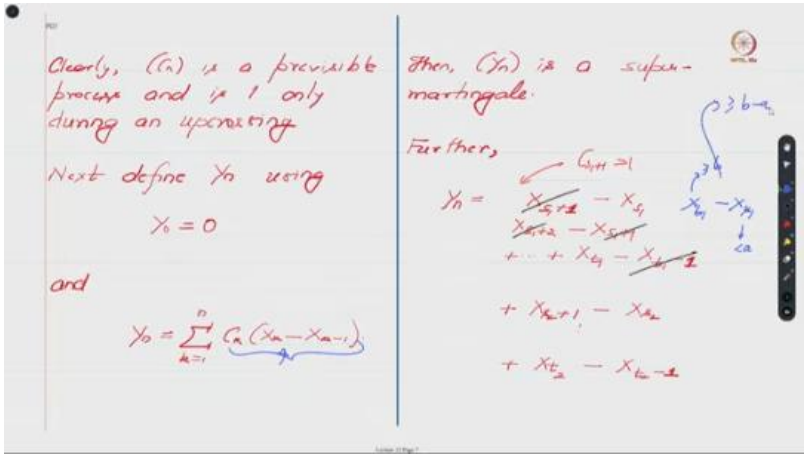


So, this additional terms correspond to this up crossing which began but could not complete, right? Which began but could not complete. So, this is the last term that we have over here, right? So, this term, so recall that on the left hand side, you have yn, right? So, this Yn will equal you know things related to your k up crossings and finally some term which corresponds to the last up crossing which started but did not complete.

So, again from the telescopic nature, we would have Xn minus Xk plus 1. So, this plus 1 is over 1. k itself, whereas this plus 1 is over after sk, right? So, this is sk plus 1, and this is multiplied by the indicator that, you know, your sk plus 1 is less than or equal to n. So, let us understand why we need this indicator. So, it is possible that the last, you know, up-crossing never started, in which case this indicator will be 0.

On the other hand, the last up-crossing did not I mean, started but did not complete, in which case this indicator will be 1, and we will have this term over here, right? So, now, as I mentioned over here, you know, because of this telescopic nature, this whole expression will only end up containing XT1 minus XS1. And by our definitions of T1 and S1, right, this expression will be above B, and this expression will be strictly less than A, and hence this term will be greater than B minus A, right? And one can see that every such up-crossing will lead to a term which is greater than or equal to B minus A. And hence, until this point, one can conclude that the expression is B minus A times the number of up-crossings.





And then we have, you know, this tail part which comes over here. So, at the end, one can conclude that this yn term is greater than or equal to the number of up-crossings. I think I missed a term here. This should be B minus A times this. minus this last part.

So, why is this minus? It basically follows because of this expression and I leave it as an exercise to show why this expression is less than equal to this part over here.

$$Y_0 = 0$$

$$Y_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1})$$

$$Y_n = X_{S_1+1} - X_{S_1}$$

$$+ \cdots + X_{t_1} - X_{t_1 - 1}$$

$$+\cdots+X_{S_2+1}-X_{S_2}$$

$$+X_{t_2}-X_{t_2-1}$$

$$+\cdots$$

$$+X_{S_k+1}-X_{S_k}$$

$$+\cdots$$

$$+X_{t_k}-X_{t_k-1}$$

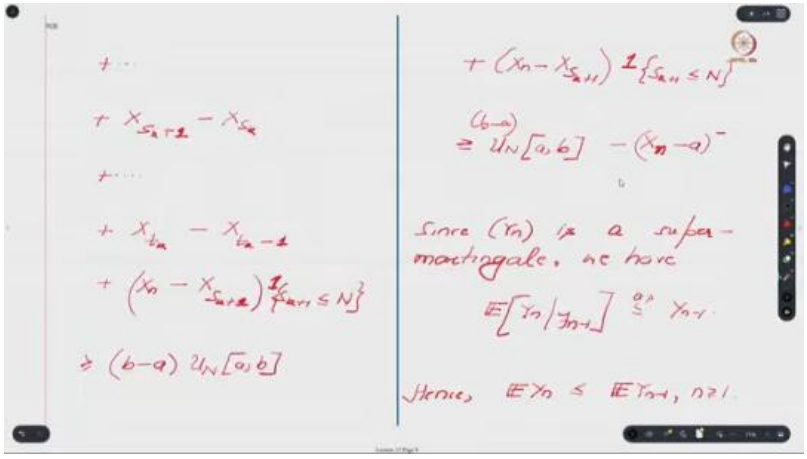
$$+\left(X_n-X_{S_k+1}\right)^1\{S_{k+1}\le N\}$$

$$\ge (b-a)\mathcal{U}_N[a,b]$$

$$+\left(X_n-X_{S_{k+1}}\right)^1\{S_{k+1}\le N\}$$

$$\ge (b-a)\mathcal{U}_N[a,b]-(X_n-a)^-$$

And I will just show you some hints on how you can prove this. So, let us say you have A here and let us say this is X minus Now, xk plus 1 should definitely be you know your sk plus 1 should be strictly less than a. However, your xn you know either can be here, either can be here or a.



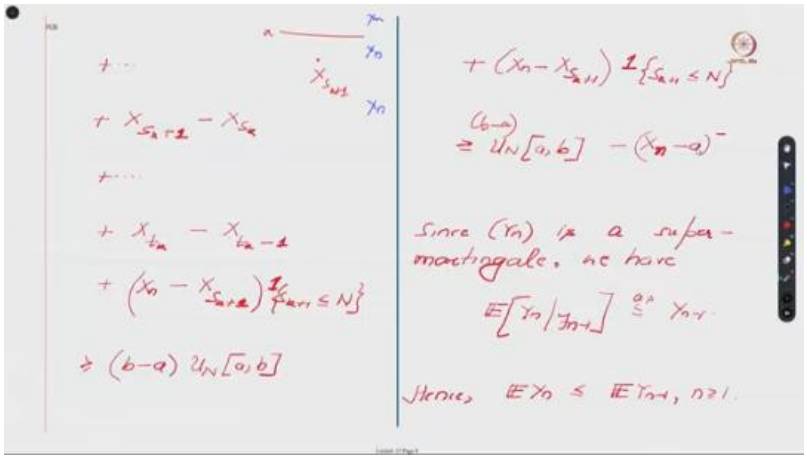
It can be above your line A. So, you consider these three different cases and in either of these cases, you know, check whether this expression is greater than equal to minus of Xn

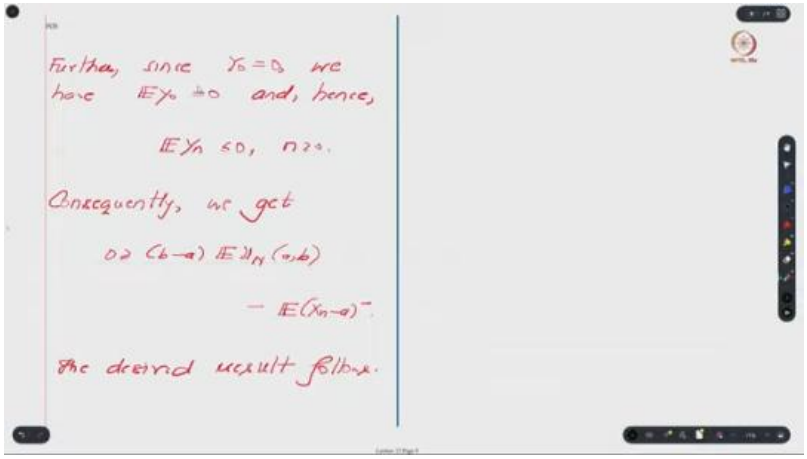
minus A with a superscript minus. Now, we can use the fact that Yn is a super martingale to then recall that the conditional expectation of Yn given Fn minus 1 is almost surely less than Yn minus 1 which if you take expectations on both sides implies expected value of Yn is less than equal to expected value of Yn minus 1. So, you can use this iterated expectation property to conclude this.

$$\mathbb{E}\left[Y_n|_{\mathcal{F}_{n-1}}\right] \leq Y_{n-1}$$

$$\mathbb{E}Y_n \leq \mathbb{E}Y_{n-1}, n \geq 1$$

Now, your Yn is actually bigger than equal to this and we also know that your Y0 is 0. So, Y0 is 0 and your expectation satisfy this property and Yn is bigger than equal to this.





So, if you put all of them together one can conclude that expected value of Yn is less than equal to 0. 0 must be greater than equal to this. So, if you move this to this side one gets the desired result ok.

$$\mathbb{E}T_n \leq 0, n \geq 0$$

$$0 \geq (b-a)\mathbb{E}\mathcal{U}_N(a,b) - \mathbb{E}(X_n - a)^-$$

So, this brings us to the end of lecture 12. Let me quickly summarize what we have studied in this class.

We looked at a new way to construct a supermartingale or a martingale from a given martingale or a supermartingale. In particular, we made use of this pre-visible process and, you know, looked at these successive differences multiplied by this pre-visible process to come up with a new process, and we saw that that process is actually either a supermartingale or a martingale, depending on whether the XK process itself was a supermartingale or a martingale. And then we proved what is called Dub's upcrossing lemma. So, we will, you know, in the next few lectures, see how this upcrossing lemma can be used to prove the convergence of a supermartingale and hence also of a martingale. In particular, in this upcrossing lemma, we were able to relate the number of times a process crosses this interval [a, b], right, in terms of, you know, the value of the process at time n, right.

So, let me just show you that last expression. I think there is a mistake here. Sorry, I should have written little n over here. So, then for all n greater than or equal to 1, this holds.

So, the way to read this process is that the number of upcrossings between 0 and n is upper bounded by the value of the process at time n in terms of this function. So, you will see that this result actually plays a very crucial role in discussing the convergence of supermartingales. With this, I would like to stop. See you in the next class. Thank you and Namaste.

Lemma 11.3 (Doob's Upcrossing Lemma)

Let (X_n) be a submartingale and $U_N[a,b]$ be the number of upcrossings of $[a,b]$ by time N.

Then, $\forall n \geq 1$,

$$(b-a)\,\mathbb{E}\,U_n[a,b] \leq \mathbb{E}\left[(X_n-a)^-\right]$$

$x^- = \max\{-x, 0\}$ | $x = -2$, $x^- = 2$. | $x = 3$, $x^- = 0$.

Proof: Define $(C_n)_{n\geq 1}$ as follows. Let

$$C_1 = \mathbf{1}_{\{X_0 < a\}}$$

and, for $n \geq 2$, let $C_n \in \mathcal{F}_{n-1}$

$$C_n = \mathbf{1}_{\{C_{n-1}=1\}}\,\mathbf{1}_{\{X_{n-1}\leq b\}} + \mathbf{1}_{\{C_{n-1}=0\}}\,\mathbf{1}_{\{X_{n-1}<a\}}$$