STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

Department of Computer Science and Automation

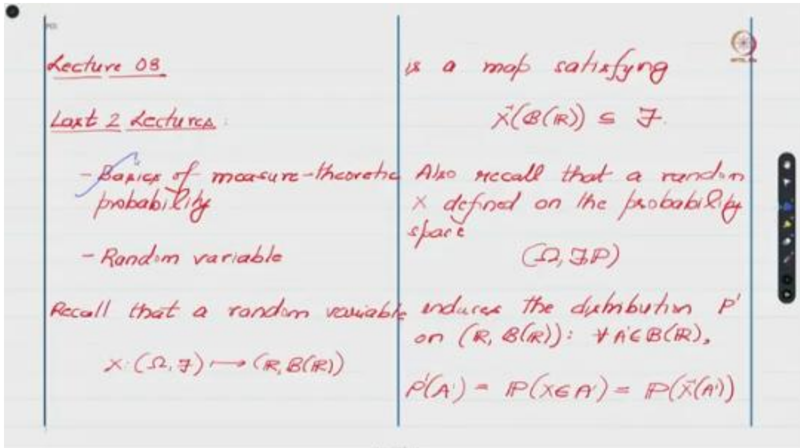
Indian Institute of Science

Lecture 8

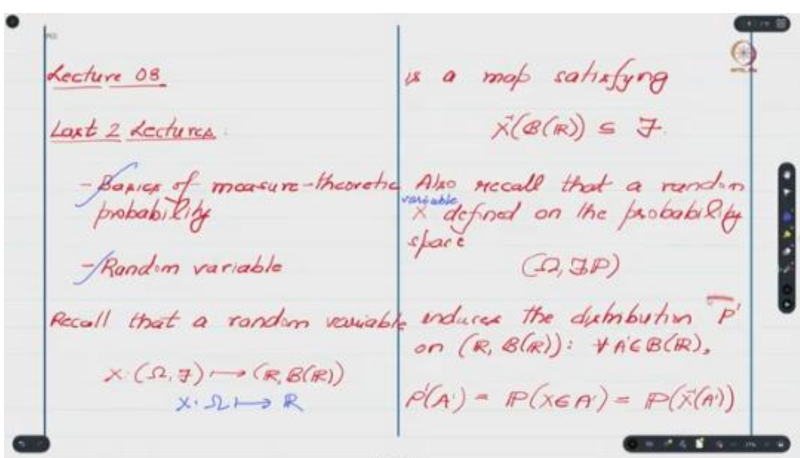
Expectation of Random Variable as Lebesgue Integration

Hello and Namaste everyone, welcome to Lecture 8 of this course on Stochastic Approximation. So, let us do a quick recap of what we did in the last two classes. As I told you, the way we handle noise is by invoking some of the limiting results, such as the strong law of large numbers and so on and so forth. In order to do so, on one hand, we could work with independent and identically distributed random variables with zero mean. However, that would limit our applications, and hence what we do is we look at the next best thing to IID zero-mean random variables, which in the previous class—or the two classes before—I mentioned that they are these martingale difference sequences. And towards understanding martingale difference sequences, we have to understand this concept of conditional probability. In order to understand a bit of conditional probability, I have been giving you some basics of probability from this textbook called Probability Path by Resnick, right?

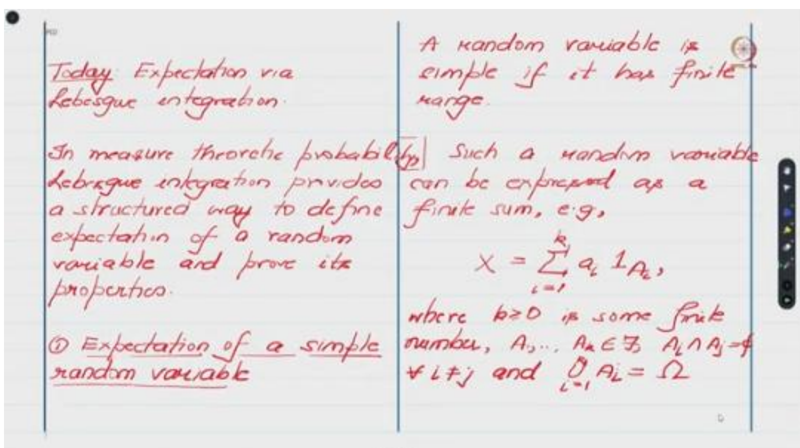
In the past two lectures, we looked at some basics of measure-theoretic probability, and then we looked at this concept of random variables, right? Okay, and recall the way I defined these random variables. I said that a random variable is a function, okay? A random variable is a function which goes from omega to R, and it is a function that satisfies this inverse relation in the following sense. So, B of R, recall, is a collection of subsets of R which together form a sigma-field, and X inverse is basically the inverse map applied to every element in this collection. So, this is another collection, and we require that this collection be a subset of F, in which case we say that X is measurable and, in which case, we say that X is also a random variable. So, again, I would like to highlight that while I write X as a map between omega F to R B R, I actually mean that X is a map from omega to R, right? And in addition, it satisfies a condition like this, right?



And then we also made a point that whenever we have a random variable X—so there is a typo here, it should be saying a random variable X—I made this point that whenever we have a random variable X, which is defined on some probability space omega F and P. Then, this random variable induces a distribution, which we denote as P prime on R, B, R. So, what do I mean by this P prime distribution? Well, you give me any set in B of R, then P prime of A prime is equal to the probability of X belonging to A prime, which is equal to the probability of X inverse of A prime. So, since A prime is an element in B of R and since X is a random variable, this element over here will belong to F. And P is a function that has been explicitly defined on every element in F; hence, P of this quantity that we have over here is well-defined, and we sort of assign this value to P prime of A prime. And often, to keep things simple, we do not want to talk about this P prime; hence, we will use this notation.



So, the summary of this discussion is that whenever we have a probability space in which we have some probability measure P, and let us say we have a random variable which is defined from this space onto R, then this random variable induces a new distribution onto, you know, this new measurable space R, B, R, and this new distribution is defined in this fashion. So, this is a summary of what we saw last time, and what we will be doing today is we will be looking at the concept of expectation. Now, I am sure you must have already studied this expectation either in your early days of probability or perhaps also in your undergraduate course. So, we will be looking at the same notion of expectation, but in a slightly, you know, measure-theoretic fashion—that is, we will be looking at this concept of expectation from this perspective called Lebesgue integration. So, when you look at it for the first time, it may look a bit difficult, and it may look very different from what you have studied and so on and so forth.



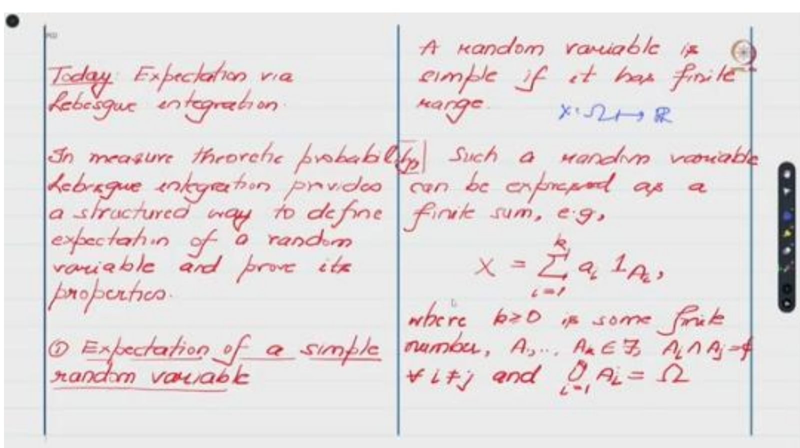
So, one can, you know, indeed ask why we should work with Lebesgue integration if you already know some notion of expectation. So, the answer to this question is, you will soon see that, you know, this way of defining expectation for random variables allows us, on one hand, to define expectation for a much more general class of random variables, and also it will turn out to be very easy to use this definition to prove several of the nice properties that expectation satisfies. Now, I would not have the time in this course, and in particular in this class, to prove all the results that we will be looking at, but as I have already mentioned, we will be relying on this book called Probability Path to quickly brush up our basics of measure-theoretic probability, and you know, when we go to conditional expectations and in particular when we talk about martingale differences and their convergence, these basics will hopefully help us, you know, get through the material quickly. Alright, so now, how do we define expectation under this Lebesgue integration idea? In the Lebesgue integration idea, in some sense, we define expectation in three stages.

In the first stage, which I denote by 1 over here, we begin by defining the expectation of what we refer to as the simple random variable. So, when is a random variable said to be simple? Well, a random variable is said to be simple if it has a finite range. Finite range means, recall that X is a map from omega to R, right? So, range means the set of values that X represents.

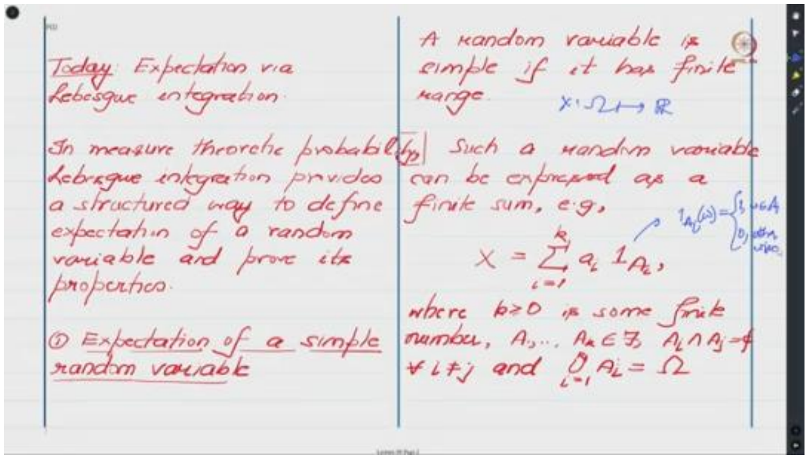
takes right, and we say X is a simple random variable if the range of values it takes is finite. Right? Now, for such random variables, one can show that one can have a description of the following form, right. So, this random variable, whenever it is finite, has a description of the following form, which is given over here.

$$X = \sum_{i=1}^{k} a_i \ 1 \, A_i$$

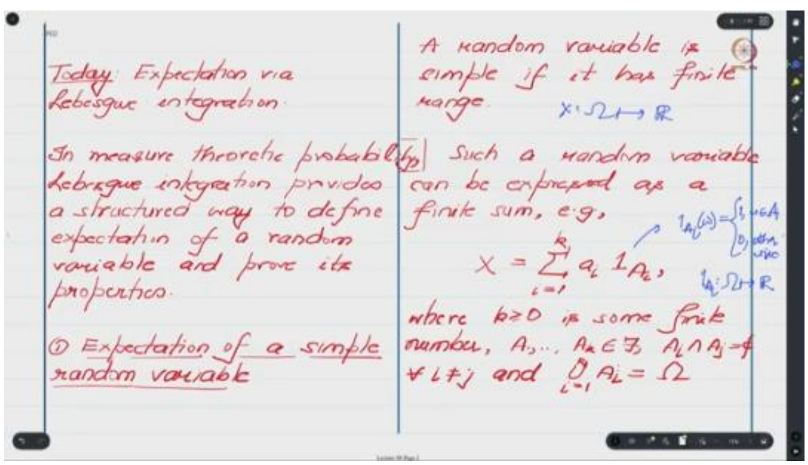
So, you can see that we have a finite sum. So, you see that there is a k over here, and k can be any finite number like 5, 10, 1 million, 1 billion. So, it can be as large a number as we want, but we require that it be finite, and this X



has the form AI indicator AI. So, recall that this indicator random variable over here is given by: you give me any little omega in capital Omega, then you give me any little omega in capital Omega, then this 1 Ai of omega will be 1 if omega is in Ai and 0 otherwise. So, what does this tell us? Well, the indicator of A, or the indicator of AI, is also a map from omega to R, and it is a 0, 1 map, right? So, for some inputs, it takes the value 1.



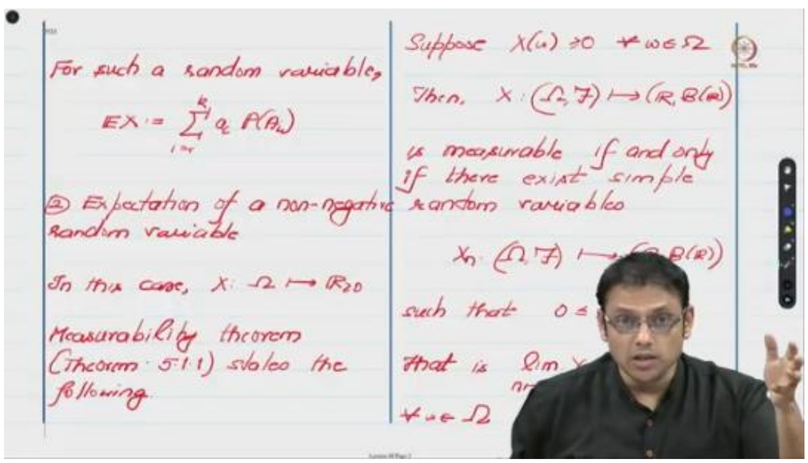
In particular, when the input is from the set AI, it takes the value 1, and it takes the value 0 when the input to this function, which is this indicator AI, is outside the set AI. Right. So, the claim is that whenever you have this random variable X, and if you say it is simple, it means that one should be able to express it in this fashion. Okay. All right.



And we require that these sets AI satisfy some special properties. On the one hand, we require that all these elements, that is A1 to AK, be in your sigma field, which is defined on your original space. In addition, we require that the pairwise intersection be empty. And furthermore, we require that their union equals the whole sample space omega. In other words, we require that A1 to Ak form a partition of your sample space omega.

Okay, so let me again summarize. What is the first stage of defining expectation via Lebesgue integration? Well, you define the expectation of what is called a simple random variable. So, what is a simple random variable? A simple random variable is any random variable whose range is finite, and whenever its range is finite, we can express it as a sum of indicator random variables scaled by the scalars ai.

So it is easy to see that whenever omega lies in capital AI, x will take the value small ai. That is, whenever omega lies in capital AI, x of omega will be little ai, and one can see that this definition extends for any value of ai where i ranges from 1 to k. And because A1 to Ak, that is, these capital A1 to Ak, form a partition of omega, it implies that if you give me any little omega in capital omega, it indeed has to lie in one of these A1 to Ak. So, in that sense, x is completely defined using this expression. And for such a random variable, we sort of rely on our basics of probability to define its expectation in the following way.



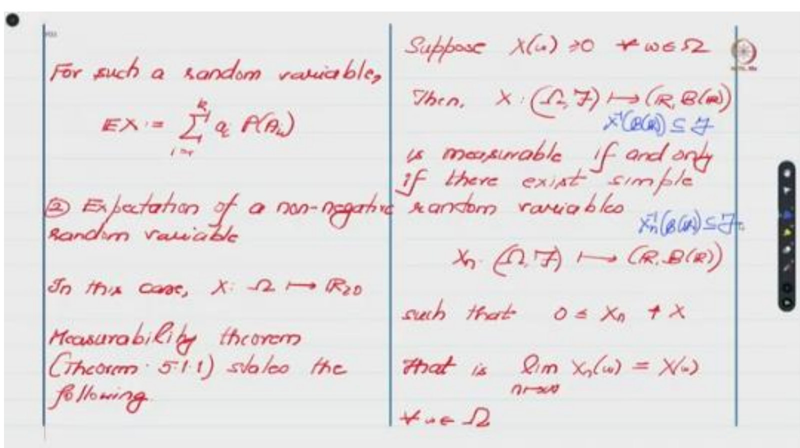
Now, because Ai lies in your sigma field calligraphic F, P of Ai is well-defined, and we define the expectation of X to be this finite sum of Ai times little ai times the probability of capital Ai.

$$\mathbb{E}X := \sum_{i=1}^{k} a_i \; \mathbb{P}(A_i)$$

So, this completes the first stage of defining the expectation of a random variable. So, using this expression over here, we can define the expectation of a simple random variable. So, now we can move on to the second stage, which involves defining the expectation of what is called a non-negative random variable. Right, so in this case, X is again a map from Omega, but this time X is allowed to take any value in the set of non-negative integers, right? And we do not restrict X to only have a finite range; it can, for example, take every value in this set of non-negative real numbers, right?

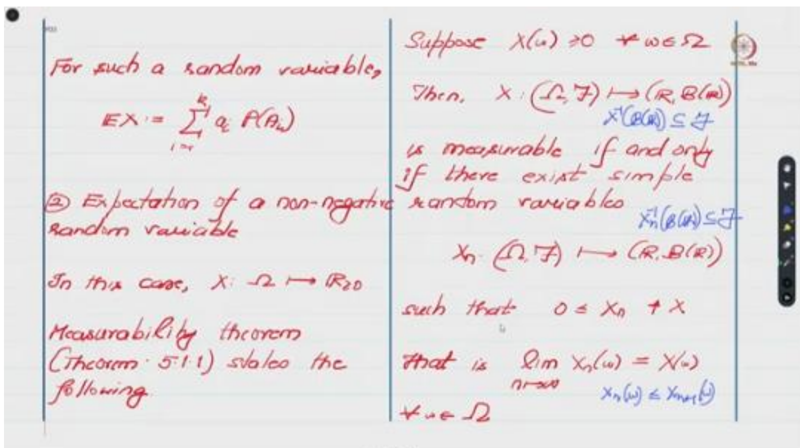
So, we only require that the output of X be a non-negative real number, right? And then there is this very nice theorem called the measurability theorem, which in the probability path textbook is numbered as 5.1.1, which states the following. So, what does it say? Suppose we have a non-negative function, which means that X of omega is greater than or equal to 0 for all little omega and capital Omega. So then this result says that X is measurable in the following sense. It means that X inverse of b of r is a subset of f if and only if there exists a sequence of simple random variables.

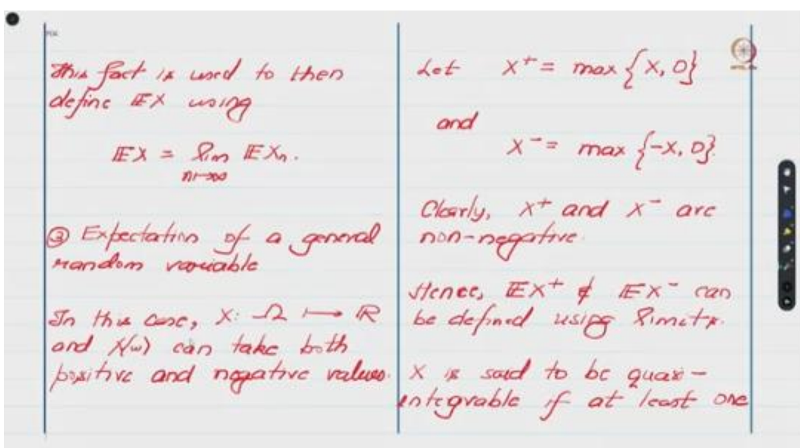
So, the sequence is made up of xn, and each xn is measurable in the following sense, that is, you have xn inverse of b of r is a subset of f. So, this condition is satisfied by every n. And we require that, or I should say, X is measurable if and only if there exists a sequence of simple random variables Xn. So each Xn is a simple random variable, and they satisfy this condition that is specified over here. So, what does this condition say?



It says that every random variable is non-negative, which means that for every little omega, Xn of omega takes a non-negative value. Furthermore, Xn monotonically increases and converges to X. So what it means is that you have this Xn of omega, so this will be less than Xn+1 of omega, and this will continue. That is, X0 of omega will be less than X1 of omega, which will be less than X2 of omega, and so on and so forth. So for every omega, this monotonicity will hold. Furthermore, for every omega, we will also have this limiting condition that is written over here. That is, if you fix little omega and take little n to infinity, then the limit of Xn of omega will be X of omega.

So, let me summarize this theorem again. It says that if you give me any non-negative random variable, then there exists a sequence of simple random variables which are non-negative and monotonically converge to X. Is this okay? So, this is what it says. Now, this fact can be used to define the expected value of X, right? So, what is the idea?



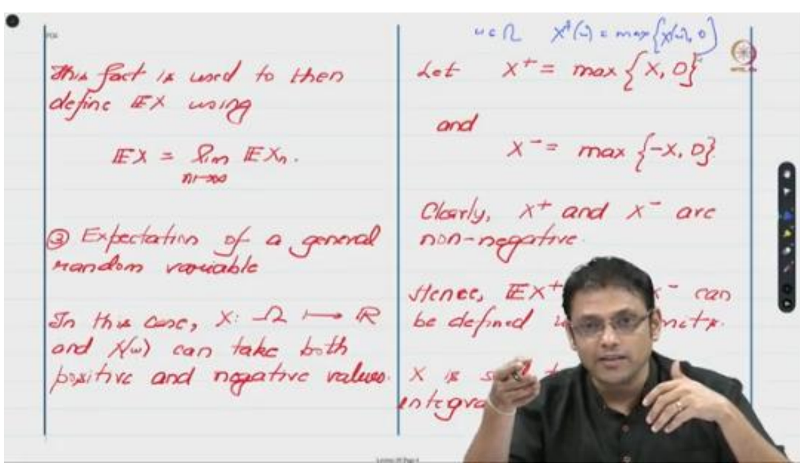


Now, since Xn is a simple random variable, its expectation was already defined in the first stage, right? So, we defined the expectation of a simple random variable in the first stage. So, we can use that definition to define the expected value of Xn. Now, for a non-negative random variable X, its expectation is defined as the limit of the expected value of Xn. Is this okay? So, this completes the definition of the expectation of a random variable up to the second stage.

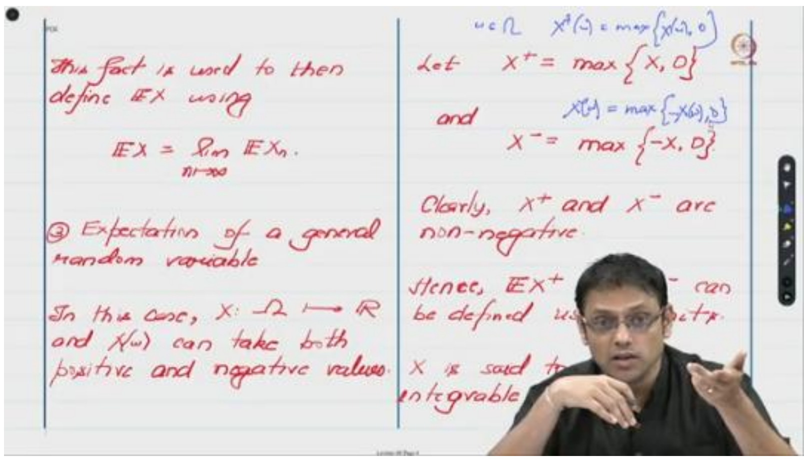
So, now we know how to define the expectation of a simple random variable, and based on that, we know how to define the expectation of a non-negative random variable. So, now comes the third stage, where we ask how we can describe the expectation of a general random variable. So, what do I mean by a general random variable? Again, X is a map from capital omega to R, and this time, X of omega can take positive, negative, and also zero values. So, it can take any number in R, and again, its range need not be finite.

It can, for example, take every value in R. So, the question in the third stage is how we define the expectation of such a random variable. So, in the measure-theoretic sense, what we do is we first define two random variables, X plus and X minus, right? And X plus is defined to be the max of X comma 0. So, what does this mean? You give me any omega in capital omega; then, X plus of omega is defined to be the max of X of omega and 0.

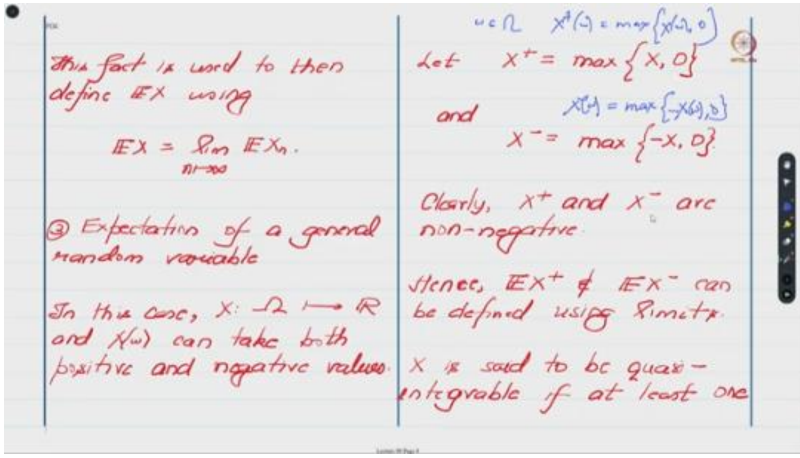
So, what it means is that if X of omega for the given omega is 5, then X plus of omega will also be 5. On the other hand, let us say for some omega, X of omega is minus 2. In that case, X plus of omega will be 0. So, because of this description, one can see that X plus is actually a non-negative random variable. And similarly, we have this X minus definition, which is the max between minus X and 0.



Again, the interpretation here is that x minus of omega is the max of 0. x minus of x minus of x of omega and 0. So, if your x of omega is the given function x, and if x of omega is 5, then x minus of omega will be 0. On the other hand, if x of omega is, let us say, minus 2 for your input omega, then x minus of omega will be plus 2. So, I hope you are now able to understand what this and this means, and in this chapter on random variables in Resnick's textbook, you can see why if X is a random variable, then X plus and X minus are also random variables.



So, it is very easy to show that if X is a random variable, then both X plus and X minus are also random variables. Now, because of this definition of X plus and X minus, one can see that both X plus and X minus are non-negative. Now, because they are non-negative, we can look at or make use of the second stage of our definition of expectation of a random variable and use that to define the expected value of X plus and the expected value of X minus. So, you see that we were able to define the expectation of a non-negative random variable as the limit of the expectation of a sequence of simple random variables. So, because X plus and X minus are non-negative, we can use that same idea to define their expectations as well.

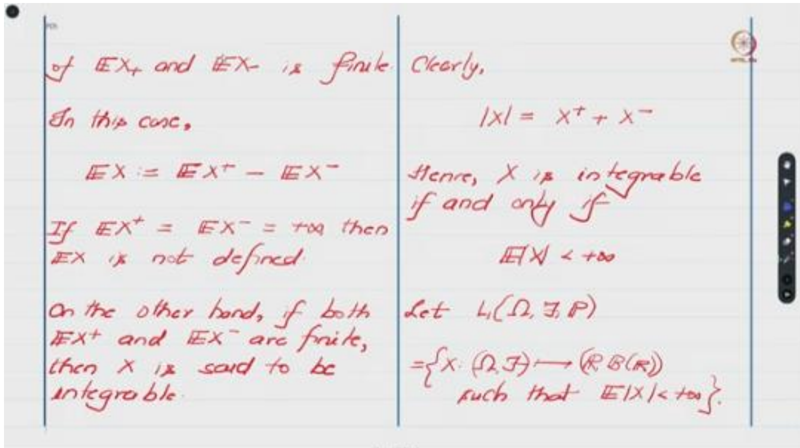


So, now the question is: how can we use the expected value of X plus and the expected value of X minus to define the expected value of X? This is the question that we have in mind, right? So, toward that, we will introduce some concepts. The first among them is what is called quasi-integrability. So, we will say the given random variable X is quasi-integrable if at least one of the expected value of X plus or the expected value of X minus is finite, right?

So, if one among them is finite, we will say that your random variable X is quasi-integrable. And in that case, we will define the expected value of X to be the expected value of X plus minus the expected value of X minus, right?

$$\mathbb{E}X := \mathbb{E}X^{+} - \mathbb{E}X^{-}$$

And you can already see that because one of them is finite, this difference is well defined. Now, it would be challenging if both these values were infinity. If both these values were infinity, then we end up in this scenario of infinity minus infinity, which is not defined.



So, what we will do is if the expected value of X plus and the expected value of X minus are both infinity, then in that scenario, we will say that the expected value of X is undefined. On the other hand, if both the expected value of X plus and the expected value of X are finite, we will say that X is integrable. So, let us just summarize. So, for the general random variable—by general, I mean the one which can take both positive and negative values—we first define this X plus and X minus.

Now, X plus and X minus are non-negative random variables; hence, their expectation is always defined. Now, if one among them is finite, then we will say X is a quasi-integrable random variable, in which case we can define the expected value of X using this relation that we have over here. On the other hand, if both the expected value of X and the expected value of X minus are infinite, then this expectation is undefined, right? And we will say X is integrable if both the expected value of—sorry—the expected value of X plus and the expected value of X minus are finite. Now, it is easy to see that the absolute value of X—now, the absolute value of X is also a function.

So, this goes from omega to R, and how is this defined? Well, if you give a little omega as input, it is equal to the absolute value of X of omega. So, this is how you define this function over here, and one can see that this function is basically the sum of these two functions, that is, X plus and X minus.

$$|X|: \Omega \mapsto \mathbb{R}$$

$$|X|(\omega) = |X(\omega)|$$

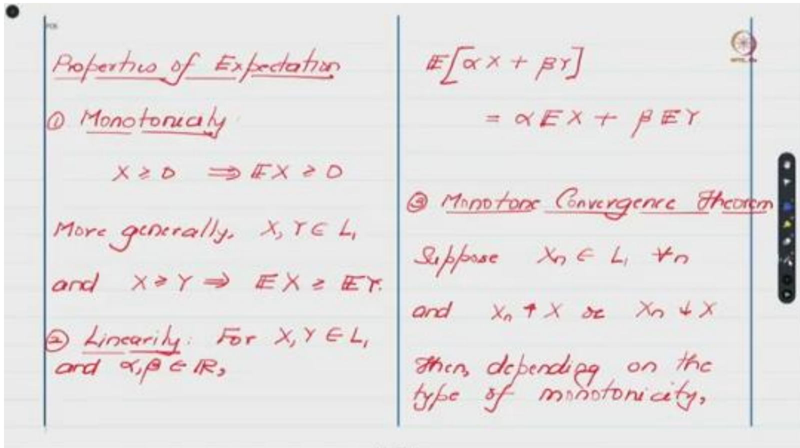
$$|X| = X^+ + X^-$$

So, recall that the original X that we had over here, okay? So, I think I have not mentioned it, so let me do that. Okay, so the original X, I hope you agree, satisfies this relation. Okay, so capital X is equal to X plus minus X minus.

Okay, so it is very easy to see from the descriptions of X plus and X minus, and using the same idea, one can see that the absolute value of X equals X plus plus X minus. And one can see that the random variable X is integrable if and only if the expected value of the absolute value of X is finite. Now, one can ask where we have defined the expected value of the absolute value of X. Well, X plus is a non-negative random variable, X minus is a non-negative random variable, hence their sum will also be a non-negative random variable. And because it is a non-negative random variable, the expected value of the absolute value of X is always defined, whether X is integrable or not, and hence one can use this description to check whether X is integrable or not. In particular, we will say X is integrable if the expected value of the absolute value of X is less than or equal to infinity, right?

And often, we will collect all random variables which are integrable and put them in this set, okay? So, we will say L1. So, omega fp is a given probability space, and L1 omega fp refers to the collection of all random variables such that their expectation is finite. In other words, L1 omega fp consists of all integrable random variables. And whenever the probability space is clear, I will refer to this set simply as L1.

So, whenever I say X belongs to L1, I mean that X is an integrable random variable. So far, we have defined the expectation of a random variable, and just to quickly summarize, we did it in three stages. We started with simple random variables, then we moved on to non-negative random variables, and then we finally established it for general random variables. Alright, so now we will look at the properties that this expectation operation satisfies, right? So, I have chosen a subset of the properties. The expectation operation actually satisfies several more properties, and I request the listener to refer to this probability textbook to look at the other properties.



In this collection, I have chosen some of the important properties, and in particular, I have chosen those which we will often be using throughout this course. The first of these properties is what is called the monotonicity property. So, what does the monotonicity property say? It says that suppose X is a non-negative random variable; then one can show that its expectation is always non-negative. Now, recall that whenever we have a non-negative random variable, its expectation is always defined.

It can be infinity, but it is defined, right?

$$X \geq 0 \Rightarrow \mathbb{E}X \geq 0$$

So, the claim is that if your X is non-negative, then the expected value of X is non-negative. More generally, suppose we have two random variables in L1, which means that these two random variables have finite expectations. And suppose X is greater than or equal to Y. So, what does this condition mean? So, X and Y are functions.

So, this condition over here implies that x of omega should be bigger than y of omega for all little omega and capital omega. So, if you have such a condition, the monotonicity condition or property of expectation implies that the expectation of X should be bigger than or equal to the expectation of Y.

$$X \geq Y \Rightarrow \mathbb{E}X \geq \mathbb{E}Y$$

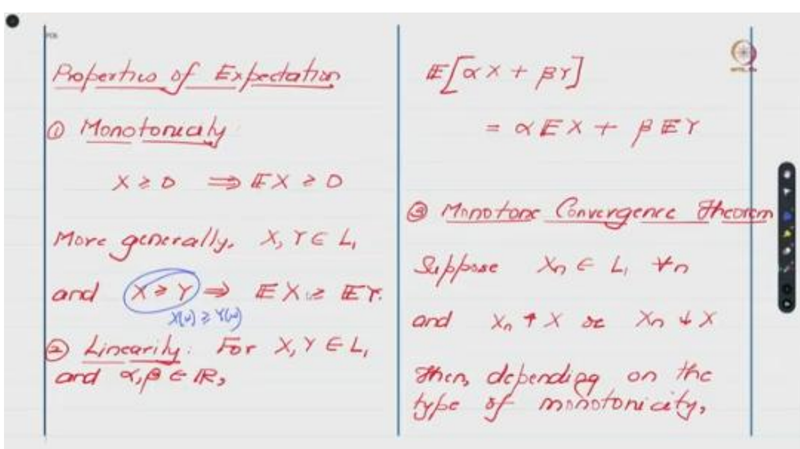
$$X(\omega) \geq Y(\omega)$$

So, this concludes the description of the first property, and the way we go about proving these properties is that you first show it for simple random variables and then for non-negative random variables. So, as I told you at the beginning, when you initially look at these Lebesgue integrations and so on and so forth, it may look a bit intimidating, but once you start working with these properties, you will see that we can actually do it in a very automated fashion. You just follow some standard recipe, and one can prove all these properties just by following that recipe.

The next property that we will look at is that of linearity. The linearity property says that suppose you have two integrable random variables x and y, and let us say we have two real numbers alpha and beta. Then, the expected value of this linear combination of X and Y dictated by these weights alpha and beta is the linear combination of their individual expectations, which means that the expected value of alpha X plus beta Y equals alpha times the expected value of X plus beta times the expected value of Y.

$$\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}X + \beta \mathbb{E}Y$$

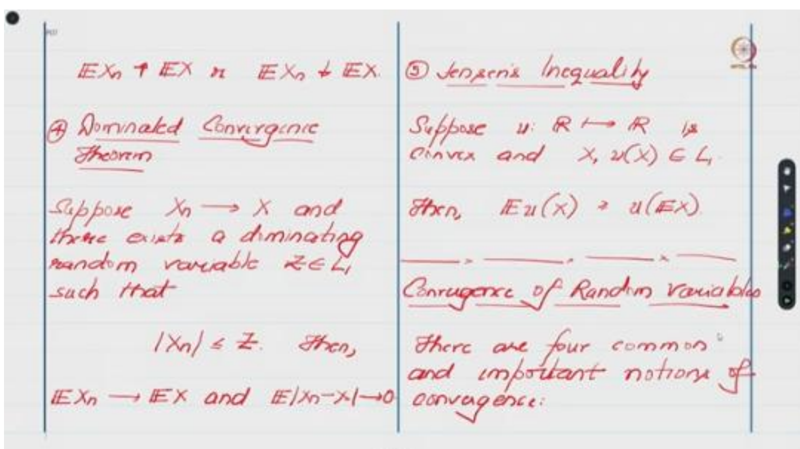
Again, this can first be proved for simple functions, then for non-negative functions, and then finally for integrable functions. For those who are not used to such calculations, I encourage you to do it once on your own and really see how easy these properties are to verify.



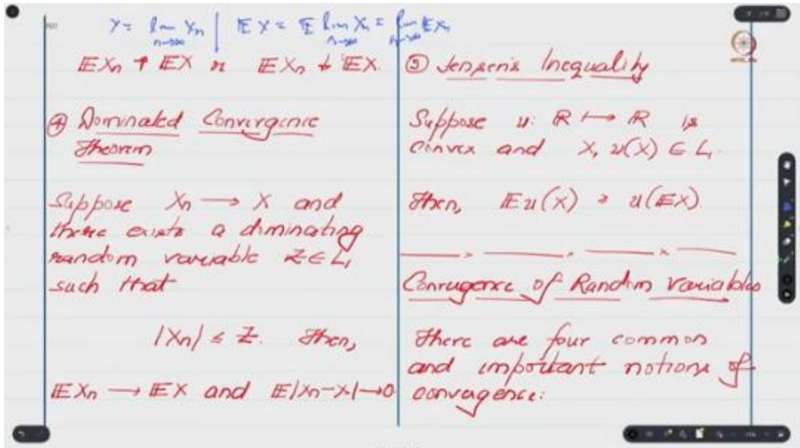
The third of the properties, and in fact one of the most powerful properties, is that of the monotone convergence theorem. The monotone convergence theorem says the following:

suppose you have a sequence of integrable random variables. So, you have Xn belonging to L1 for all n. Furthermore, suppose that these Xn's either monotonically increase to X, or there is another scenario where there is another random variable X to which these Xn's monotonically decrease. So, this is one scenario; this is another scenario.

Then, depending on the type of monotonicity, one can show that the expected value of Xn also monotonically increases to the expected value of X. Or vice versa, that is, the expected value of Xn will monotonically decrease to the expected value of X. So, what this result says is that if you have a sequence of random variables which monotonically increases to X, then the limit of the expected value of the limiting random variable will be the limit of the expectation. So, let me just write that down. So, X is equal to the limit as n tends to infinity of Xn, and this monotone convergence theorem says that the expected value of X equals the expected value of the limit as n tends to infinity of Xn.



This is equal to the limit as n tends to infinity of the expected value of Xn. So, you can see that this monotone convergence theorem allows us to exchange this limit and expectation. The first equality is by the definition of X, and the second equality is the consequence of the monotone convergence theorem. It says that this expectation and limit can actually be interchanged. So, this is a very powerful statement we are making over here.

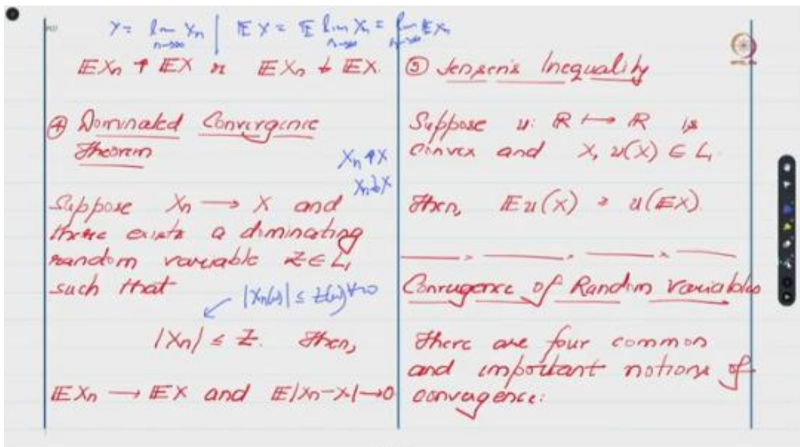


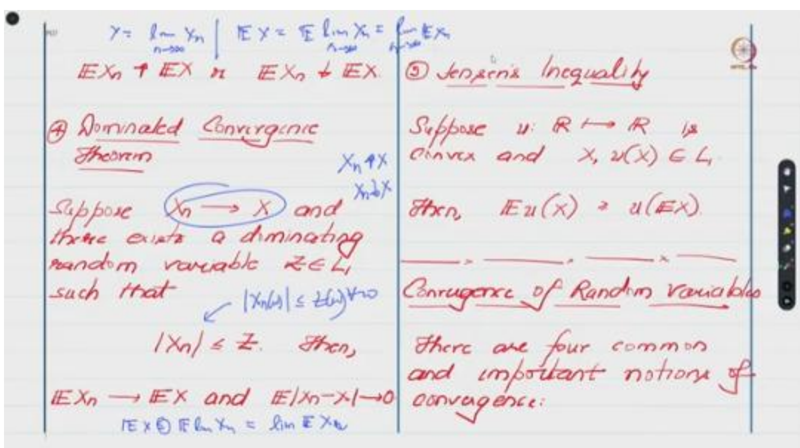
Right. So, the next property is that of the dominated convergence theorem. Okay. There is one more point that I want to talk about regarding the monotone convergence theorem. Some of you may think, 'Oh, we had used some property in the second stage and so on.'

So recall that there, these accents were simple random variables. In this monotone convergence theorem, these Xn's need not be simple random variables, and we are saying that this limiting nature of the expectations continues to hold even for general random variables. So, we now move on to the fourth of the set of properties we are looking at, and this includes the dominated convergence theorem. So, here the setup is that we have a sequence of random variables, and Xn converges to X. However, notice that here I have not said that Xn monotonically increases to X or Xn monotonically decreases to X. Rather, you know, it can fluctuate around the true value of X of omega and eventually converge to X of omega.

For every little omega, Xn of omega could sometimes be above X of omega, below X of omega, and these values keep fluctuating and eventually settle on X of omega. So, now the question we are asking is, in such a scenario, is this interchange of limit and expectation still true? And the dominated convergence theorem says that yes, it is true if you have what is called a dominating random variable. So, a dominating random variable is one which itself is integrable and is an upper bound to the absolute value of Xn for all n. So, what this condition implies is that the absolute value of Xn of omega is less than Z of omega for all omega.

So, in such a scenario, we will say that your Z is a dominating random variable, and in that case, one can show that the expected value of Xn goes to the expected value of X. Again, the interpretation here is that the expected value of X equals the expected value of the limit of Xn. This equality again follows from the given fact that Xn goes to X, and the dominated convergence theorem is basically saying that you can actually exchange the limit and the expectation. So, the limit and the expectation. So, then this is the fourth of the important properties that we need to know, and the last of the important properties that we need to know is that of Jensen's inequality.





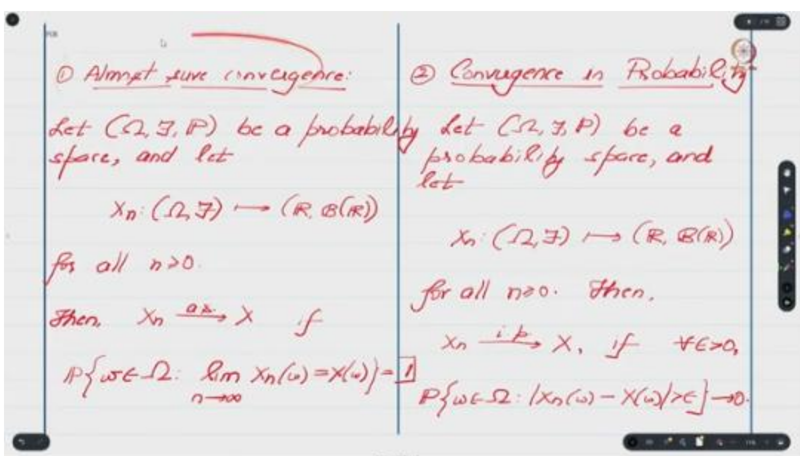
So, although there is a J over here, we often refer to this name as Jensen's inequality. So, what does Jensen's inequality say? It says that suppose you have a function, little u, which goes from the set of real numbers to the set of real numbers and is a convex function. And furthermore, let us say X, which is a random variable, and U of X, which is another random variable. Both these random variables are integrable random variables, right?

So, this expectation and this expectation are well defined. So, Jensen's inequality states that the expected value of U of X is always bigger than U of the expected value of X.

$$\mathbb{E}\, u(X) \geq u(\mathbb{E}X)$$

So, this is a very powerful inequality, and often when you know it, you can exploit it in several ways. So, this is the last of these properties that I would like to talk about. Again, as I said, there are several properties that expectation satisfies, and for the complete list of properties, I request you to look at this textbook by Resnick, which is called A Probability Path. So, the next thing we will be talking about is what is called the convergence of random variables.

As I told you, we want to talk about the strong law of large numbers in the context of noise. In particular, we want to talk about the strong law of large numbers for the sequence of martingale difference sequences. So, when we talk about these limiting results, we want to make sure we understand what we mean by the convergence of random variables and so on, and here is a quick brush-up of what this means. So, there are four common or important notions of convergence. The first of them is what is called almost sure convergence.



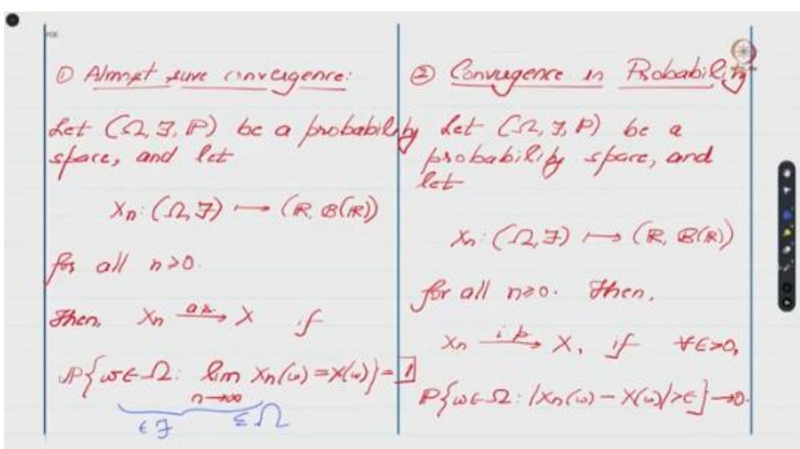
So, in almost sure convergence, we are given a probability space. And we have a sequence of random variables, all defined on the same probability space. So, that is very important. And we will say Xn converges almost surely to X, which in notation is given over here, right? So, on the arrow, we write A dot S dot to indicate almost sure convergence.

So, this convergence holds if this probability is 1. So, what is this probability? Well, it is the set of all little omega in capital omega, where the limit of Xn equals capital X of omega. So, it is the collection of little omegas where Xn of omega goes to X of omega. So, this collection of sets—I mean, this collection of little omegas—I hope you agree that this is a subset of capital omega. One can show that if X is a measurable, if Xs are measurable random variables, then this limit is also a measurable random variable.

In other words, one can show that this set is actually an element of calligraphic F, which is given over here, right? So, one can actually show that, and hence we can talk about the probability of that event. So, we will say Xn almost surely converges to X if the probability of this event is actually 1.

$$\mathbb{P}\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = 1$$

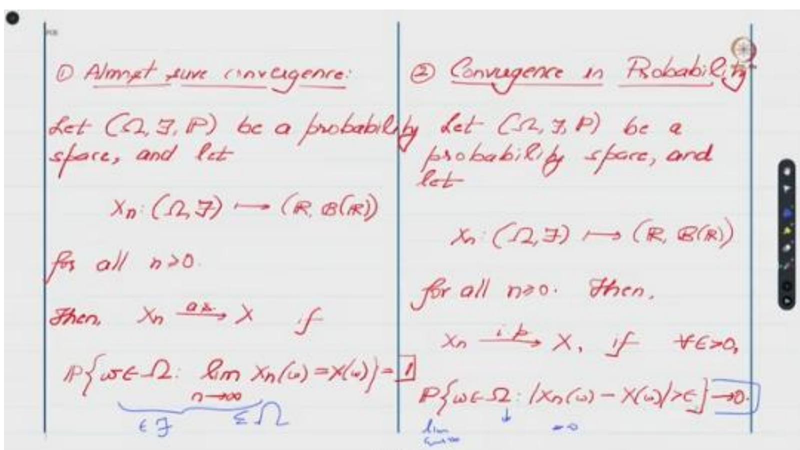
So, now one can ask: does it not imply that Xn goes to X of omega for all omega? Well, the answer is no, and I encourage you to think about it and see that, on the collection of little omegas whose probability is 0, this convergence need not hold.

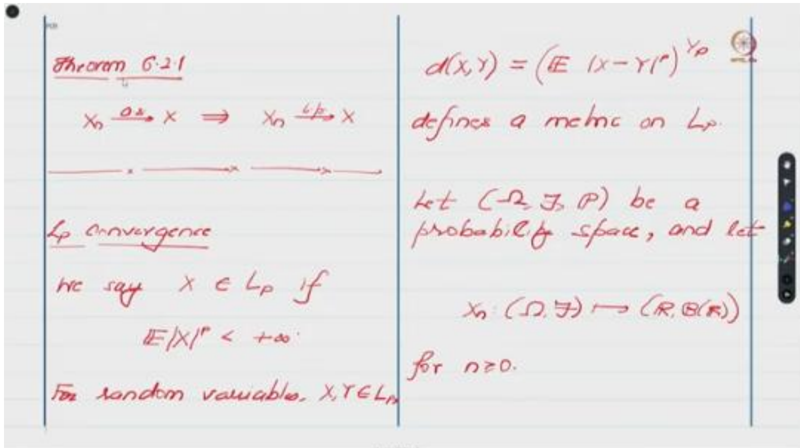


So, this covers the first notion of convergence. The second notion of convergence is that of convergence in probability. Right here again, we are given a probability space, and we require that the sequence of random variables be all defined on this common probability space. And we will say this sequence Xn converges to X if, for every epsilon greater than 0, this probability goes to 0. So, let us understand what this probability is.

So, notice that in the almost sure convergence sense, we had a limit inside the probability itself—that is, the collection of omegas where a certain property held, and that property itself had this limit included in it. Here, on the other hand, you can see that the event or the input or the set which is given as input to this P function does not involve any limit. It is only for some n, right? So, this function that we have is a function of n, and we require that the limit of this function, as n tends to infinity, should be 0, as we have highlighted over here, right? So, what does this collection of omega imply?

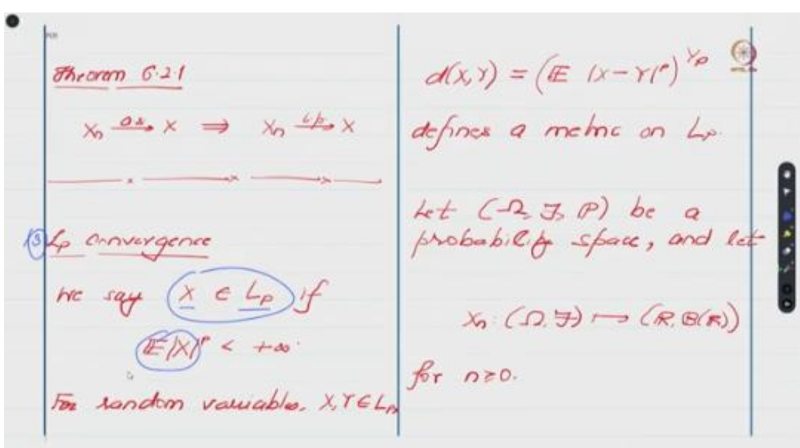
So, this is the collection of omega where Xn of omega minus X of omega exceeds this given threshold epsilon, right? So, you fix an epsilon. And you collect all those omegas where Xn of omega minus X of omega exceeds epsilon. And you look at the probability of that event, and one can show that the probability of this event goes to 0 if Xn converges in probability to X. In other words, this is the definition of Xn converging in probability to X. And one can look at Resnick's textbook—in particular, one can look at Theorem 6.2.1—to conclude that almost sure convergence actually implies convergence in probability.





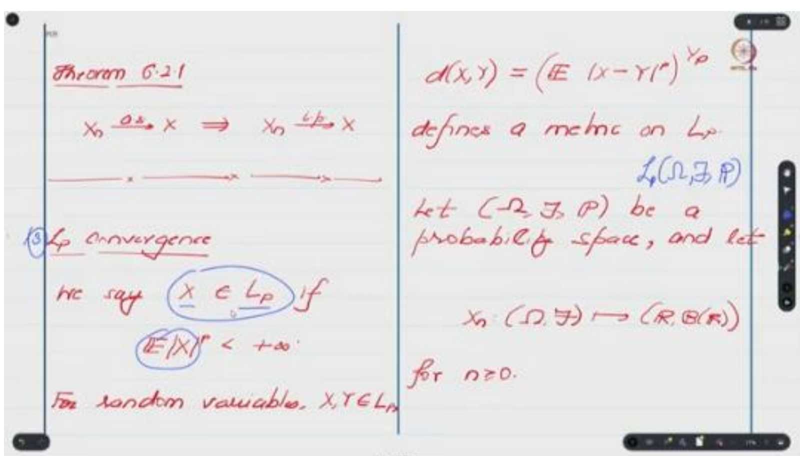
In particular, this result says that if you have a sequence converging almost surely to X, then that sequence also converges in probability. The converse direction is not true; that is, if you have a sequence of random variables converging in probability to X, that does not automatically imply that there is also almost sure convergence. So, the third notion of convergence that we will be talking about is that of Lp convergence. So, we will say a random variable X is in Lp. So, recall that we had said that X belongs to L1.

If the expected value of the absolute value of X is less than infinity, alright. So, the third notion of Lp convergence we will now discuss. So, towards that, we first need to define this concept of X belonging to Lp. We will say, so Lp again is a collection of random variables, and X is said to belong to Lp if it satisfies a condition like this. That is, the expected value of the pth power of the absolute value of X is less than infinity.

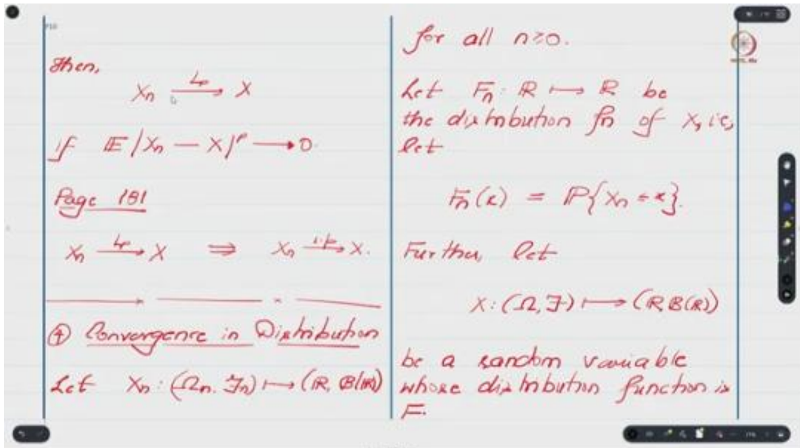


Now, in the initial few discussions, we have already seen what it means to say X belongs to L1. So, here we are going a step beyond and are now defining when X belongs to Lp for any p which is greater than or equal to 1. And so now Lp is a collection of random variables; for any two random variables X, Y which belong to this collection, we can now define a notion of distance between them. So, we will say the distance between X and Y in the Lp sense is the expected value of the pth power of the absolute value of X minus Y, the whole thing raised to 1 over p. So, one can show that this description actually defines a metric on the space of these random variables. And now, once this metric is defined, and let us say we have a collection of random variables—sorry, let us say we have a probability space and a collection of random variables Lp omega Fp.

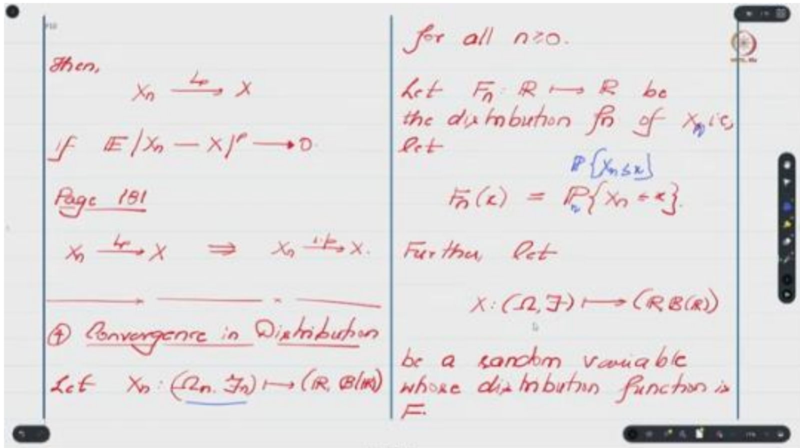
So, this Lp(Ω, F, P) is the collection of random variables for which the expected value of the pth power is less than infinity. Here, I ignored the probability space. So, here it was implicit; here I am expressing it explicitly. So, Lp(Ω, F, P) is basically the collection of all random variables whose expectation of the pth power is finite. And we will say a collection of random variables Xn.



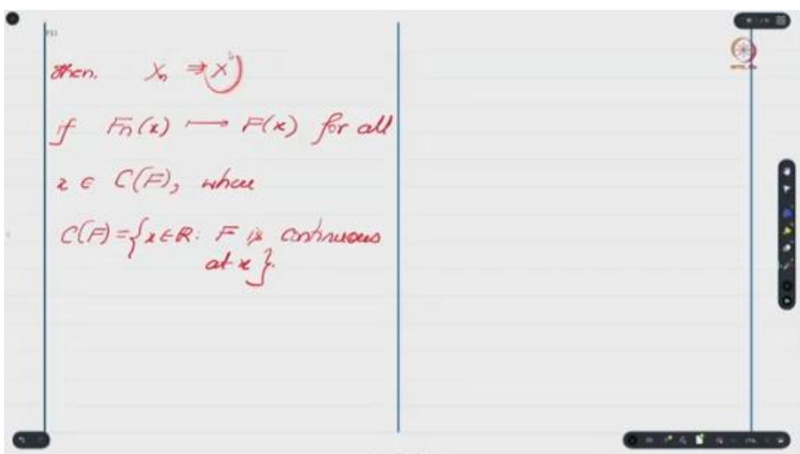
Again, notice that all these random variables are defined on the same probability space. It converges in the Lp sense to X if the distance between Xn and X in this metric sense converges to 0. And again, from this probability path textbook—in particular, if you look at page 181—one can show that Lp convergence actually implies convergence in probability. So, with this, we have covered three notions of convergence: almost sure, then in probability, and Lp. The key thing to note here is that all these three notions of convergence require that the random variables be all defined on the same probability space. Now, this condition is not necessary when we look at the fourth notion of convergence, which is that of convergence in distribution, right?



So, here, suppose you have a sequence of random variables which may be defined on different probability spaces, okay? So, these may be defined on different probability spaces, right? And let us say capital Fn is the distribution function of Xn. So, I think I should emphasize this over here: that capital Fn is the distribution function of Xn—that is, for every little x in ℝ, Fn of x— So, for every little x in ℝ, Fn of x is defined in this fashion. And again, you know, this can depend on n itself—this probability function—because these probability spaces differ with respect to n. The probability could also differ with respect to n. However, we will often write this in some generic sense as the probability of Xn less than x. So, suppose we define Fn in this fashion for every x in ℝ, and let us say there is another random variable X which is measurable with respect to some other probability space—let us say Ω and calligraphic F.



Then, we will let us say capital F is its distribution function, which is again defined in this sense but with respect to X. Then, we will say Xn converges in distribution to X if their corresponding distribution functions converge. In particular, we will say Xn converges in distribution to X if capital Fn of little x converges to capital F of X. For all those x's where your capital F function is continuous. So, in this way, we define this notion of convergence in distribution. So, this brings us to the end of this class.



I will give a quick summary now of what we have discussed. We first discussed the expectation of a random variable via the Lebesgue integration idea. Therein, we first described the expectation of a simple random variable, then we extended that definition to non-negative random variables, and finally, we defined the expectation of general random variables. Then we discussed several interesting properties of these random variables, such as monotonicity, the monotone convergence theorem, the dominated convergence theorem, Jensen's inequality, and linearity.

So, these five properties we discussed, and thereafter we discussed four notions of convergence, which included the almost sure convergence, convergence in probability, Lp convergence, and convergence in distribution. So, again, I have quickly gone over all these topics because the point of these discussions is to just give you familiarity with all these topics. So, that when we talk of convergence of a sequence of martingale difference terms, you will know what I am talking about. With this, let me stop. Thank you.