STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

Department of Computer Science and Automation

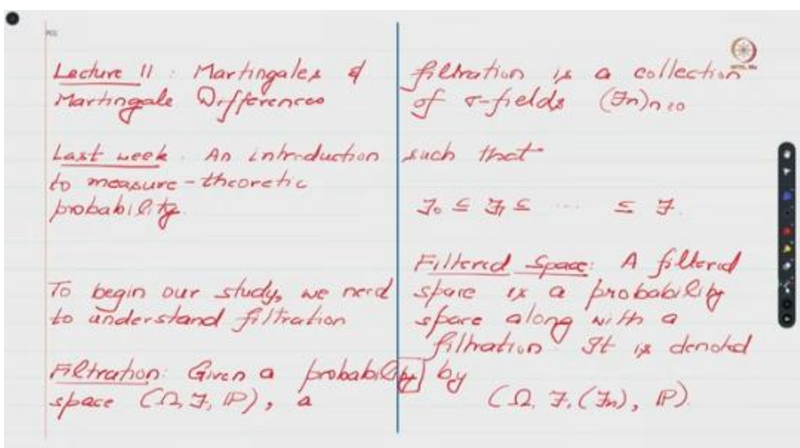
Indian Institute of Science

Lecture 11

Martingales: Definition and Examples

Hello and Namaste, everyone. Welcome to Lecture 11 of this NPTEL course on Stochastic Approximation. Last week, we quickly skimmed over the basics of measure-theoretic probability. This week, we will Now, go a bit slowly and thoroughly understand the convergence of martingale difference sequences by first understanding the convergence of a related stochastic process called martingales.

All right. So, to begin our study, we will need to understand this concept called a filtration. So, what is a filtration? Well, suppose we have been given a probability space (Ω, F, P). I hope with our introduction from the previous week, this notation should now be clear. But nevertheless, let me give a quick recap of what the three things mean.



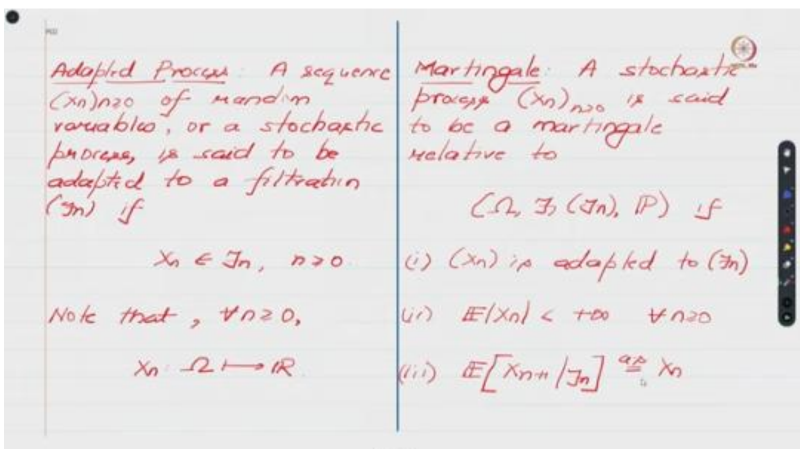
This is your sample space. This is the sigma-field of sets of subsets of Ω, and this is the probability measure defined on F. So, given a probability space (Ω, F, P), A filtration is a sequence of sigma-fields with the property that F_0 is a subset of F_1. F_1 itself is a subset of F_2, and so on and so forth.

And every element in the sequence of sigma fields is a subset of F, right?

$$F_0 \subseteq F_1 \subseteq \cdots \subseteq F$$

So, this is what a filtration is. So, the filtration is basically the sequence Fn. And henceforth, we will refer to what is called a filtered space when we have a probability space along with a filtration. So, what is a filtered space?

A filtered space is a probability space along with a filtration, and we will denote it in the following way with omega, F, Fn, and P. So, now given a filtered space, we will define what is called an adapted process. So, a sequence Xn of random variables, or equivalently a stochastic process, is said to be adapted to a filtration Fn if it satisfies the following properties. That Xn belongs to Fn, right? So, again, when I write this notation, I do not mean that 'belongs to' is in the set notation context.



What I mean is that Fn is a sigma field and Xn is measurable with respect to Fn. And recall that Xn is basically a map from omega to R. So, all Xn's are defined on the same probability space, and Xn belongs to Fn, which basically means that your Xn inverse of B of R is a subset of Fn, which itself is a subset of F.

$$X_n \in F_n, n \geq 0$$

$$\forall\, n \geq 0$$

$$X_n : \Omega \mapsto \mathbb{R}$$

$$X_n^{-1}\big(B(\mathbb{R})\big) \subseteq F_n \subseteq F$$

So, this is what I mean by an adapted process. So, with these definitions in place, we can now define what a martingale is.

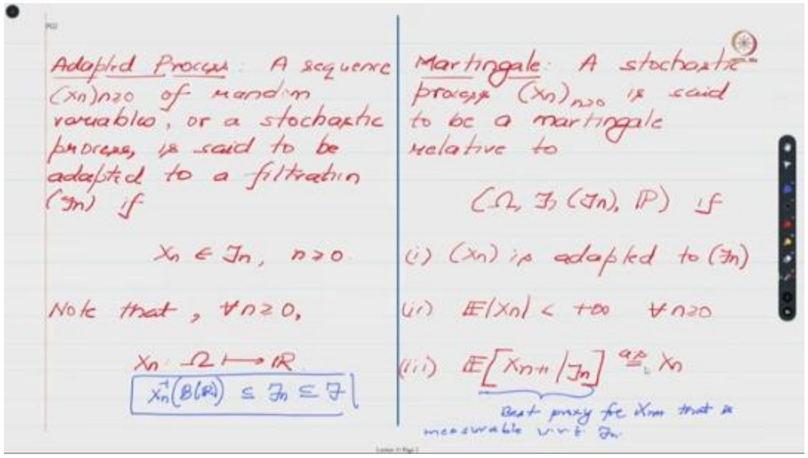
So, in our study of stochastic approximation, we would require this concept of a martingale difference sequence. In order to understand a martingale difference sequence, we will first understand the concept of a martingale. So, for this, we begin with a filtered space, and we will say a stochastic process Xn is a martingale relative to the given filtered space if it satisfies three properties. The first property that we require is that Xn be adapted to this filtration Fn. The next property we require is that all these Xn's are integrable.

In other words, we require that the expected value of the absolute value of Xn is less than infinity for all n. And lastly, the most important property we require is that the conditional expectation of Xn plus 1 given Fn is almost surely Xn, that is, on average, Xn plus 1 behaves like Xn, right? And recall this conditional expectation notation, right? This, at a very loose level, is the best proxy for Xn plus 1 that is measurable with respect to Fn.

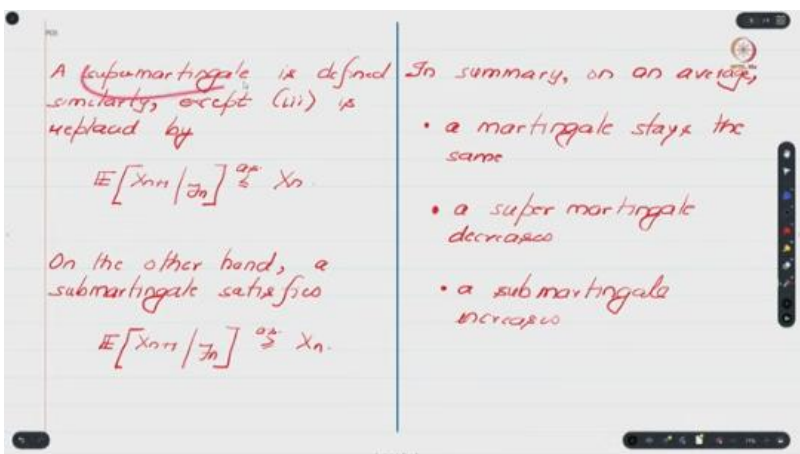
$$\mathbb{E}|X_n| \subset +\infty, \forall n \geq 0$$

$$\mathbb{E}\big[X_{n+1}|_{F_n}\big] = X_n$$

So, the conditional expectation is the best proxy for Xn plus 1 that is measurable with respect to Fn. And we require that the best proxy for Xn plus 1 that is measurable with respect to Fn be Xn itself. And since we have presumed that Xn is adapted to Fn, this Xn will indeed be measurable with respect to Fn. So, let us quickly recap what a martingale sequence is. A martingale is a stochastic process that has three properties.



The first is that the stochastic process is adapted to Fn. The next is that each element in this sequence is integrable, and the third is that the conditional expectation of Xn plus 1 given Fn equals Xn, and this property should hold for all n greater than or equal to 0. If all these properties hold, then we will refer to Xn as a martingale. Now, in a similar spirit, we can define what is called a supermartingale or a submartingale. In a supermartingale, we again require that the first two properties hold.



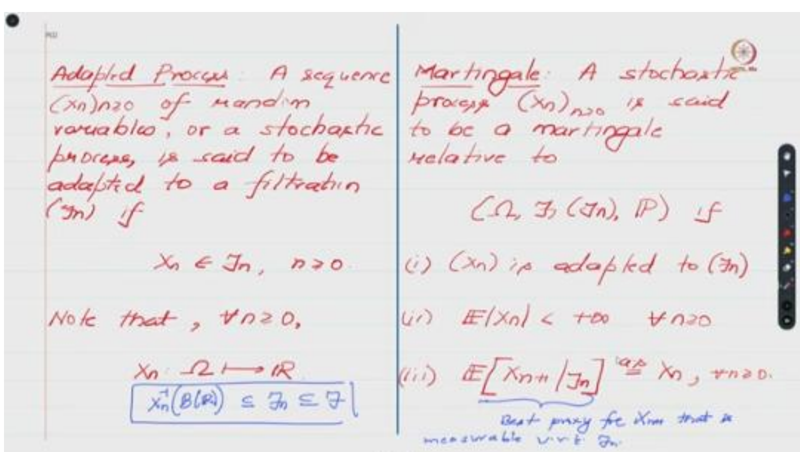
Instead of the third property, we require that Xn satisfies this inequality. For a martingale, recall that we had an equality over here. In a submartingale, we require that the equality be replaced by a less-than-or-equal-to sign, right? So, what is a supermartingale? A supermartingale is a sequence Xn which is adapted to Fn, is integrable, and its conditional expectation satisfies this inequality. In other words, on average, Xn plus 1 is less than or equal to Xn. If the inequality is reversed, we have what is called a submartingale.

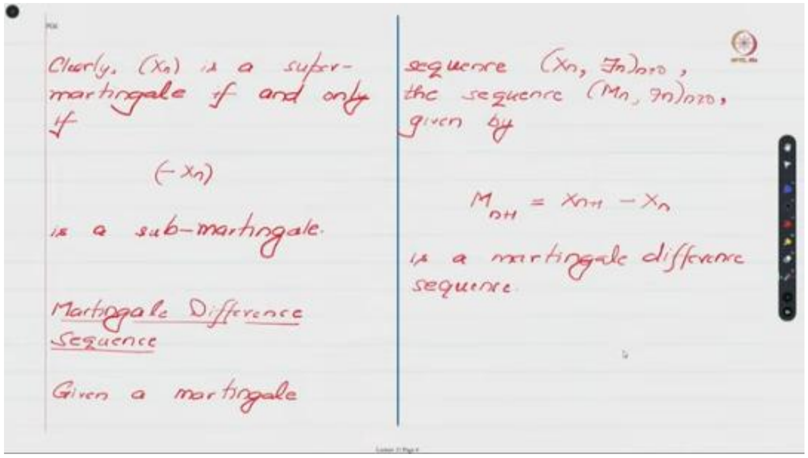
$$\mathbb{E}[X_{n+1}|_{F_n}] \leq X_n$$

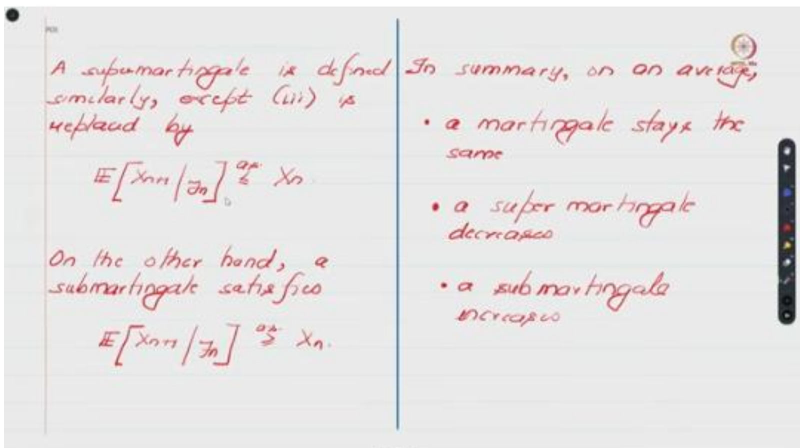
$$\mathbb{E}[X_{n+1}|_{F_n}] \geq X_n$$

So, in summary, a martingale stays the same on average, a supermartingale decreases on average, whereas a submartingale increases. And in all these relations, you would have seen I have written 'almost sure' here and similarly I have written 'almost sure' here.

Because, as I told you in the definition of the conditional expectation, if we change the definition of a random variable over a set of measure 0, then these definitions do not change. Hence, we only require that these relations hold except for a set of probability measure 0. So, I will soon be giving you some examples of martingales, but before that, let us make a few points clear. So, it is obvious to see that a sequence Xn is a supermartingale if and only if minus Xn is a submartingale. So, it is easy to see that if a sequence Xn satisfies this inequality, then if you take the negative of the sequence, it will satisfy this inequality, that is, the inequality would get reversed.







Hence, if Xn is a supermartingale, then the negative of that process will be a submartingale. Now, as I told you in stochastic approximation, the stochastic process that would be of prime importance is not exactly the martingale sequence, but rather it is the martingale difference sequence. So, now let me quickly define what a martingale difference sequence is. So, we begin with a martingale sequence, and henceforth, I will often denote a martingale sequence in the following way: that is, Xn comma Fn.

So, Fn is the filtration, and Xn is your stochastic process, right? So, given a martingale sequence Xn comma Fn, the sequence Mn comma Fn given by Mn plus 1 equals Xn plus 1 minus Xn, right?

$$M_{n+1} = X_{n+1} - X_n$$

So, if you consider this process, then this process is what is referred to as the martingale difference sequence, right? So, now I hope the name makes sense.

Xn is your martingale sequence, and now we are looking at the difference between successive elements of your martingale sequence. Hence, this new process is referred to as the martingale difference sequence. Now, one can easily see that your martingale difference sequence satisfies the following property. If you look at the conditional expectation of Mn plus 1 with respect to the sigma field Fn, right, where Fn is, let us say, the sigma field with respect to which your Xn is adapted, right. So, Xn is adapted with respect to Fn. So, this is the sigma field that is given to you, and let us look at the conditional expectation of Mn plus 1 with respect to this sigma field Fn.

So, from the definition of Mn plus 1, we know that Mn plus 1 equals Xn plus 1 minus Xn. Hence, this conditional expectation equals this conditional expectation. Now, from the linearity of expectation. Conditional expectation, sorry, this expectation equals the conditional expectation of Xn plus 1 given Fn and the conditional expectation of Xn given Fn. However, Xn is measurable with respect to Fn.

Hence, the conditional expectation of Xn with respect to Fn is Xn itself, right? Now, from the martingale property, the conditional expectation of Xn plus 1 given Fn is Xn, and from this, we can see that we have Xn minus Xn, which is 0. And I would like to again highlight that because we are working with conditional expectations, all these equalities actually only hold in an almost sure sense, right?

$$\mathbb{E}\left[M_{n+1}|_{F_n}\right] = \mathbb{E}\left[X_{n+1} - X_n|_{F_n}\right]$$

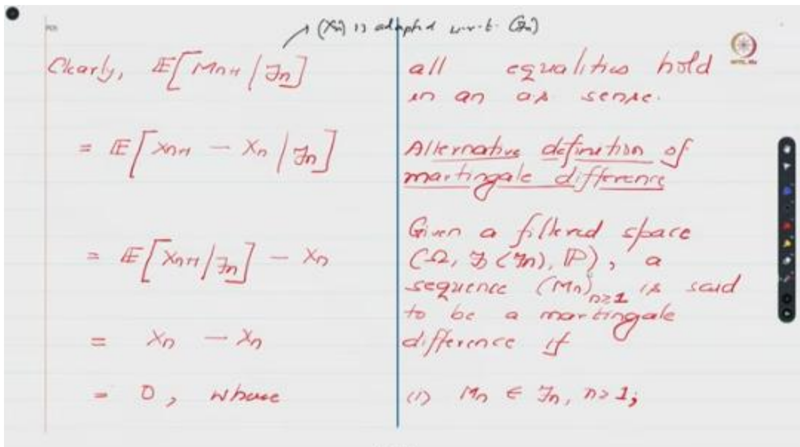
$$= \mathbb{E}\left[X_{n+1}|_{F_n}\right] - X_n$$

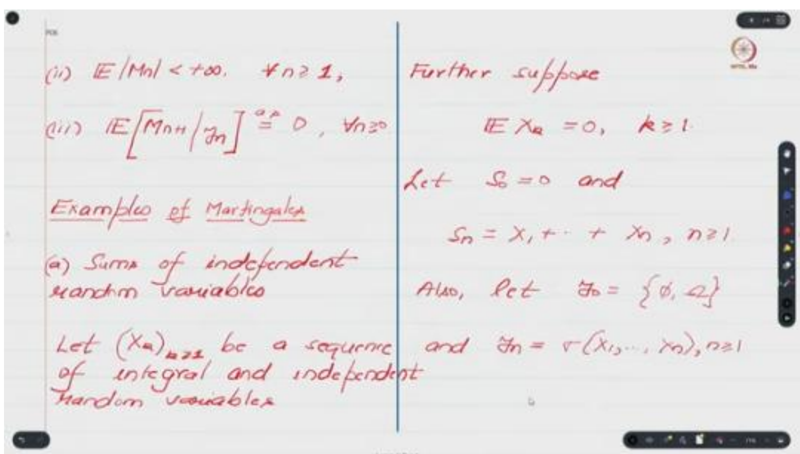
$$= X_n - X_n$$

$$= 0$$

So, let us just summarize what we have seen. That if we have been given a martingale sequence and if we look at the difference between the successive elements, then their conditional expectation of the successive differences, that is, the conditional expectation of Mn plus 1 given Fn, is 0 almost surely.

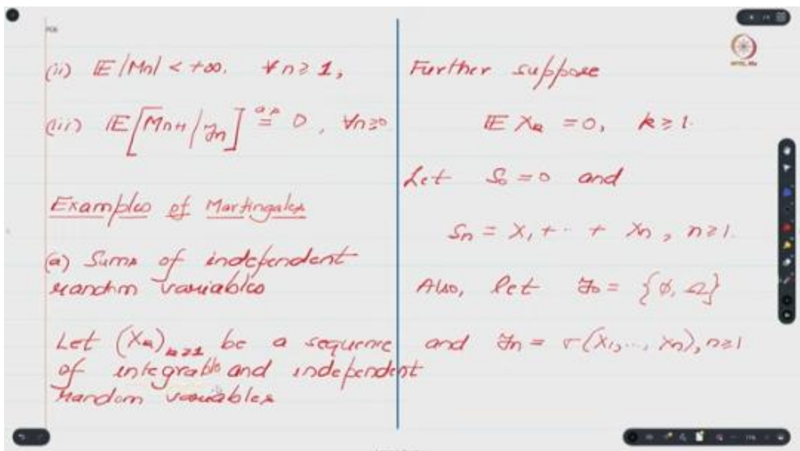
So, your martingale difference sequence actually satisfies this special property. Hence, we can also come up with an alternative definition of a martingale difference sequence, which is given now. So, given a filtered space omega f, sequence Fn and P, a sequence Mn is called a martingale difference if it satisfies the following properties: that is, Mn is adapted to Fn, right? Or, in other words, for every n, your Mn is measurable with respect to Fn. Furthermore, your Mn is integrable for every n, and importantly, the conditional expectation of Mn plus 1 given Fn is 0 almost surely, right?





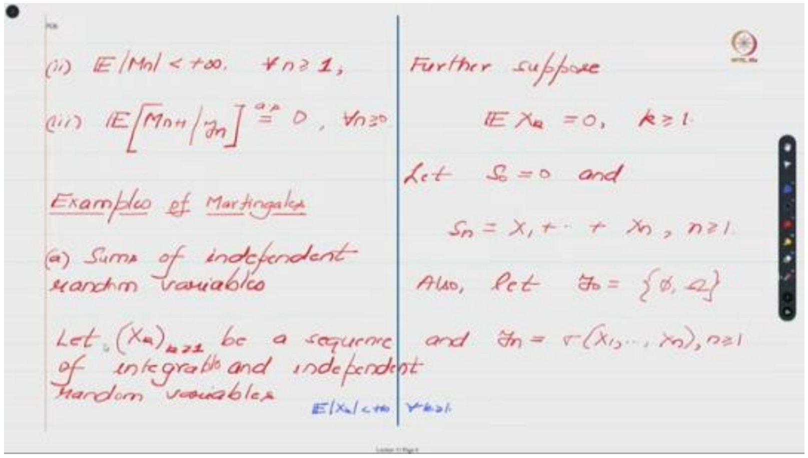
And as an exercise, I would encourage you to check whether a sequence of 0-mean independent and identically distributed random variables is a martingale difference sequence or not, okay? So, now, as promised, we will look at some examples of martingales, and once you see martingales, you can easily cook up examples of martingale difference sequences. The first of the examples that we will be looking at corresponds to sums of independent random variables. So, let us look at the example. Let us say we have been given a sequence xk.

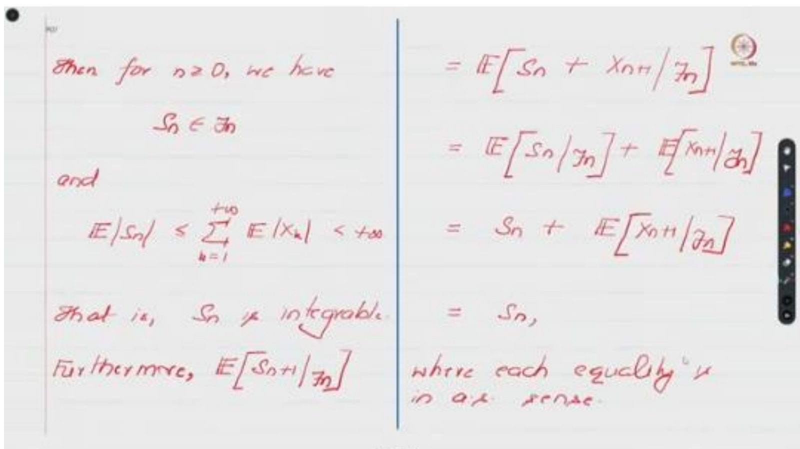
So, I should, I think, correct this; this should be. So, let xk be a sequence of integrable and independent random variables. So, integrable means that your expected value of xk is less than infinity for all k greater than or equal to 1. Further, suppose that each of your xk's has 0 mean. Now, with this definition of xk, let us cook up another sequence of random variables as follows.



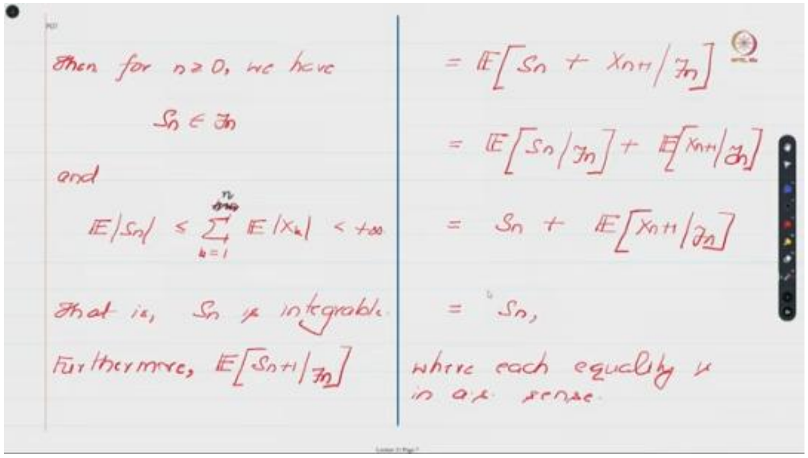
This new sequence is made up of your Sn random variables, and let S0 equal 0 and Sn equal the sum of the first n elements of the sequence xk. So now we have S0, S1, S2, S3, and so on. So we now have a sequence of random variables. Now the claim is that this new sequence is actually a martingale sequence. Now as soon as I say that something is a martingale sequence, the first thing that should arise in your mind is: with respect to which filtration?

So, let us answer that question, and towards that, let us define the following filtration. So, recall that a filtration is a sequence of sigma fields which are monotonically increasing. So, the filtration that we work with is the one where the 0th sigma field is basically the trivial sigma field made up of only the empty set and the sample space, and for every n bigger than or equal to 1, fn equals the sigma field that is generated by the first n random variables, that is, x1 to xn. And because x1 to xn are all measurable with respect to fn, and Sn is their sum, it is easy to see that Sn is measurable with respect to Fn, right?





So, for those who are not able to see why that is the case, I encourage you to look at this textbook on random variables from 'Probability Path' by Resnick, all right? And then, from some property that is satisfied by the expectation, which is referred to as the triangle inequality, one can show that the expected value of the absolute value of Sn is upper bounded by the expected value of the absolute value of K, where K goes from 1 to n. I think I made a mistake here. This should not be infinity; rather, it should be n. Okay? So, by one property satisfied by the expectation, one can see that the expected value of the absolute value of SN is upper bounded by the sum of the expected value of the absolute value of XK, where K ranges from 1 to N. And because each xk is integrable, each expectation is finite.



And since this is a sum of finite elements, each of which is finite, the sum itself is finite. In other words, the expected value of the absolute value of Sn is less than infinity. And hence, we can conclude that Sn itself is integrable for each n, right? So, we have now verified two of the three properties for when a sequence Sn can be claimed to be a martingale sequence.

$$\mathbb{E}|S_n| \leq \sum_{k=1}^{n} \mathbb{E}|X_k| < +\infty$$

So, we now move on to the third and final property, which is that of the conditional expectation of Sn plus 1 given Fn.

So, again from the definition of Sn, one can express Sn plus 1 as Sn plus Xn plus 1, and now if you take its conditional expectation, from the linearity property, one can express this conditional expectation as the conditional expectation of Sn given Fn and the conditional expectation of Xn plus 1 given Fn. Now, because Sn is measurable with respect to Fn, as we saw over here, this conditional expectation results in only Sn, and we rewrite this as above. Then we make use of the fact that Xn is independent with respect to Fn. Hence, this quantity over here equals the expected value of Xn plus 1. And furthermore, we know that each of these random variables has 0 mean, and hence this expression over here equals 0, and hence this whole expression equals Sn.

$$\mathbb{E}\left[S_{n+1}|_{F_n}\right] = \mathbb{E}\left[S_n + X_{n+1}|_{F_n}\right]$$

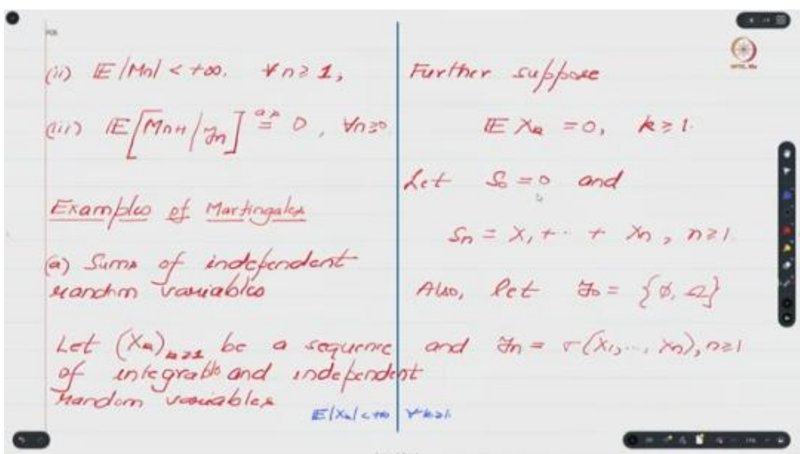
$$= \mathbb{E}[S_n|_{F_n}] + \mathbb{E}[X_{n+1}|_{F_n}]$$

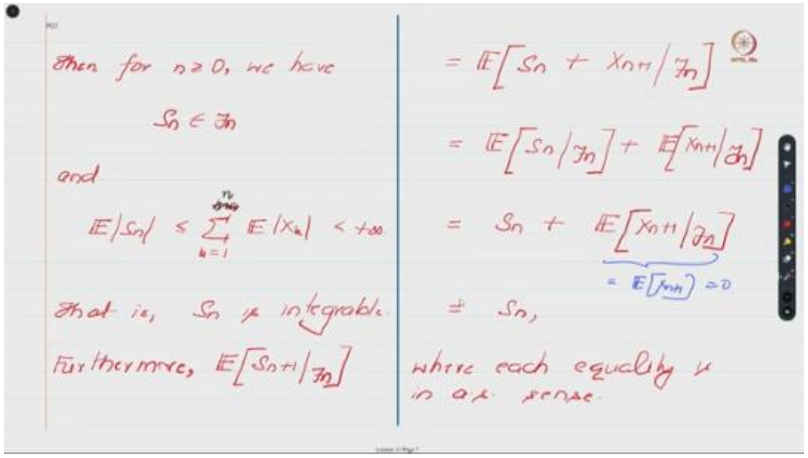
$$= S_n + \mathbb{E}[X_{n+1}|_{F_n}]$$

$$= S_n$$

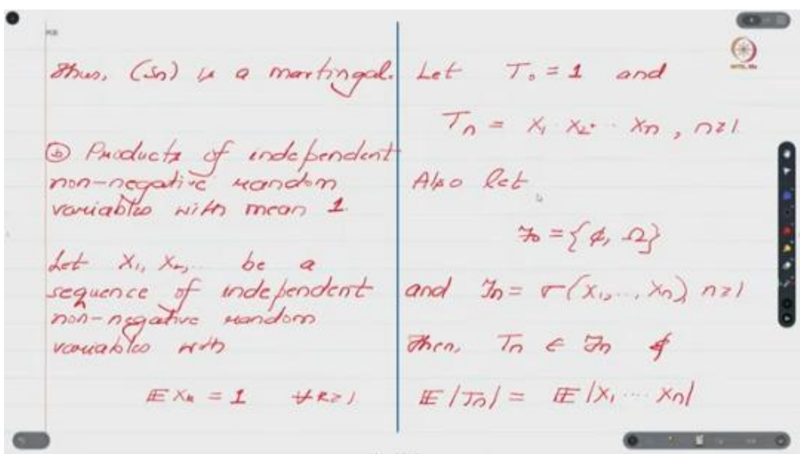
And again, I would like to highlight that all these four equalities actually hold only in an almost sure sense.

So, what have we seen so far? If I take the conditional expectation of Sn plus 1 given Fn, then it equals Sn, and from this, we can conclude that Sn is actually a martingale sequence. So, we see that the sum of independent zero-mean random variables actually forms a martingale sequence. So, from this, one may perhaps start believing that maybe to get a martingale sequence, we need this additive form, right? The additive form that we saw over here, right? So, using the next example, I would like to show that no, that need not be the case, okay?





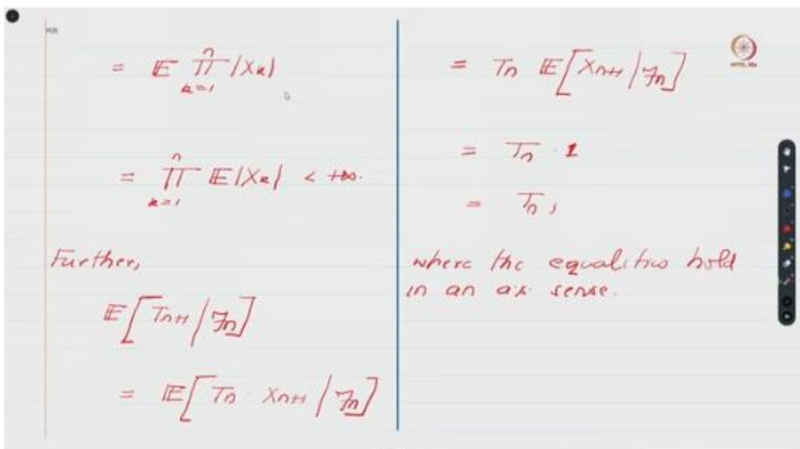
So, the next example looks at products. In the first example, we looked at sums. In the next example, we are going to look at products of independent non-negative random variables with mean 1. So, let us look at this example in more detail. Let us say we have a sequence of independent non-negative random variables.



With the property that the expected value of Xk equals 1 for all k greater than or equal to 1. So, with this property, we are now going to construct a martingale sequence in the following way. Let T0 equal 1 and Tn equal the product of the first n many elements, that is, Tn equals X1 times X2 times X3, all the way up to Xn. And now the claim is that Tn is a martingale sequence. Again, one should ask with respect to which filtration.

So, let us construct that filtration. So, let F0 be the trivial sigma-field generated by the empty set and the sample space. And for every n greater than or equal to 1, let Fn be the sigma-field generated by the first n elements of this sequence, X1, X2, and so on. So, the first n elements will comprise X1, X2, and so on, all the way up to Xn. And since Tn is a product of X1 to Xn, again one can see easily that Tn is measurable with respect to Fn.

So, that sort of verifies the first property necessary for a sequence to be a martingale sequence. The next property that we require is that Tn be integrable, right? So, towards that, let us look at the expected value of the absolute value of Tn, and since Tn is the product, this expectation equals the expected value of the product of X1 to Xn, right? Which in compact form I have written in the following way. And now, we make use of the fact that each of these random variables is independent, and because they are independent, the expectation of the product becomes the product of expectations, and hence this thing equals this.



And we know that each Xk is integrable; hence, the expected value of the absolute value of Xk is less than infinity, and since it is a product of finitely many numbers. The product will also be less than infinity, allowing us to conclude that Tn is integrable.

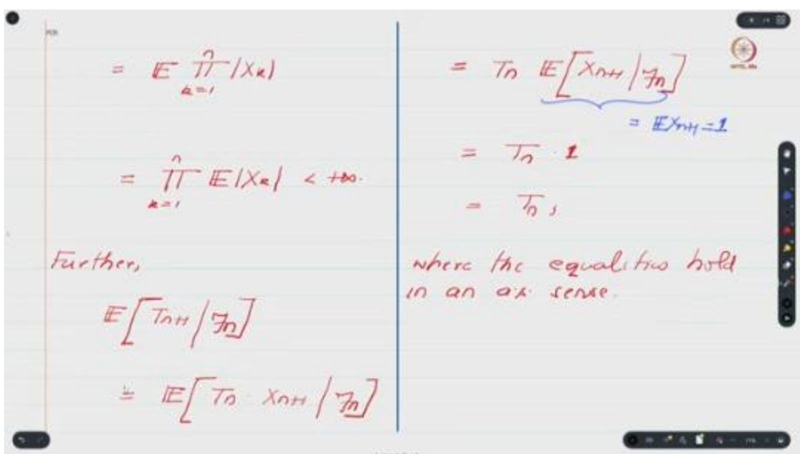
$$\mathbb{E}|T_n| = \mathbb{E}|X_1 \dots X_n|$$

$$= \mathbb{E} \prod_{k=1}^{n} |X_k|$$

$$= \prod_{k=1}^{n} \mathbb{E}|X_k| < +\infty$$

So, this completes the verification of the second property of a martingale sequence, and it now remains to verify the third property, which basically concerns the conditional expectation of Tn plus 1 given Fn. So, let n greater than or equal to 0 be arbitrary, and let us look at the conditional expectation of Tn plus 1 given Fn. And from the definition of Tn plus 1, we can break this into Tn times Xn plus 1.

Now, the conditional expectation of this product equals Tn times this conditional expectation, and this happens because Tn is measurable with respect to Fn. So, I hope you recall the properties of conditional expectation, which we had seen in the previous week, which we are invoking here to be able to write it in this fashion. And again, from the independence of the Xn sequence, one can see that this conditional expectation equals the expectation of Xn plus 1, and we know that each of these expectations equals 1. Hence, this quantity equals 1, and that is what I have written over here, and Tn times 1 is Tn itself. And again, I would like to highlight that all of these equalities, you know, hold in an almost sure sense.



So, since we are dealing with such examples for the first time and since we are going over these equalities for the first time, I am emphasizing this almost sure sense.

$$\mathbb{E}\left[T_{n+1}|_{F_n}\right] = \mathbb{E}\left[T_n \cdot X_{n+1}|_{F_n}\right]$$

$$= T_n \mathbb{E}\left[X_{n+1}|_{F_n}\right]$$

$$= T_n \cdot 1$$

$$= T_n$$

So, now let me summarize today's class. In today's class, we looked at the definition of a martingale, and we also looked at the definition of a martingale sequence. Then we looked at two very simple examples of martingales. In particular, we looked at the sum of zero-mean independent random variables, and we also looked at the product of independent random variables with unit mean.

So, the two examples were taken from standard textbooks, and their purpose was to illustrate that a martingale sequence need not just be of the sum form; it can also have a product form. In the next class, we will look at some additional properties of martingales. With that, let me thank you for joining this lecture. Namaste.