STOCHASTIC APPROXIMATION: THEORY AND APPLICATIONS

Dr. Gugan Thope

Department of Computer Science and Automation

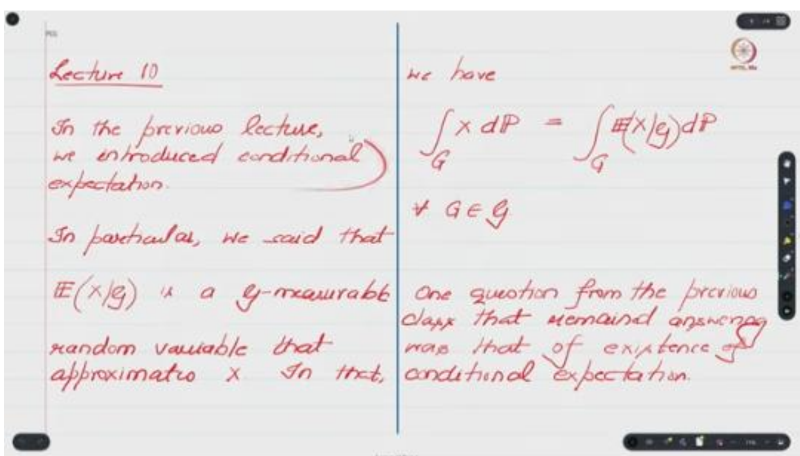
Indian Institute of Science

Lecture 10

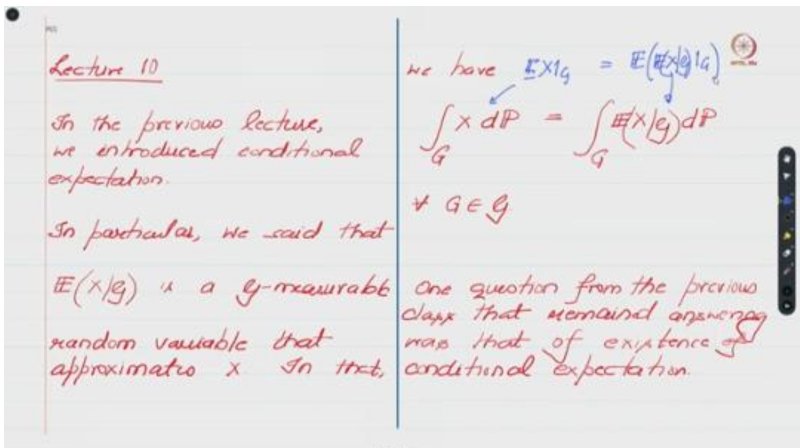
Properties of Conditional Expectation

Hello and Namaste everyone. Welcome to lecture 10 of this NPTEL course on Stochastic Approximation. So, just a quick recap of what we have been doing. This week, we have been covering some basics of measure-theoretic probability. In particular, we have been defining some concepts, making you familiar with certain notations and definitions.

So that we can talk about conditional expectations, and in the following week, we can talk about martingale differences and their convergence. Is that okay? So, let us do a formal recap of what we did in our previous class. So, in the previous class, we introduced the concept of conditional expectation. In particular, we said that if you have a random variable X and a sub-sigma-field G, then this conditional expectation is another random variable with the property that this new random variable is G-measurable.

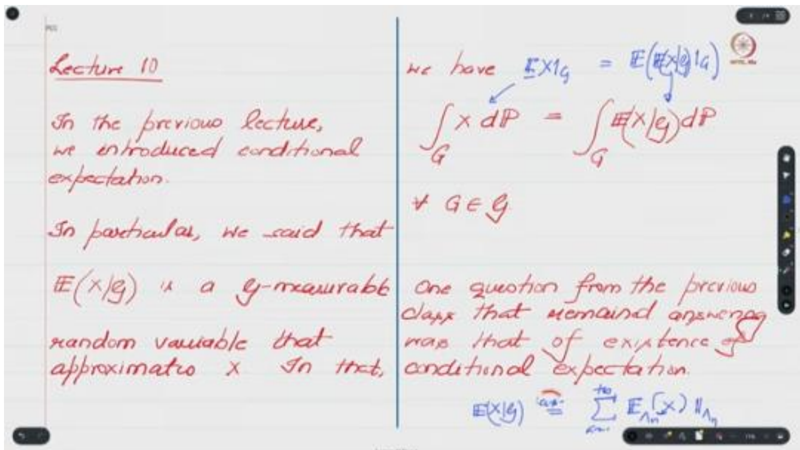


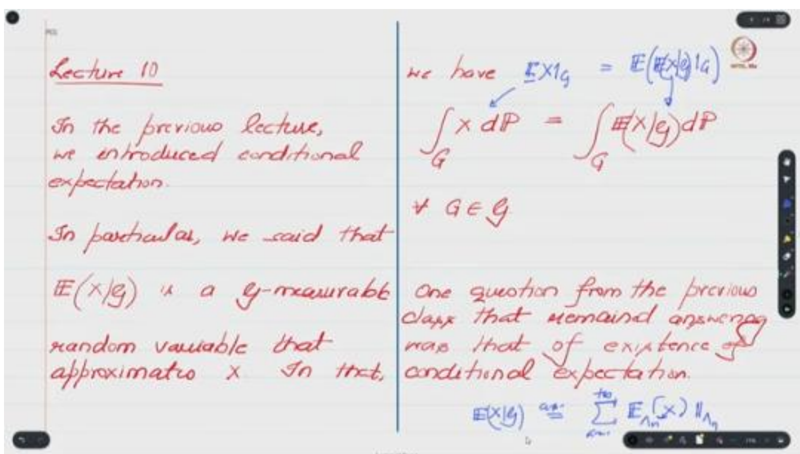
And it approximates X in a certain sense. In particular, we require that these two numbers be the same for every G in calligraphic G. So, what is the left-hand side? Well, the left-hand side, as we have described, is the expected value of X indicator G. This is the left-hand side, and the right-hand side is the expected value of this new random variable, the expected value of X given G times indicator G. And we require that this new random variable's expectation with respect to G, or the expectation of the product of these two random variables with respect to G, should be the same as the expectation of X times indicator G for every G in calligraphic G. So, this is what we mean by the conditional expectation approximating X. Whenever we say that, we mean precisely this: that instead of taking the expectation of X times the indicator of G, you can take the expectation of the conditional expectation times the indicator G, right?



So, these two things will be the same, and these two things will be the same as long as your G belongs to calligraphic G. So, this relation will not hold or need not hold outside this calligraphic G, okay? So, this is what we discussed last time, and you know, I had posed two questions with regards to this definition. One was, you know, what is the intuition behind this definition, and we had, you know, looked at a very simple case and used that to illustrate what this conditional expectation is talking about, and so on and so forth. And there was one thing that I forgot to explain in the previous class. I had said that, you know— sorry, not this but— I had said that, you know, your conditional expectation in that special case, this thing equaled almost surely this infinite sum that we have. So, in that example that we had discussed, I had indicated that this almost sure equality holds. So, let me first recall the relation that we had written.

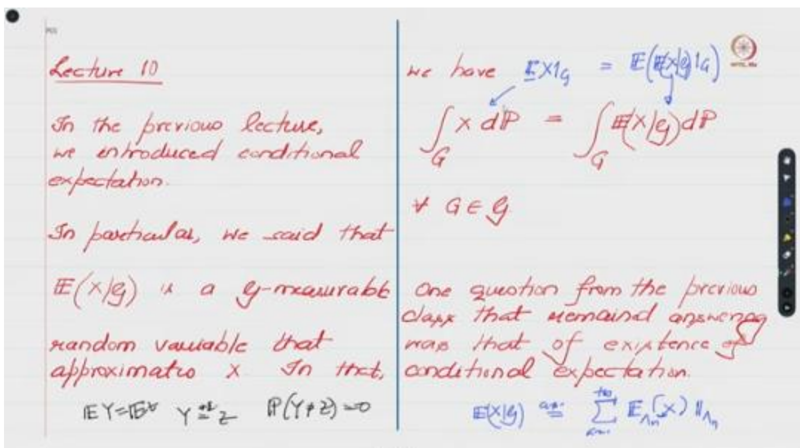
Now, the question is, why did I write almost surely equal? Well, you have one random variable over here and one random variable over here, and you know, if you look through, you know, this probability part textbook by Resnick, you will see that if you have two random variables, let us say Y and Z, which are almost surely the same, which means that the probability Y not equals Z is 0. So, let us say you have two random variables Y and Z, and we will say that Y almost surely equals Z if a probability condition like this holds— that is, the probability that Y not equals Z equals 0. Now, from the properties of expectation, one can show that even though Y and Z, right, can be different on a 0-measure set, their expectations— will be the same.

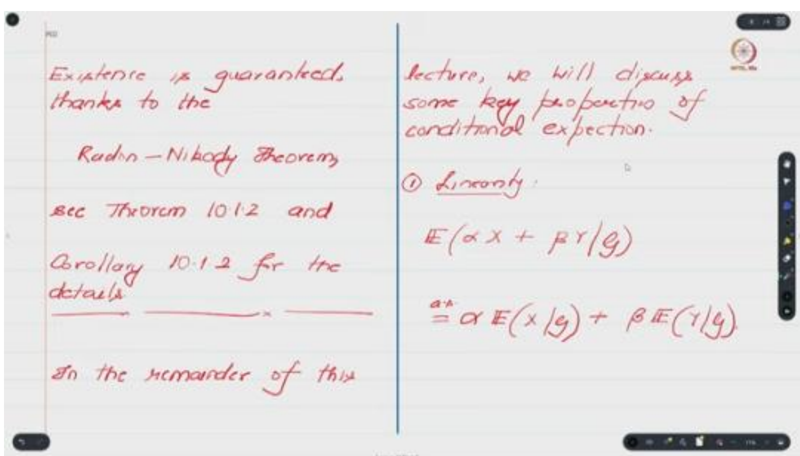




So, the expected value of Y equals the expected value of Z. So, coming back over here, why did I write almost sure here? It basically means that except for a set of measure 0, these two things could be different, but whenever you exclude this 0-measure set, these two functions should be identically the same. And why do we insist on almost sure? The reason is just that, you know, someone can, you know, change this right-hand side on a 0-measure set and conclude that, oh, this is not equal to what you have written over here. So, this is just to safeguard against such a discussion. It is to ensure that you can change this right-hand side over a set of measure 0, but apart from that, no one would be able to change this beyond it. So, this is something that I think I forgot discussing in the previous class.

There was another question that remained unanswered from the previous class, which is that of why such a conditional expectation should exist. So, conditional expectation, recall, is a random variable which is g-measurable and satisfies these two properties. So, why does such a random variable—basically this thing—exist? So, while I would not be able to give full details, I can give enough details so that if you are interested in answering such questions, you have references to things you can read and study on your own to figure out the answer. So, the existence of such random variables is guaranteed thanks to this theorem called the Radon-Nikodym theorem.





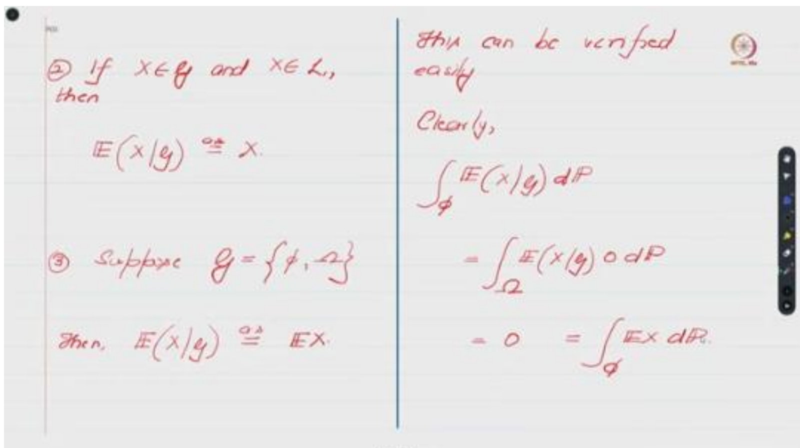
I should write M over here—Radon-Nikodym theorem—and you can look at Theorem 10.1.2 and Corollary 10.1.2 in this textbook called A Probability Path by Resnick. So, if you look at those theorems, you can indeed get access to the details of this Radon-Nikodym theorem and, in particular, how this theorem guarantees the existence of this conditional expectation, right? So, if this were a course on probability, I would have given you full details. However, in this stochastic approximation course, I am just trying to give a quick overview of these concepts. I am only providing you a reference, and I encourage you to look it up if you are interested in knowing details about this existence.

So, in the remaining part of this lecture, we will discuss some key properties of conditional expectation that we will repeatedly use throughout our discussion on stochastic approximation. So, the first of these properties is the linearity property. And I would like to emphasize that for a small subset of these properties, I will tell you how to prove them, and in some cases, I will in fact prove them. But for most, I would not have the time to provide exact details of the proof, and I encourage you to look at Chapter 10 of this textbook called A Probability Path by Resnick. So, he would have provided you with all the details. So, the first property that we want to look at is that of linearity.

So, what does the linearity property say? It says that suppose you have two random variables X and Y; let us look at their linear combinations. That is, alpha X plus beta Y, where alpha and beta are real numbers, and let us look at the conditional expectation of this random variable. The linearity property says that the conditional expectation of this property would equal alpha times the conditional expectation of X and beta times the conditional expectation of Y. Again, this equality holds in an almost sure sense—apart from a set of measure 0, you cannot change this right-hand side.

$$\mathbb{E}\left(\alpha X + \beta Y|_{g}\right) = \alpha\mathbb{E}\left(X|_{g}\right) + \beta\mathbb{E}\left(Y|_{g}\right)$$

So, I will not be giving you the proof, but the proof is very straightforward; it follows from the linearity property of expectation that we had stated before. So, you just invoke that property and verify the two conditions that were there in the conditional expectation definition, which is that the right-hand side should be G-measurable and if you take the product of this right-hand side with indicator G and look at its expectation, then that expectation should equal the expected value of this alpha X plus beta Y random variable times indicator G. So, whenever these two expectations hold, that will follow from the linearity of expectation, and one can conclude in that case that this indeed plays the role of a conditional expectation. So, the next property, while it is very trivial to state, is often very important. What this says is that suppose you have some random variable X and you have been given a sigma-field G, and suppose it turns out that the given random variable is measurable with respect to G. So, the given random variable itself is measurable with respect to G. In that case, the conditional expectation of X with respect to this sigma-field G is basically X itself, apart from a set of measure 0.

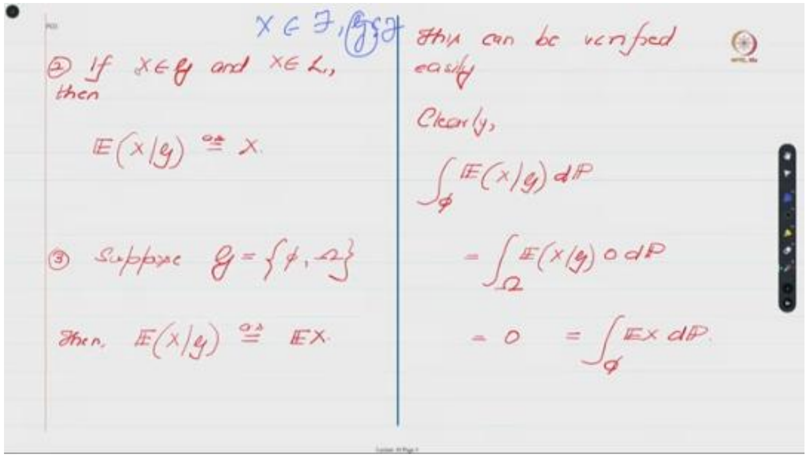


So, that is why I have written 'a.s.' over here, and this makes sense.

$$\mathbb{E}(X|_g) = X$$

What is conditional expectation? Well, it is the best G-measurable random variable. Now, if X itself is G-measurable, then that will be the best proxy for X. And I would like to distinguish this case from what we have seen so far.

And what we have seen so far is that we had some parent sigma field F, and X was measurable with respect to F. Your G was basically a sub-sigma field defined with respect to F, and we were looking at the conditional expectation of X with respect to this G. And in this special case, what we are considering is that the given random variable X is measurable actually with respect to G itself. Is this okay? So, of course, since G is a subset of calligraphic F, if you are measurable with respect to G, you automatically become measurable with respect to F. But measurability with respect to F need not imply that X is also measurable with respect to G. So, in that sense, this is a special case.

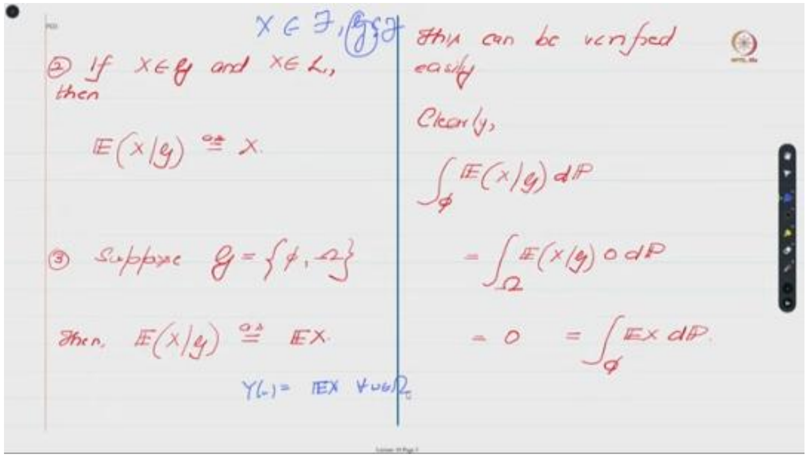


So, if you have measurability with respect to G, then one can very easily check that the expected value of X given G is actually X itself. So, now let us move on to the third property. The third property looks at the case where the sigma field, with respect to which you want to define the conditional expectation, is trivial in that it only includes the empty set and the sample space omega. And in this case, the claim is that the conditional expectation of X with respect to such a G will actually be a constant, and it will be the actual expectation of X itself. So, when you take the conditional expectation of X with respect to G, it is a constant expected value of X. And since this is a very easy proof, I will verify this over here so that you know how to verify some of these results.

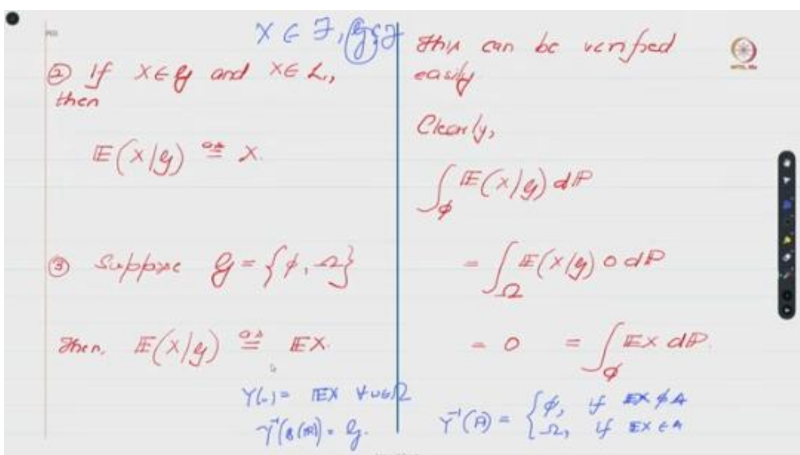
$$\mathbb{E}(X|_g) = \mathbb{E}X$$

Now, the first thing we need to verify is that this right-hand side over here is G-measurable.

Now, let us define a function Y which is equal to the expected value of X for all omega in capital omega. So, because of this definition, one can see that Y inverse of any set A, which is a subset of R, this will either be the empty set or omega. Is that okay? And when it will be the empty set if your set—okay, or let me write it the other way—the expected value of X, which is a real number, does not belong to A. On the other hand, if the expected value of X belongs to A, then Y inverse of A will be omega.



And in this sense, you can see that this constant thing that we have on the right-hand side is indeed G-measurable because the inverse of any A is either the empty set or omega. In particular, we can conclude that Y inverse of B of R is actually G, which is what is given over here. So, this is G-measurable. So, we have verified the first condition in the definition of the conditional expectation. Next, we have to verify that for every set G in this calligraphic G collection, the suitable expectations take the same value. So, let us check that.



So, for the case where G is the empty set. So, for the case where G is empty. In that case, we have to take the conditional expectation, and we are claiming that this conditional expectation is actually the expected value of X. So, we have to check that this thing over the empty set actually equals the integral of X dP over the empty set. So, we have to verify this, and again, as I told you, the right-hand side is actually the expected value of X times the indicator of this set, but since this is empty here, right, this is actually 0.

So, X times 0 is actually 0, and the expected value of 0 is basically 0. So, this can be checked via this definition of the simple function, the expectation of a simple function that we had looked at. In particular, the 0 can be viewed as 0 times the indicator omega. So, the expectation of this quantity will basically be 0 times the probability of omega, which is 0. And now, we have shown that this quantity over here, when you have the empty set over here, is 0. In the same way, this quantity can also be shown to be 0, and since both the left-hand side and the right-hand side are equal. We can conclude that at least for the case where G is empty, these two things are the same. One can similarly verify that this quantity and this quantity are the same when this empty set is replaced by omega, which is the second element in your calligraphic G, and then conclude that indeed these two things are one and the same.

$$\int_{\phi} \mathbb{E}(X|_g) d\mathbb{P} = \int_{\Omega} \mathbb{E}(X|_g)\, 0 d\mathbb{P}$$

$$= 0 = \int_{\phi} \mathbb{E}\, X d\mathbb{P}$$

$$\int_{\phi} \mathbb{E} X d\mathbb{P} = \int_{\phi} X d\mathbb{P}$$

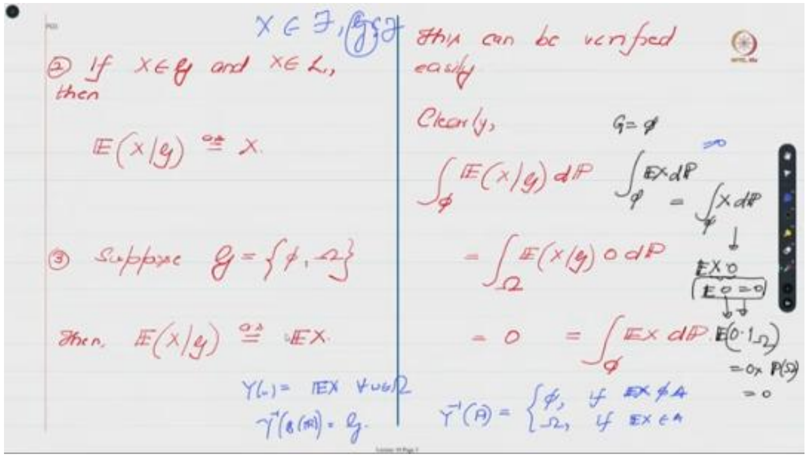
$$= \mathbb{E} \cdot 0$$

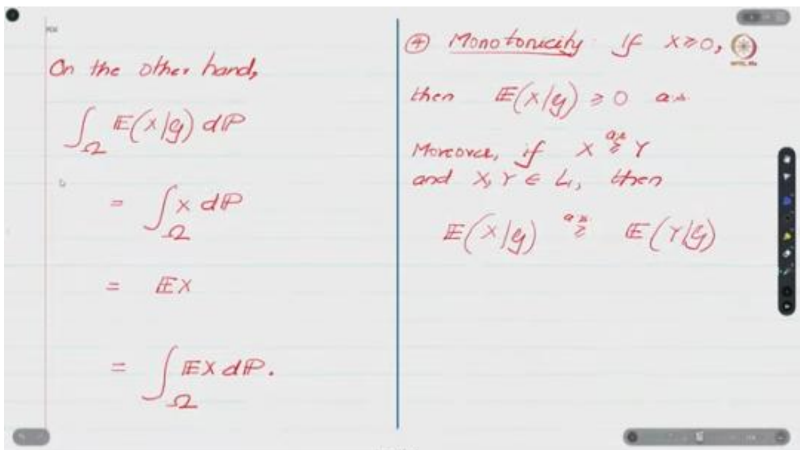
$$\mathbb{E}0 = 0$$

$$\mathbb{E}(0.1\Omega) = 0 \times \mathbb{P}(\Omega)$$

$$= 0$$

So, we verified first that this expected value of x is measurable with respect to g and for a set in your G, one can actually conclude that these two things are one and the same. So, this is what we have verified here for the case where you have omega. So, let me just quickly go through it. We require that the conditional expectation of X with respect to G must equal this expression because this is how we defined the conditional expectation.





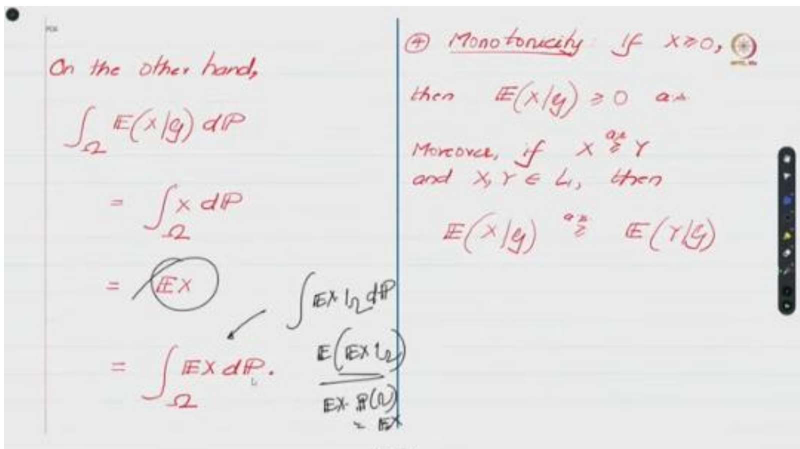
So, this is what we have. And by definition, this is the expected value of X, and you know the expected value of X can be written this way. Using our definition of expectations of simple random variables, this is actually the expected value of X indicator omega, right? dP or, in other words, this is the expected value of the expected value of X indicator omega, and this quantity is basically the expected value of X times the probability of omega, which is the expected value of X as desired over here, right? So, we have verified this, and hence one can conclude that when you are given a sigma subfield made up of only the empty set and the whole sample space, then the conditional expectation is actually the true expectation, right? OK.

$$\int_{\Omega} \mathbb{E}(X|_{g})\, d\mathbb{P} = \int_{\Omega} X d\mathbb{P}$$

$$= \mathbb{E}X$$

$$= \int_{\Omega} \mathbb{E}X d\mathbb{P}$$

There is this very important property of monotonicity, which says that if your given random variable X is non-negative, then the conditional expectation will also be non-negative, right?



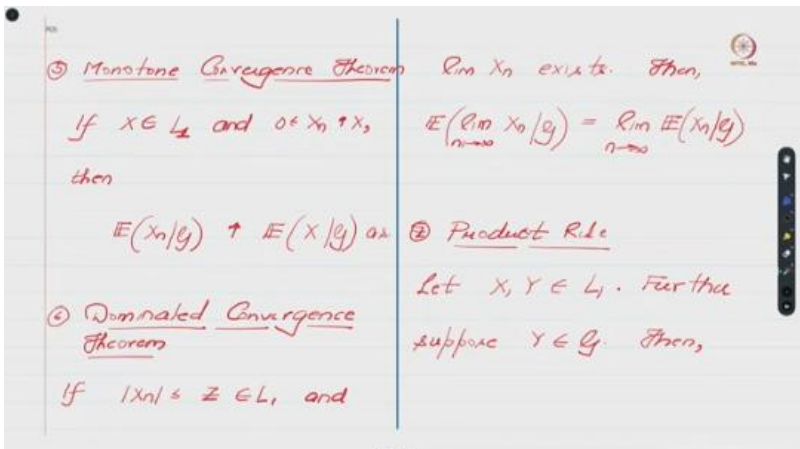
So, recall that this is a random variable, and we are saying that this is non-negative, which means that if you look at every omega in capital omega, this left-hand side has to be non-negative. However, as I told you, if two random variables differ only on a zero-measure set, then their expectations are the same. Hence, it is possible that even though this is non-negative for every little omega in capital omega, the conditional expectation may be non-negative on every omega in capital omega except for a zero-measure set. So, that is why this almost sure acronym is added to this statement over here.

$$\mathbb{E}(X|_g) \geq 0$$

$$\mathbb{E}(X|_g) \geq \mathbb{E}(Y|_g)$$

And this monotonicity idea can actually be generalized to the case where X is almost surely bigger than Y, and suppose that X and Y are in L1 so that the conditional expectations are well-defined.

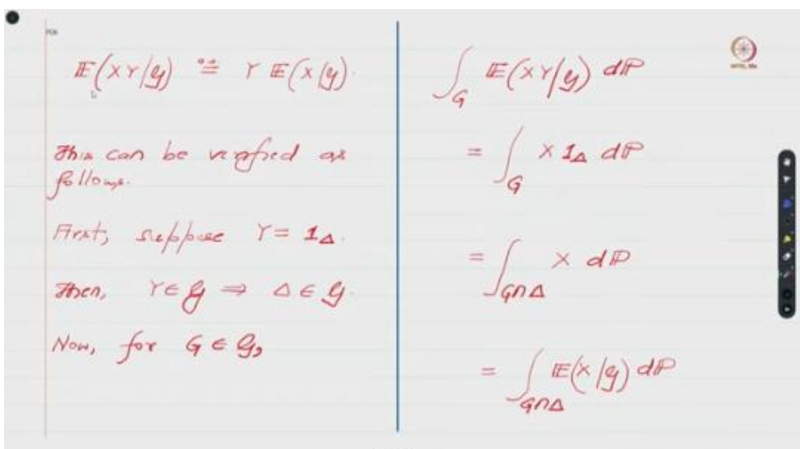
And the monotonicity property says that if X is bigger than or equal to Y, then the conditional expectations also satisfy some relation like this in an almost sure sense. So, we have seen some monotone convergence theorem for expectations. A similar monotone convergence theorem also exists for conditional expectations, and it is what we would expect. That is, suppose you have an L1 measurable random variable X, and in addition, suppose you have non-negative random variables Xn which monotonically increase and converge to X. Then, the conditional expectations also monotonically increase and converge to the conditional expectation of X given G. Of course, all of this holds in an almost sure sense. So, just as we have a similar extension of the monotone convergence theorem, we also have an extension of the dominated convergence theorem.



So, the dominated convergence theorem says that suppose you have a sequence Xn such that every element in that sequence—that is, Xn—satisfies this relation: the absolute value of Xn is less than or equal to Z, and this should hold for every little omega in capital Omega. Furthermore, suppose that this Z itself is L1 integrable, that is, the expected value of Z is less than infinity. Right? And because of this condition, we trivially get that each Xn is L1 integrable. Furthermore, suppose that your Xn converges. So, unlike the monotone convergence theorem where we had Xn monotonically increasing, here we are not insisting that Xn monotonically increases. Here, we are just saying that Xn converges.

So, it is possible that whatever the limit is, your Xn may go up and down for different values of omega. But we are only insisting that the limit exists, and in that case, again, the limit and the conditional expectation operation can be interchanged, just like in the monotone convergence theorem. So, the dominated convergence theorem says that if you have a dominating random variable, then the conditional expectation of the limit equals the limit of the conditional expectation. So, such a statement also holds when you have something like that—that is, the monotone nature—and in that case, this relation is referred to as the monotone convergence theorem. So, the next key property that we would be relying on is that of the product rule.

And this property, since it is very important and we will often be relying on it, I will prove this very explicitly and, you know, go over it a bit slowly. I will also tell you, or highlight, the advantage of working with this measure-theoretic perspective. So, what does the product rule say? Suppose X and Y are two random variables, and let us say they are both integrable, and suppose one of these random variables, in this case Y, is G-measurable. And the question is, given that Y is G-measurable, what can we say about the conditional expectation of the product? In one of the properties that we studied before, we did not have Y, and we showed that if X itself is G-measurable, then the conditional expectation of X given G was X.



In this case, we are having two random variables, and we are only saying that one of those random variables is G-measurable. The question that we are interested in asking is, what can we say about the conditional expectation of the product of X and Y? The claim is that this conditional expectation is actually Y times the conditional expectation of X given G. That is, whenever Y is G-measurable, you can pull Y out of this conditioning and write it outside. Is this okay?

$$\mathbb{E}(XY|_g) = Y \in (X|_g)$$

So, let us verify this property. So, the way we verify this, and in general the standard way of verifying several of the results in measure-theoretic probability, is to follow the following recipe. Whatever is desired to be shown, you first show it for an indicator random variable, then you extend it to simple random variables, then you extend the proof to non-negative random variables, and finally, you extend the idea to general random variables.

So, let us, you know, follow this recipe. So, the goal is to show this, and towards that, what we will do is we will presume that Y has a special form, that is, Y equals indicator delta, where this delta belongs to G. So, by ensuring that delta belongs to G, we are ensuring that Y is measurable with respect to G, as required in the statement of this claim. So, now we have to show that if Y has this special form, then this condition actually holds. So, let us verify that.

So, the first thing that we need to verify is that this product is actually G-measurable right now because delta belongs to G, Y is G-measurable, and this conditional expectation by definition is G-measurable. So, we have a product of two G-measurable functions. And hence, the product is also G-measurable. So, these are things that one can show using some of the discussions, you know, random variables from Chapter 5 of this Probability Path textbook. So, now it remains to verify whether if you take the expectation of this times indicator G or whether you take the expectation of this times indicator G, they remain the same.

So, let us verify that. So, the left-hand side, that is the expected value of X, Y given G, when it is integrated with respect to the specific G in calligraphic G. By its definition, because this is a conditional probability, this must equal the integral of X, Y indicator dP G. So, this is exactly what I have written over here by substituting that Y equals indicator delta. Is that OK? Now, because you have indicator delta, this delta can be moved here by, you know, introducing an intersection. Hence, this quantity equals the integral of X

dP with respect to G intersection delta. In other words, this is basically the expected value of X indicator G intersection delta. But G belongs to calligraphic G, and delta belongs to calligraphic G. And because this calligraphic G is a sigma-field, one can conclude that G intersection delta also belongs to calligraphic G. Now, since G intersection delta also belongs to calligraphic G, this integral that we have will actually equal this integral from the definition of the conditional expectation of X given G.

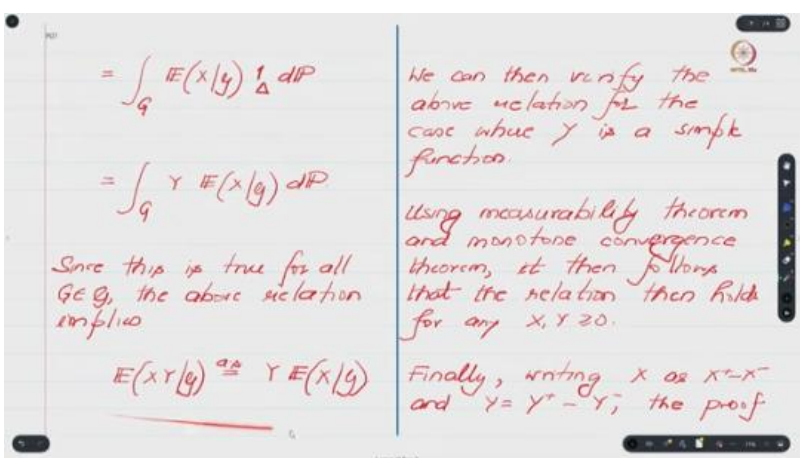
$$\int_G \mathbb{E}(XY|_g)d\mathbb{P} = \int_G X1_\Delta d\mathbb{P}$$

$$= \int_{G\cap\Delta} Xd\mathbb{P}$$

$$= \int_{G\cap\Delta} \mathbb{E}(X|_g)d\mathbb{P}$$

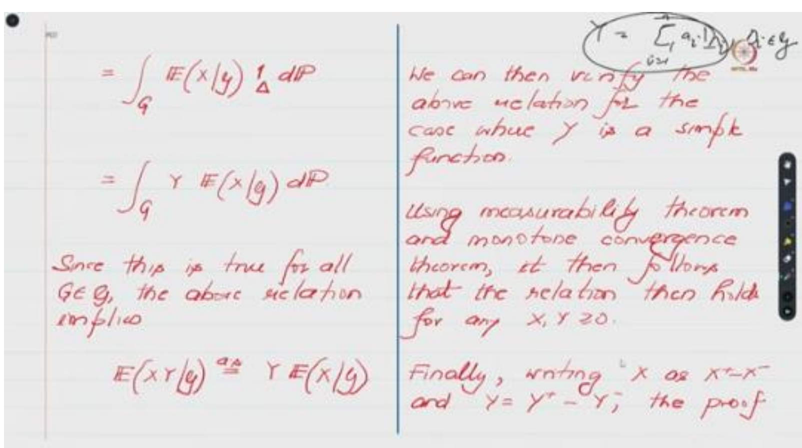
This just follows from the fact that, you know, if you have a conditional expectation, then this integral must equal this integral for every set in calligraphic G, and since we have shown that G intersection delta belongs to calligraphic G, these two things must indeed be the same.

So, now what we can do is whatever was there in G intersection delta, that intersection delta, we can again move it above in this sense. So, we can write it as indicator delta, and I have just written it here as y, and hence we have concluded that, you know, this expectation that we have here equals this thing. Right. And since this g was arbitrary, one can conclude that indeed for the special case where y equals indicator delta, it indeed holds that these two things are one and the same, right. And the way we verified this relation, we can use the same proof technique to conclude that, you know, this relationship also holds when y is actually a simple function, that is, y is

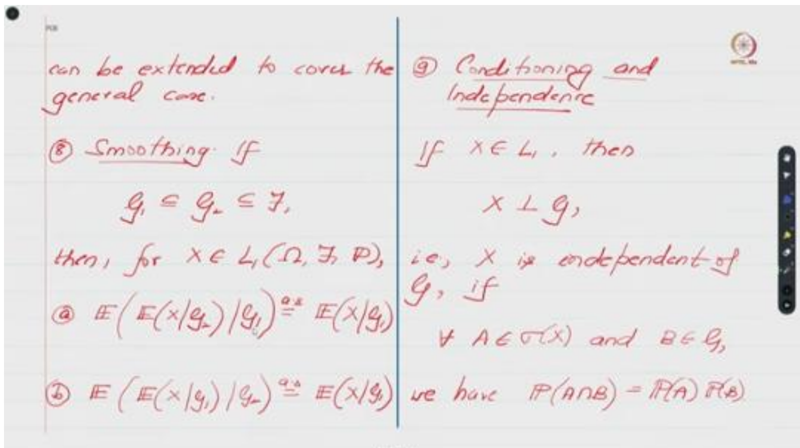


summation i equals 1 to n a i indicator lambda i or delta i, where delta i belongs to g. So, if y had this form, then one can conclude that such a relation also exists just by following the same proof. And then we can, you know, invoke the measurability theorem and monotone convergence theorem. to show that if X and Y now are more general, that is, X and Y no longer have any special form, but rather they are some general random variables which are non-negative, right. So, then, you know, by measurability theorem, one can show that there is a sequence of simple functions that converges to Y, and from the monotone convergence theorem for conditional expectations, one can then conclude that this also holds. And finally, we can extend whatever proof that we have discussed here to a more general scenario where both x and y need not be non-negative.

In that case, you know, we just use this following idea that we can express x as x plus minus x minus, and this x plus and x minus are non-negative random variables. Similarly, y plus and y minus are non-negative random variables. Hence, if you look at x plus times y plus, x plus times y minus, and similarly x minus ys and x minus y inverse, y inverse, sorry, y minus, sorry, right. So, all those four products are non-negative, and hence whatever we have proved over here will apply to that setting, and you know one can show that in that case, you know, this property also holds.



So, this is something that is, in some sense, a standard recipe for working with conditional expectations—for proving properties related to conditional expectations. In this case, I sort of went over as many details as possible to verify one of these properties. So, the next property, which is again a very key property that we will repeatedly use with respect to conditional expectation, is that of smoothing. So, what does the smoothing property say? It says that suppose you have two sub-sigma fields—in particular, let us say you have G1, G2, and F. Such that G1 is a sigma field, G2 is a sigma field, and F is a sigma field, and in addition, G1 is a subset of G2, which is a subset of F. Right?



And now, suppose that X is an integrable random variable that has been given to you. And now, if you do an iterated conditioning—that is, you first take the conditional expectation of X with respect to G2. So, recall that G2 sits over here, and then this is a random variable, and you take its conditional expectation with respect to G1. So, the smoothing property says that if you take this iterated conditioning, then that must equal the conditional expectation of X with respect to this smaller sigma field. So, this thing over here vanishes, and of course, this equality holds only in an almost sure sense.

If you take this conditioning in this sense—that is, you first take the conditional expectation with respect to G1 and then take whatever you get, its conditional expectation with respect to G2. Now, this second property is very easy to show because this conditional expectation is actually G1-measurable. And because it is G1-measurable and G1 is a subset of G2, we get it for free that this is also G2-measurable. And since this random variable is G2-measurable, when we take the conditional expectation of this random variable with respect to G2, we get back this.

$$g_1 \subseteq g_2 \subseteq \mathcal{F},$$

$$X \in L, (\Omega, \mathcal{F}, \mathbb{P}),$$

$$\mathbb{E}\left(\mathbb{E}\left(X|_{g_2}\right)|_{g_1}\right) = \mathbb{E}\left(X|_{g_1}\right)$$

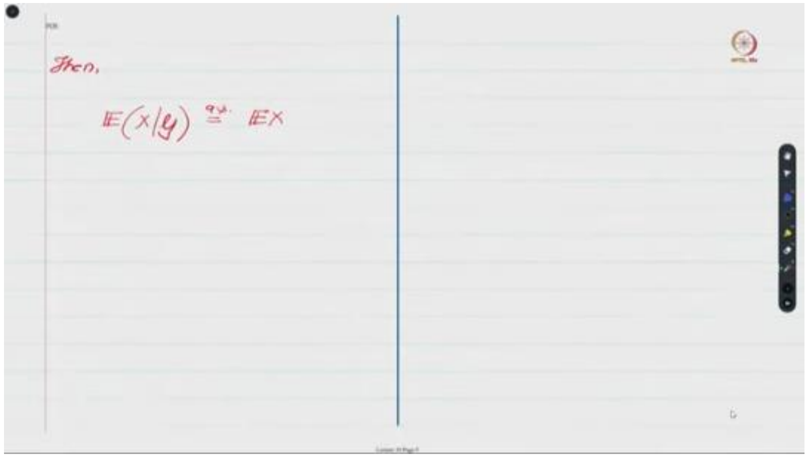
$$\mathbb{E}\left(\mathbb{E}(X|g_1)|_{g_2}\right) = \mathbb{E}\left(X|_{g_1}\right)$$

So, the summary is that whenever you have G1 and G2, which satisfy some relationship like this, then if you do this iterated conditioning, that will equal the conditional expectation with respect to the smaller sigma field.

And we now move on to the last property that we will repeatedly use in the discussion of conditional expectations throughout this course, which is that of conditioning and independence. So, first, let me tell you what independence means. In this lecture, I am going to define independence of random variables with respect to a sigma field. So, what do we have here? We have an X, and I want to discuss when X is independent with respect to a sigma field. I will use this notation to denote this independence. So, we have a random variable X, a sigma field G, and we want to say when X is independent with respect to this sigma field.

So, we will say X is independent of this sigma field if, for every A in the sigma field generated by X and every B in G, we have this relation. So, this relation is something very familiar to us. This is the definition for two events being independent. Now, we are using this definition to discuss when a random variable X is independent with respect to a sigma field.

So, X is independent of G if, for every A in the sigma field generated by X and every B in the sigma field generated by G, we have that A and B are independent in the usual sense. So, in this case, one can show that the conditional expectation of X with respect to G actually equals the expected value of X itself. Is this okay? Because X is independent with respect to G, the best proxy for X, which is G-measurable, is actually the expected value of X itself. So, with this, we come to the end of this class.



So, let us quickly summarize what we have done throughout this week. This week, we have quickly gone over this entire textbook on probability path. In particular, we have covered measure-theoretic probability, which is one of the prerequisites of this course, and I wanted to quickly brush up on some of the ideas and concepts that we will require throughout this course. Since I wanted to review, I have gone over it very quickly. In subsequent classes, we will cover the material more slowly, and in particular, we will prove all the results in significant detail. So, having said that, let me now summarize what we have done this week. We looked at probability spaces, then we examined random variables, and then we studied the expectations of random variables.

In particular, we defined the expectation of a simple random variable, then generalized it to non-negative random variables, and so on. Then we discussed the properties of this expectation, and in the last two lectures, we introduced the notion of conditional expectation. In this class, we have covered some properties of conditional expectation. Next week, we will discuss martingale differences and the conditions under which we can show their convergence. Having said that, thank you, namaste, and hopefully, you will join me for the third week soon. Thank you.