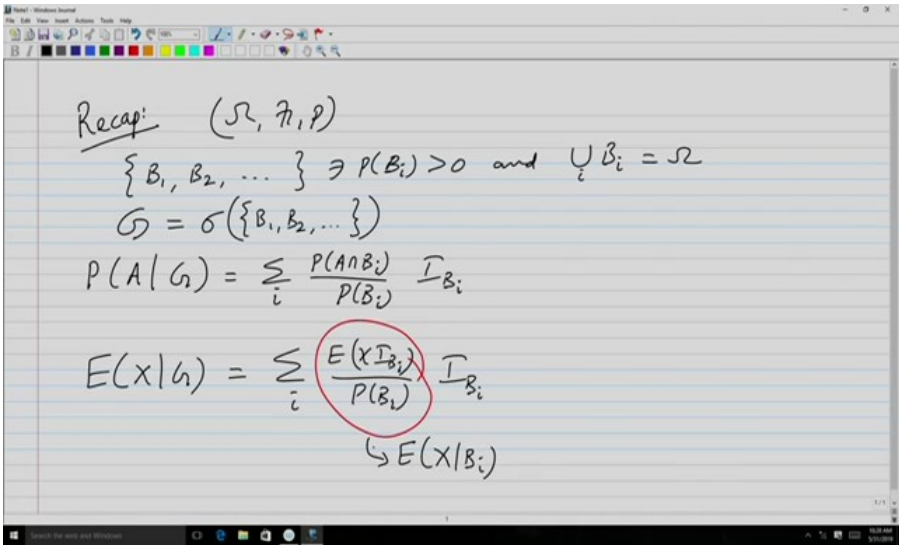
Mathematical Finance
Risk-Neutral Pricing in Discrete-Time
Lecture 21: Properties of Conditional Expectation

Professor N. Selvaraju[1] and Professor Siddhartha Pratim Chakrabarty[1]

[1]*Department of Mathematics, Indian Institute of Technology Guwahati, India*

Hello everyone. In the last lecture we have seen we have ended with the definition of conditional expectations. So let us just briefly recap what we have said.

(Refer Slide Time: 00:43)



So, there is a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and there is a sequence or a collection of sets $\{B_1, B_2, \ldots\}$ which are either finite or countable in nature such that each of them have positive probability and $\bigcup_i B_i = \Omega$ and the $\mathcal{G}$ is the sigma field generated by this collection, .i.e, $\mathcal{G} = \sigma(\{B_1, B_2, \ldots\})$. The conditional probability of A given $\mathcal{G}$ is given by

$$\mathsf{P}(X|\mathcal{G}) = \sum_i \frac{\mathsf{P}(A \cap B_i)}{\mathsf{P}(B_i)} I_{B_i}$$

On the set B_i this will take $\mathsf{P}(A|B_i)$. So, remember this quantity which we are written inside the sum, $\frac{\mathsf{P}(A\cap B_i)}{\mathsf{P}(B_i)}$ is same as $\mathsf{P}(A|B_i)$.

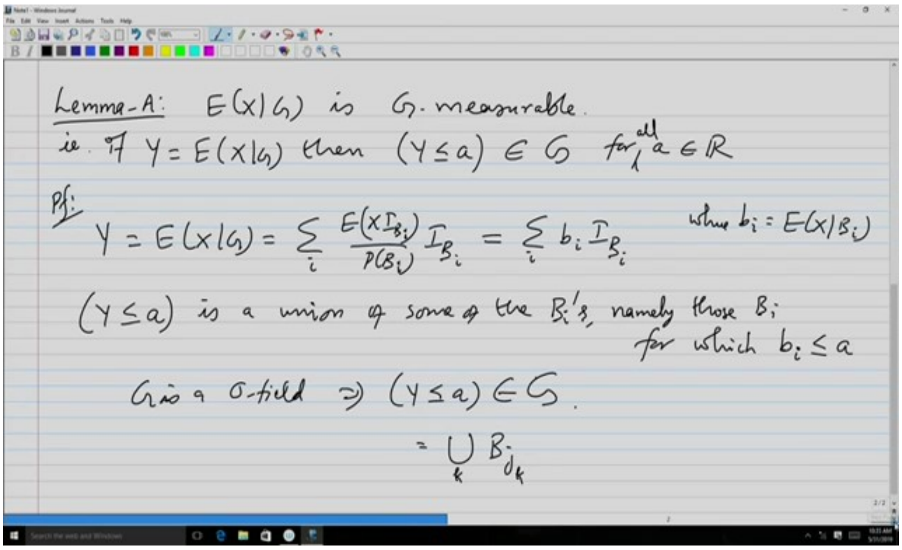
Extending the definition of $E(X|A)$ to $E(X|\mathcal{G})$ along similar line, we define for a random variable X, which is defined on this probability space $(\Omega\mathcal{F}, \mathsf{P})$. $\mathcal{G}$ is obviously you can see that it is a subspace of this $\mathcal{F}$, it could be even $\mathcal{F}$ or it is a proper subspace of this $\mathcal{F}$.

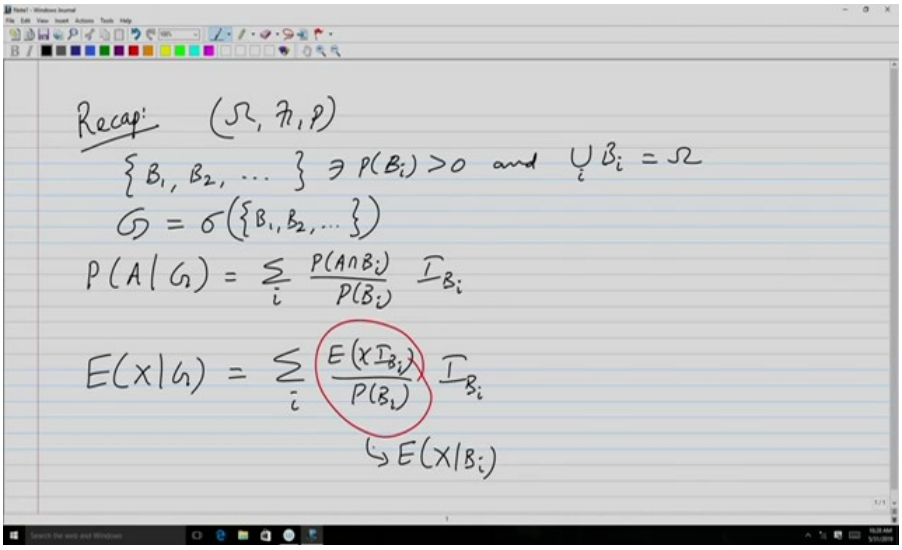
And we defined by extending the definition $E(X|\mathcal{G})$ to be this

$$E(X|\mathcal{G}) = \sum_i \frac{E(XI_{B_i})}{\mathsf{P}(B_i)} I_{B_i}$$
$$= \sum_i E(X|B_i) I_{B_i}$$

Now, keeping this definition in mind now, let us go through some of the properties that this conditional expectation would satisfy. They are a bit technical in nature. Proper understanding of that properties will help, in a way when we move further into the course. So, this understanding and the comfortness that we are going to gain with respect to these properties would play a major role. So, and it is not so difficult to understand, so you will be able to see what is going on, when we look at the properties.

(Refer Slide Time: 04:38)





So, the first property you know, which let me call as a lemma A. We are calling now this particular quantity $E(X|\mathcal{G})$ as a random variable, because it takes one particular value which is this $E(X|B_i)$ is a simple value and this is the indicator function I_{B_i}. So, on the sets B_i, it takes this particular value, $E(X|B_i)$, is what this random variable is and when we recall the examples that we have set.

So, our first property is that $E(X|\mathcal{G})$ is $\mathcal{G}$ measurable, what does that mean? That is if I denote, so if I denote this quantity as some $Y = E(X|\mathcal{G})$ then recall the definition of measurability . So, this $(Y \leq a)$ is a set in $\mathcal{G}$ for each real number a.

So, this is what we mean when we say that this is some random variable that is measurable with respect to some sigma field, so this is what first claim. So, let us look at why and how this is true. So, we recall by definition this $\mathcal{G}$ is again, you know, the one which is going to be generated by those class of sets B_1, B_2 and so on. They could be either finite in nature or it could be countable in nature.

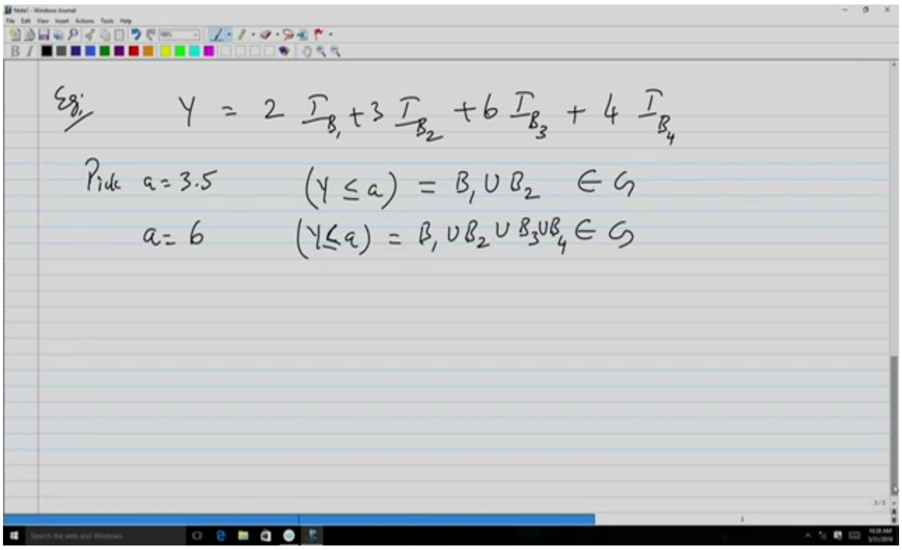
For this idea if may not work or will not work when we extend this to beyond that level when you have an uncountable many sets but as long as you have sets of that nature this idea will work.

So, that we need to worry when we move to continuous time or general probability spaces which we will revisit that time. Currently, now what we see is by definition, this is nothing but then, so this let us call since this particular quantity $Y = E(X|\mathcal{G})$is $\sum_i b_i I_{B_i}$ where my $b_i = E(X|B_i)$.

Now, if I look at this set, $(Y \leq a)$, because I need to show that this belongs to $\mathcal{G}$, this is nothing but, is union of sum of the B_i's namely those B_i for which this $b_i \leq a$. If this particular set suppose if I could able to write as; so this $(Y \leq a)$ is nothing but union of sum of the B_i's namely those B_i's for which this $b_i \leq a$.

So, it will include all those $Bi's$. But $\mathcal{G}$ is a sigma field which is closed under countable union. Since G is a sigma field, it implies then this Y is also belonging to this which is nothing but union of sum B_{j_i}. So, this is essentially if I have to write it has to write $\underset{k}{\cup} B_{j_k}$. Since G is sigma field this is done. So, you could see that this is a G measurable random variable.

(Refer Slide Time: 11:00)



Now, let us look at an example which will help. Suppose, if I could write this Y as,

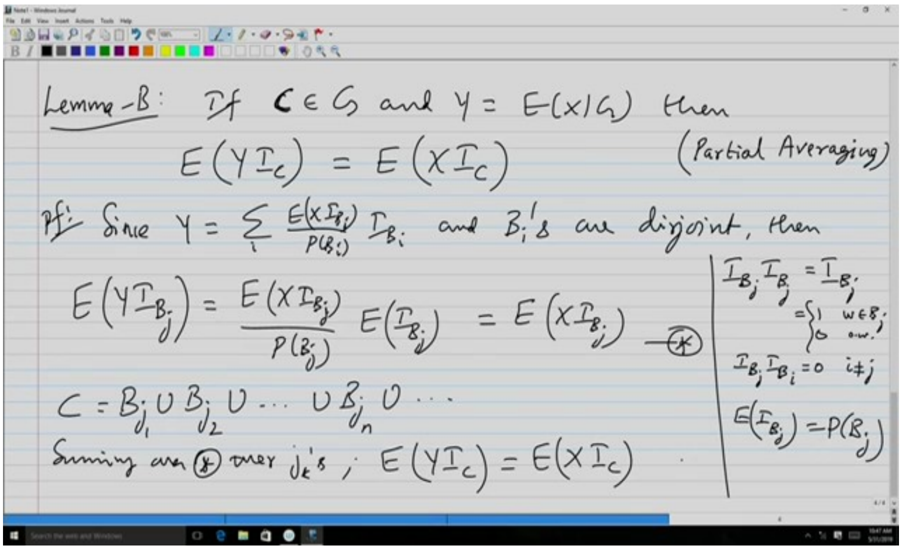
$$Y = 2I_{B_1} + 3I_{B_2} + bI_{B_3} + 4I_{B_4}.$$

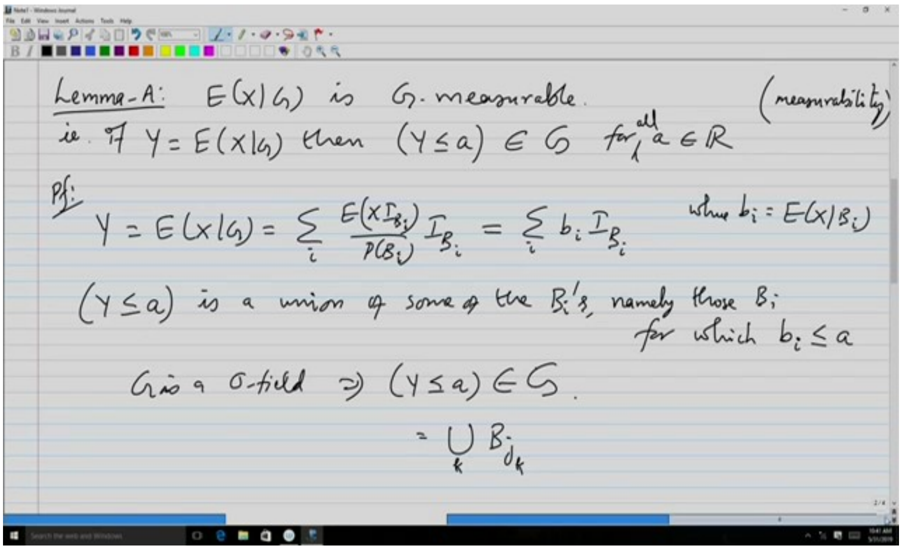
Suppose, if I assume that you know, I have a partition with respect to this, B_1, B_2, B_3, B_4 is the collection and then my Y can be written in that form. I could write it in this form, so even if I have more then, I could assume that there is 0 times that it is not an issue. Now, pick a as some value say 3.5. Now, if I look at the set $(Y \leq)$, what are those sets? This will be this is then nothing but my B_1 because it takes value 2 on B_1 and it takes value of the random variable Y takes value 3 on B_2 so that is 8.

Suppose, if I pick my a to be 6 for example, then this set is $B_1 \cup B_2$ and now with this will include B_3 also. So, this is in $\mathcal{G}$. So, you could see that, so since Y can be written in this form because we have just said that this is the definition, first and what are these 2, 3, 6 and a they are, these are these are essentially expectation of X given B_i, this values 2, 3, 6 and 4 that we have computed.

Suppose, if this is the conditional expectation expression then you could see immediately that for any a for any . So in this case, it will also be include B_4 because B_4 is 4. So, this is all, so which essentially if I can assume this could also be an omega or otherwise. So, this is what is then you will see as the measurability condition. So, the first property is that $\mathsf{E}(X|\mathcal{G})$ is measurable with respect to $\mathcal{G}$ which we always say $\mathcal{G}$ measurable, random variable.

(Refer Slide Time: 14:00)





And the second property which we state as lemma B is if I take an element in $\mathcal{G}$ and my $Y = E(X|\mathcal{G})$ then what you see is $E(YI_C)$ is will be equal to $E(XI_C)$. So, this for any set C in the sigma field $\mathcal{G}$.

In a way if I look at it so what you are saying is that if I take this to be equal to omega, then the left side is simply $E(Y)$ and this $E(X)$, which is the average of X and average of Y you can claim but now what we are doing: we are restricting this average to the subset $C(\omega)$. So, you could call this as partial averaging property.

If I go back to the previous one, so this I could call this as measurability, measurability. So, this property we would refer that this to be as the measurability property and the property that we are just we are stating which as partial averaging property. So, let us quickly see the proof which is just a pretty simple one.

Now, since $Y = E(X|\mathcal{G})$ and my $Y = \sum_i \frac{E(XI_{B_i})}{\mathsf{P}(B_i)} I_{B_i}$. Here, B_i's are disjoint. Since B_i's are disjoint, what this implies that if I take YI_{B_i} this sum will become this sum over i. So I take some particular B_j, then this sum will become $E(I_{B_j})$. It will become $\mathsf{P}(B_j)$and finally this would imply $E(XI_{B_j})$.

So, here, we are using an certain idea with respect to this indicator, which I will just recall. My $I_{B_j}I_{B_j} = I_{B_j}$ which is equal to 1 if $\omega \in B_j$ or 0 otherwise. $I_{B_j}I_{B_i}$ multiplied by is equal to 0 for $i \neq j$.

So, we are using that and hence this sum goes away, only one term would remain and $E(I_{B_j})$ again you know that this is nothing but $\mathsf{P}(B_j)$. So,what we seen is that B_j's are disjoint, now on disjoint sets are on the atoms for this sigma field $\mathcal{G}$, $E(Y|I_{B_j})$ where Y is the conditional expectation of X given $\mathcal{G}$ is same as this.

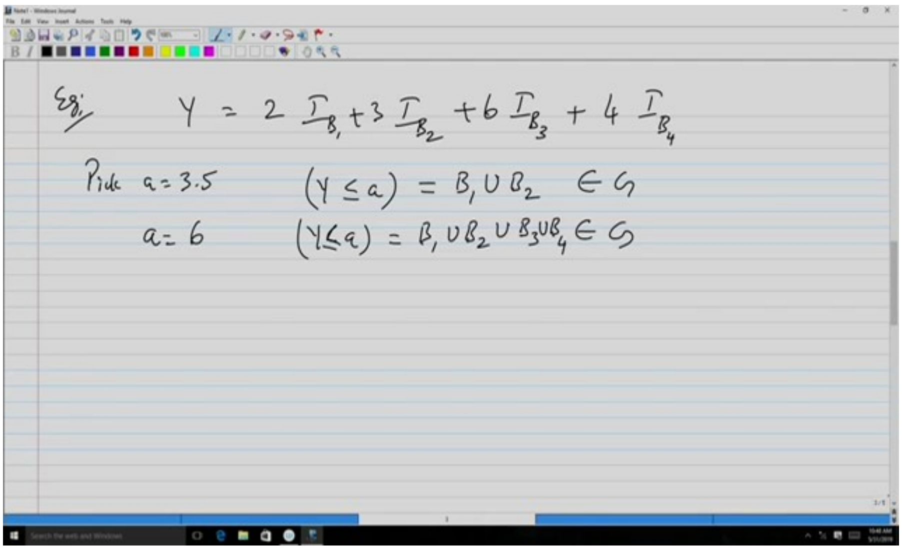
So, which mean that the partial averaging what we have just seen is the partial averaging is true on the sets B_i's but our claim is that it must be true for every C in the sigma field $\mathcal{G}$, but you will see that C is sigma field generated by the sets B_i's, so this C is then union of sum of this B_i's. B_{j_1} is the union sum B_{j_2} union and so on, union sum of B_{j_n} union and so on.

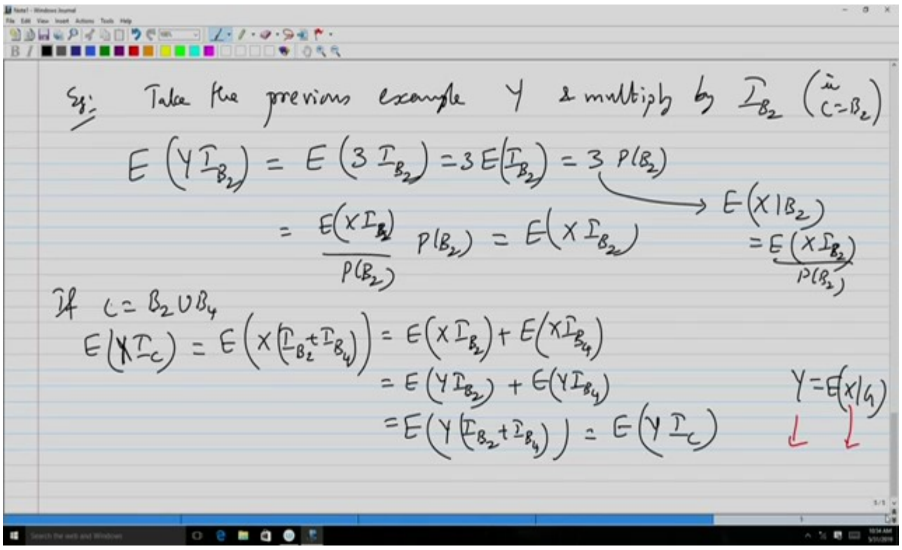
So, this any C which you pick from $\mathcal{G}$ can be represented as union of this B_i's. Now, on each of these B_i's this is true and they are disjoint, so summing over the this portion expectation of Y given I_{B_j} equal to expectation of X given I_{B_j}, so summing over; so if I call this as star over this what you call this say this jk's, what we get is expectation of $E(YI_C)$ will be equal to $E(XI_C)$, that is it.

So, what you are observing? So, how we are showing that? Instead of showing it directly for C we show it for the B_j's which generate the sigma field $\mathcal{G}$ and we see that on B_j's this is true. Now, any C is just simply the union of sum of these B_j's so summing over all such B_j it is going to be expectation of Y given I_C is equal to expectation of X given I_C.

So, this shows that this conditional expectation will satisfy this partial averaging property for on any sets which mean if I take the partial average of the conditional expectation and the original random variable X on any set in the sigma field, C then these two are one and the same.

(Refer Slide Time: 23:04)





Again let us take the previous example. In the above example, for this position and C. Now, the Y that we have considered a bit before this particular Y, now you take this Y and let us multiply this by indicator function on one of these B, say you take B_2, for example. So, take this Y and take multiply by I_{B_2}.

Now, then we see that $E(YI_{B_2})$ would be simply as only that one term will remain all other terms will become 0 when you multiply by this. So, it will be $3I_{B_2}$. So, which is nothing but $3E(I_{B_2})$ and this is nothing but $3P(B_2)$. But remember this 3 is not just any number 3, this is $E(X|B_i)$ which is then equal to $E(XI_{B_2})$ divided by $\mathsf{P}(B_2)$ which is $3\mathsf{P}(B_2)$.

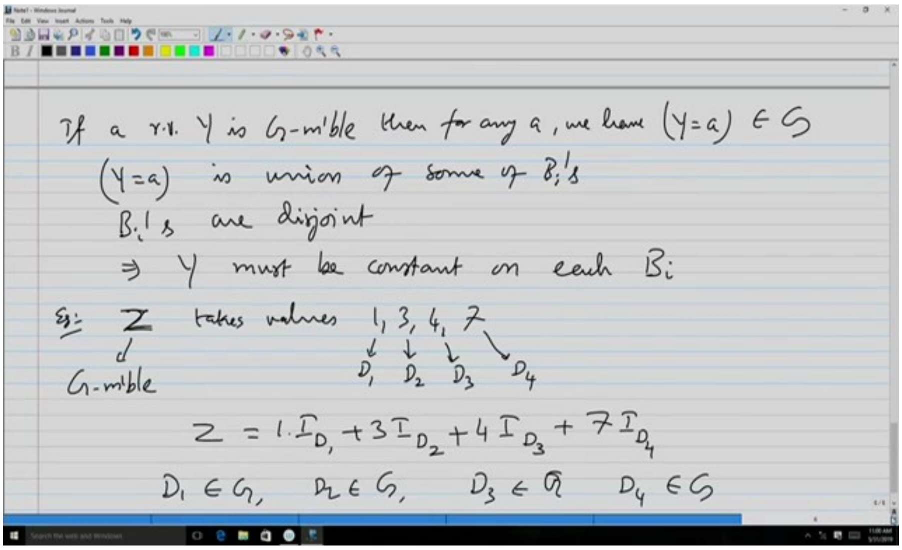
In this example what you are seeing is: what you wanted is what then you got it .If I take this particular random variable $Y = E(X|\mathcal{G})$ and this, where what we are noticing is this particular 3. This is nothing but $E(X|B_2)$ which is expectation of $E(XI_{B_2})$ by $\mathsf{P}(B_2)$, if I see that then.

Now, similarly if I take this is by taking actually, what I am taking is here, my C is B_2. Suppose, if my C is some union of this B_2, say if I take $B_2 \cup B_4$, then my expectation $E(XI_C) = E(XI_{B_2}) + E(XI_{B_4})$.

Now, by the first part, what we see is that $E(XI_{B_2})$, this particular quantity which we are writing it here is, if I use the first part this is nothing but expectation of $E(YI_{B_2}) + E(YI_{B_4})$, which is again $E(Y(I_{B_2} + I_{B_4}))$, which is nothing but $E(YI_C)$. So, if I take any C, since any other element in C can be written as union of sum of these B_i.

Suppose, if this is a example $B_2 \cup B_4$, then you can see that this is true. So, this is this example, we have seen that this is both measurable and it satisfy the partial averaging property where my Y is nothing but the conditional expectation of so this is important. For, $Y = E(X|\mathcal{G})$, if I take this, Y and this X, so this partial averaging property of this Y and the original random variable is X on the set C is what we have shown it.

(Refer Slide Time: 28:50)



Now, we will also observe certain things which will help us to prove the next result, which is, if a random variable say Y, which if it is $\mathcal{G}$ measurable, then then for any a real number a, so we have $(Y = a) \in \mathcal{G}$, which means that Y, this is the union of sum of this one or more of this B_i's, this $(Y = a)$ is one or more of this union. Because you said less than or equal to in case of discrete random variables, that could be equivalently tested using $Y = a$. So this is fine, even otherwise, you know, you will be able to see this.

Now, this is union of sum of B_i's, and B_i's or disjoint. So this implies that this Y which is the conditional random variable must be constant on each B_i. So, this is fairly easy to observe but still we will you know, explicitly state it, so that you know you do not have a problem when we go further, when we say that.

Now, if you want to look at an example how this is going to be and we can do that, let us quickly do that before we proceed further. Let us take a random variable say Z, which is taking values say 1, 3, 4 and 7 and let the corresponding sets on which this takes this particular value, if I call this as say D_1 and this is sum D_2, this is sum D_3 and this is sum D_4, this is nothing but so basically D_i.

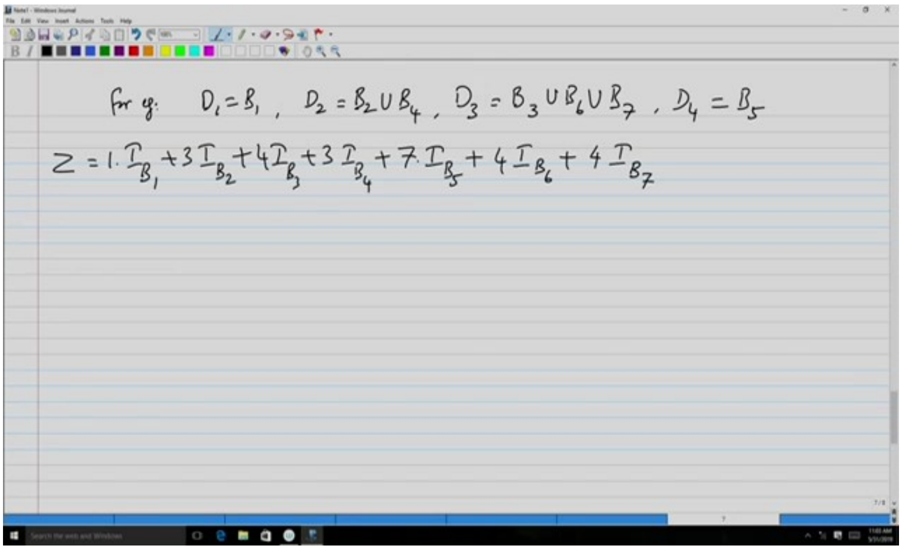
So, I cannot say D_i, so they will just say that 1, 3, 4, 7 so which means the Z, set of all ω on which Z takes value 1 is what we denote it by D_1 and so on for the other three, that means then the my $Z = 1I_{D_1} + 3I_{D_2} + 4I_{D_3} + 7I_{D_4}$.

So, this is what we say, so which means constant on each B_i, I mean we want to see and then Z take this value. Suppose if these or these, then my it means that Z can be represented in this form and you can easily see from this that my because Z is $\mathcal{G}$ measurable random variable, that is important.

These are $\mathcal{G}$ measurable random variable and D_1, D_2, D_3, D_4 then you can show or you can easily observe that they are all in $\mathcal{G}$. So, which if you want to see, so you can take $Z \leq a$ and which will be union of sum of these D_1, D_2, D_3, D_4, which you can see because this is a $\mathcal{G}$ measurable.

So, this first we can observe that D_1 belongs to $\mathcal{G}$. You can show that D_2 belongs to $\mathcal{G}$, D_3 belongs to $\mathcal{G}$ and D_4 belongs to $\mathcal{G}$. You can show. Now, because the sets in $\mathcal{G}$ are union of B_i's each of these D_i's are union of sum of those B_i's.

(Refer Slide Time: 35:08)

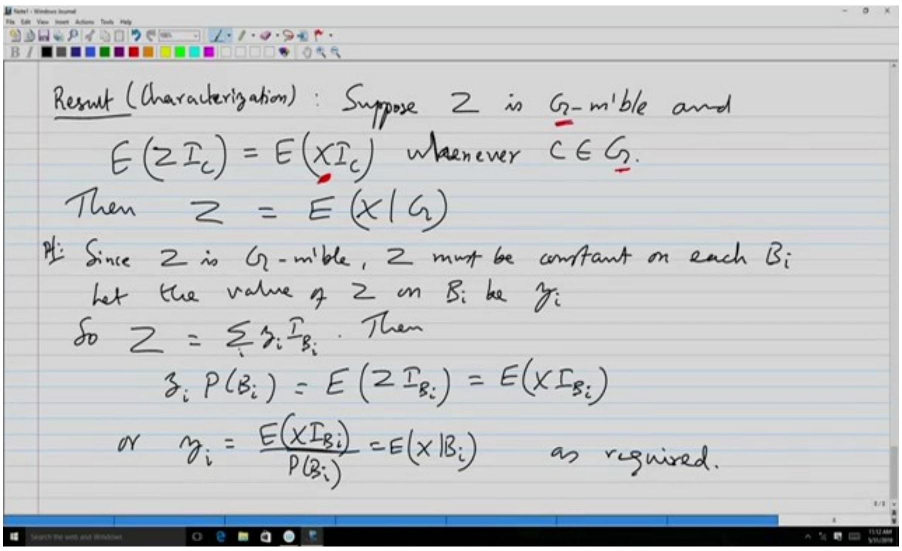


So, if you use that idea, if I use that idea, for example, if I pick my D_1 to be B_1, D_2 to be $B_2 \cup \cup B_4$ and $D3$ to be sum $B_3 \cup B_6 \cup B_7$ and D_4 this is simply an example. Suppose, if D_1 is first thing we have observed, we took a random variable which is $\mathcal{G}$ measurable and we first saw that each of those D_1, D_2, D_3, D_4, this set first belong to $\mathcal{G}$, which you can easily show by taking Z that the random variable is equal to a.

Now, since D is they are all in $\mathcal{G}$ and in any set in $\mathcal{G}$ can be written in terms of the elementary sets in a way B_1, B_2, B_3 and so on. So, if for example, if these can be written in this form, then that Z that we have written, you could see that you can write it as $1I_{B_1} + 3I_{B_2} + 4I_{B_3} + 3I_{B_4} + 7I_{B_5} + 4I_{B_6} + 4I_{B_7}$.

So here, this is what you would have. So, which means that what we are seeing here is that any Z which is $\mathcal{G}$ measurable can be represented in this form where this particular value which is either 1, 3, 4, 3, 7, 4, 4, and so on that we say they are all constants on each of these that is what we have observed. So, this you can observe easily and you see that whenever a random variable is measurable with respect to the particular sigma field $\mathcal{G}$ and if the $\mathcal{G}$ is a sigma field which is generated by this countable number of sets B_i's, then that random variable is constant in each of those B_i's. So, this is what we see here.

(Refer Slide Time: 38:27)



Now, going forward, we have seen two lemmas, one is measurability, one is the partial averaging property of this conditional expectation. Now, we climb that these two properties characterize the conditional expectation of a random variable. So, which we call it as a result which is the characterization of this conditional expectation, what we say is that we have shown like if we have defined a conditional expectation and we said that this is $\mathcal{G}$ measurable and satisfied with the partial averaging property.

Now, on the other hand what we are saying is if you pick a random variable, which is $\mathcal{G}$ measurable and which satisfies the partial averaging property then it must be the conditional expectation of X given $\mathcal{G}$, that is what we are saying. So, suppose pick a random variable Z which is $\mathcal{G}$ measurable and $E(ZI_C)$ is equal to $E(XI_C)$ whenever my C is in $\mathcal{G}$.

Means, for all C in $\mathcal{G}$, I have this partial averaging property satisfied and this Z is $\mathcal{G}$ measurable random variable and for any set in $\mathcal{G}$, I (sat) this equality satisfied for some random variable X then the claim is this Z is nothing but the $E(X|\mathcal{G})$, so where are these X and $\mathcal{G}$ comes, so X comes from here and $\mathcal{G}$ comes from here.

So, you have a random variable Z which is $\mathcal{G}$ measurable and then for any set $\mathcal{G}$ this partial averaging property is satisfied with respect to some random variable X then this Z is nothing but the conditional expectation of this X given that $\mathcal{G}$, that is it. So, the proof is goes along similar lines in a simple manner, so this is not so difficult, so what we observe is that since Z is $\mathcal{G}$ measurable, Z must be constant on each B_i. Let the value of Z on B_i be say z_i.

So, what we have this $Z = \sum_i z_i I_{B_i}$. Then what you have $z_i \mathsf{P}(B_i)$. If I consider this, this z_i is a constant and $\mathsf{P}(B_i)$. I can write it as expectation of indicator function of B_i, then this z_i can be pushed inside because this is a constant on B_i..

Then, this is nothing but since on each set in $\mathcal{G}$, Z satisfies the partial averaging property with respect to X, so this is nothing but $E(XI_{B_i})$ or my $z_i = E(X|B_i)$, as required.

So, what we have shown, if Z is $\mathcal{G}$ measurable then what we know is Z must be constant in each of this B_i's. Suppose, if that value be Z_i, then Z can be written in this form, then if I consider Z times probability of B_i which can be written or which is nothing but this quantity by observing the two facts that Z is constant on B_i and $p(B_i = E(I_{B_i}))$, then this must be equal to this and by the partial averaging property which Z satisfies this must be equal to this, otherwise Z is equal to this, which is what this is.

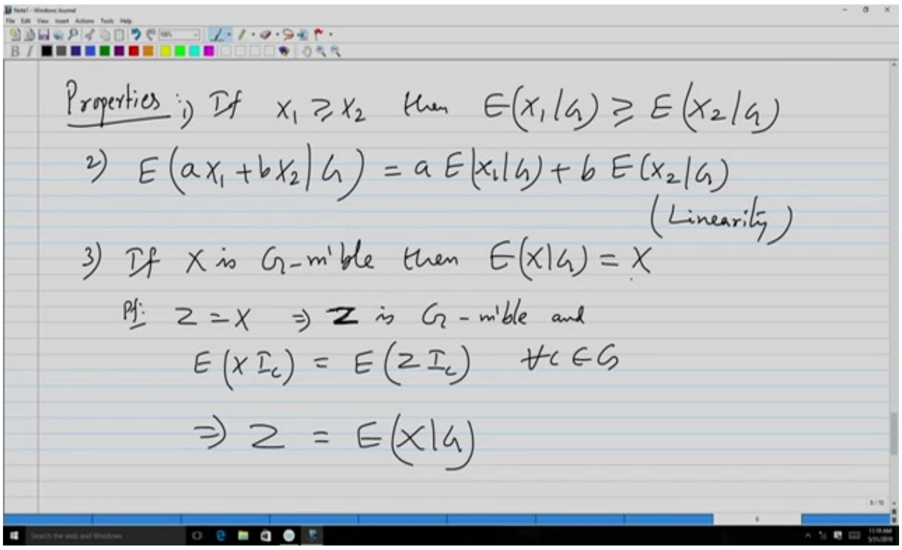
Now, if I pick this is plug in this zi here, then what I get this Z is what I define as $E(X|\mathcal{G})$. So and hence, this is the as required. So, this is what is the characterization. Now, what is shown is we defined $E(X|\mathcal{G})$ and we said that this is that particular quantity is $\mathcal{G}$ measurable and that could particular quantity that we defined is satisfy the partial averaging property.

Now, if I pick any $\mathcal{G}$ measurable random variables such that it satisfy the partial averaging property, then that random variable is must be this $E(X|\mathcal{G})$. So, all these things, you know, we are stating to discrete so we are writing this is a sum in this manner and so on so we are in this domain.

Now, this is what is the characterization, so for discrete this is why this is important I will tell you, for the discrete we have defined by extending the idea of $E(X|B)$ to $E(X|\mathcal{G})$. In continuous this is not feasible to do, so we need to look for other definition and in continuous time when we go, we will take this characterization as the definition of the conditional expectation, means it is a random variable which is $\mathcal{G}$ measurable and which satisfy the partial averaging property is the conditional expectation of X given that particular sigma field.

That is will be the original definition of definition and that you know, so it is easier in discrete time to see this and I said that when you go for general $\mathcal{G}$, which may not be count to generated one then this definition no more holds the original definition that shows but this categorization is equivalent, that is what is we mean characterization. So, this characterization will be the definition that we will carry over to continuous time so on, hence you can see like this is how then we observe this.

(Refer Slide Time: 46:44)



Now, let us see some of these properties that are important with respect to this conditional expectation before we go further, certain properties are straightforward from the definition, say for example, if X is non-negative, by the way, when we assume that the $E(X|\mathcal{G})$, you know, whether the question of whether such exists such a random variable are such a quantity.

We are taking the assumption that $E(X)$ is finite, which means expectation of X exist, as we defined earlier when we written that as sum provided the sum converges, so when that exists this always exists and it is unique and then we will see that in the later case. There is only one such random variable, which is for every X and given $\mathcal{G}$ which is the unique random variable, which is what will be X.

Now, certain properties are very easy, like which will carry over from the simple expectation concept itself, which are something like if X is non-negative then the $E(X|\mathcal{G})$ will also be non-negative. The other thing is generalization of that say suppose this is property one, if $X_1 \geq X_2$, then $E(X_1|\mathcal{G}) \geq E(X_2|\mathcal{G})$.

This is easy to prove by from the definition itself, you just write down the definition and you would see that immediately, because each of those X given B_i's would be greater than or equal to X_1, X_2 given B_i's, so that follows immediately from the definition.

Similarly, the other property which is

$$E(aX_1 + bX_2|\mathcal{G}) = aE(X_1|\mathcal{G}) + bE(X_2|\mathcal{G})$$

Thus is what we say is linearity property, of course, you know, if I look back I mean we have already used this in the proofs that we have that so far.

So, this is what linearity property is. Again as we said when we defined $E(X)$ itself that if I define that as an alternative way as we have done then the proofs of this is immediate, so we do not need to go through the proof, so this is straight forward from the definition.

Third property which is if X is $\mathcal{G}$ measurable then

$$E(X|\mathcal{G}) = X.$$

If X is $\mathcal{G}$ measurable, then $E(X|\mathcal{G})$ is nothing but X. So, how do we show? Again we appeal to the characterization and we pick a random variable which will satisfy the two properties of the characterization.

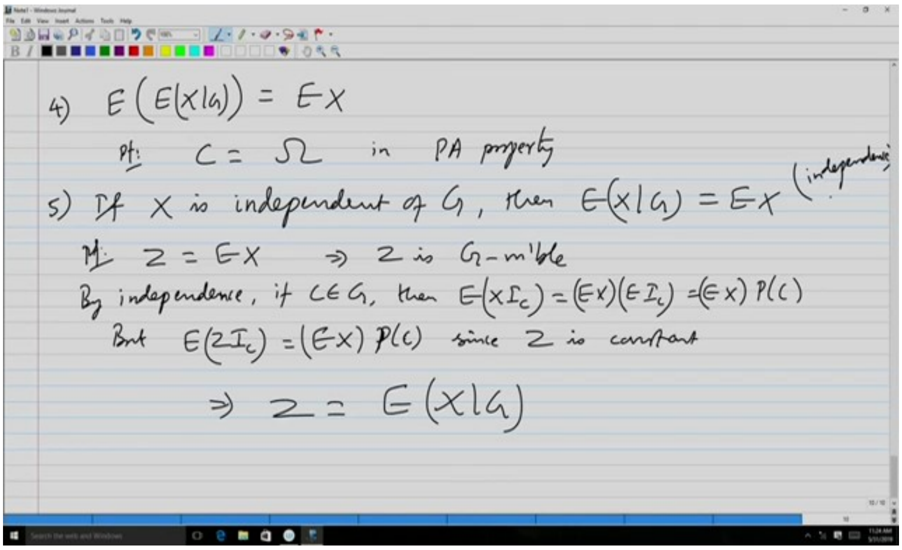
So, what we see? Suppose, if I look at the proof here, if I pick my Z to be X because our claim is, you know, $E(X|\mathcal{G})$ is X itself, so we will need to show that X has those two properties, pretty obvious but still will go through. X is, suppose, if iI want to show that Z is the one which as per the characterization satisfies those two properties.

So Z, I picked to be X and since X is given to be $\mathcal{G}$ measurable so Z is also $\mathcal{G}$ measurable. So, that implies Z is $\mathcal{G}$ measurable and if I pick expectation of XI_C, would be same as $E(ZI_C)$, for all $C \in \mathcal{G}$. So this is trivial, because my Z is equal to X, so $E(X|\mathcal{G})$ is equal to this, so which means that this is Z.

So, my Z is $\mathcal{G}$ measurable because Z I have taken to be X, X is given to be G measurable and if I look at the partial averaging of X given I_C which is nothing but $E(X|Z)$ is equal to X so this is true. So, it also trivially satisfy the partial averaging property, then by the characterization that we have made that a random variable which is $\mathcal{G}$ measurable, which is the satisfying the partial averaging property must be $E(X|\mathcal{G})$. So, this implies that my Z is simply $E(X|\mathcal{G})$, so that is what it is.

So, this is call this measurability property and later on we will see an interpretation for X given G. Now, once we have that interpretation then obviously like then you will be able to get the better picture of this property.

(Refer Slide Time: 53:07)

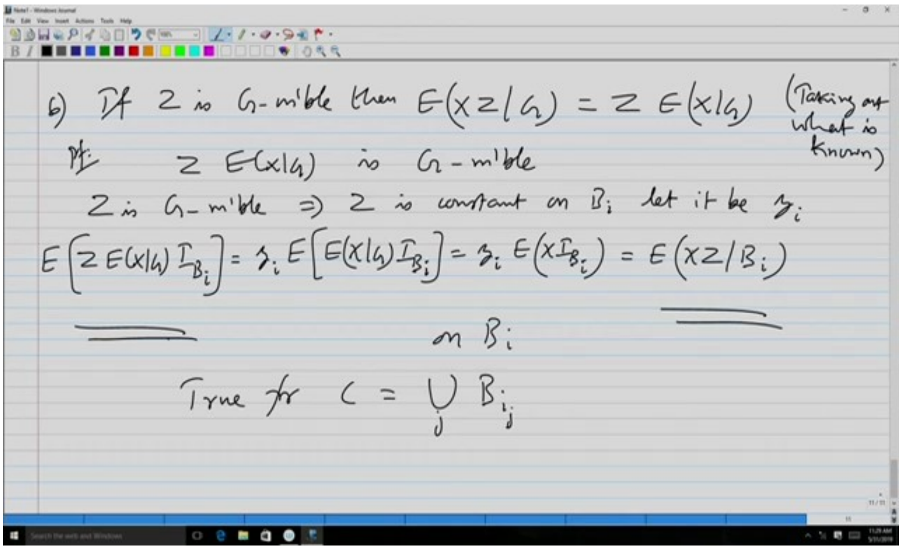


So if I, property 4 is $E(E(X|\mathcal{G})) = E(X)$. So, this is what we want to see which means that I have a random variable X and I know what is $E(X)$, you can compute. Now, what this says is that $E(X|\mathcal{G})$ for some $\mathcal{G}$ you pick, you take and you take since these are random variable, you have seen the $E(X|\mathcal{G})$ is a random variable. Now, you take expectation of that random variable that will be exactly equal to $E(X)$. Since $E(X|\mathcal{G})$ satisfies the partial averaging property, so you can take $C = \Omega$ in the partial averaging property and this is immediate.

Then we have if X is independent of $\mathcal{G}$, then $E(X|\mathcal{G})$ is nothing but $E(X)$. Again proof, we appeal to the characterization property, we take $Z = E(X)$ because this is what we want to show as $E(X|\mathcal{G})$ and we say X is independent of $\mathcal{G}$. So, Z is a constant, so immediately this is $\mathcal{G}$ measurable, because a constant random variable or a degenerate random variable is measurable with respect to every sigma field, with the trivial sigma field itself, so this is that is $\mathcal{G}$ measurable.

So, by the independence property, if C is in $\mathcal{G}$ then $E(XI_C) = E(X)E(I_C)$. Since X is independent of $\mathcal{G}$ and hence any element in $\mathcal{G}$ is independent of X, so this I can write it as $E(X)E(I_C)$, which is $E(X)\mathrm{P}(C)$, but what is this quantity? $E(ZI_C)$ is same as $E(X)\mathrm{P}(C)$, because my Z is $E(X)$ and plug it in so you get the same, so this is true, , this is because Z is constant. So, again appealing to the characterization proportion implies my Z is nothing but $E(X|\mathcal{G})$. So if X is independent of $\mathcal{G}$ and $E(X|\mathcal{G})$ is simply $E(X)$, this is property 5 and this is the this property we call it as independence property.

(Refer Slide Time: 57:49)



Now, property 6 is: if Z is $\mathcal{G}$ measurable then $E(XZ|\mathcal{G}) = ZE(X|\mathcal{G})$. So again, so what we say is that if Z is $\mathcal{G}$ measurable random variable then this Z in that case acts as it is a constant, it can be taken inside or outside of this conditional expectation quantity, that is what you know we see as far as the conditional expectation is concerned.

Let us see how this is true, so if you look at this quantity, so we want to show that $E(XZ|\mathcal{G})$ is given by this, so if we take this an appeal to the characterization and to see that whether that is exactly equal to the conditional expectation of this random variable XZ. So, what we do is that we appeal, when we first showed the partial averaging property we did it on set B_i's and then we extended to any set C.

We want to show the partial averaging property. This is the same idea that we will use here but first we notice that the first property is trivial again here, Z is $\mathcal{G}$ measurable $E(X|\mathcal{G})$ by definition is $\mathcal{G}$ measurable, so product of two $\mathcal{G}$ measurable random variables is $\mathcal{G}$ measurable, so $ZE(Z|\mathcal{G})$ is $\mathcal{G}$ measurable, so that part is so clear that the first part the measurability property is immediate.

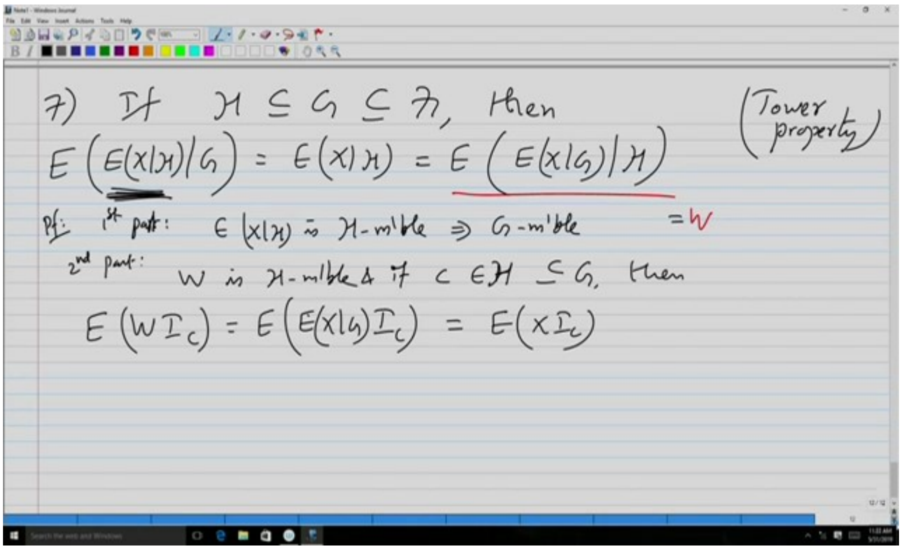
Now, for the second part, it is sufficient that if we consider only the sets B_i's and show the partial averaging property of these two quantities, then it is clear. Then for it is true for any C in $\mathcal{G}$. Now, since Z is $\mathcal{G}$ measurable, this implies Z is constant on each of these B_i's, let it be say z_i, you take a B_i then you can take $E(ZE(X|\mathcal{G})I_{B_i})$.

Since Z is constant on this B_i, so this Z takes value z_i that can be taken out essentially, so you have $z_iE(E(X|\mathcal{G})I_{B_i})$. But by definition of expectation of X given $\mathcal{G}$, this is nothing but z_i expectation of X indicator function of Bi, this is by the definition of expectation of X given G.

Now, this z_i on B_i, again it is constant, you can take it inside, so you can see that $E(XZ|B_i)$. So this is true that this part is same as this part on B_i and for so this is true for some C, which is union of sum Bij's, somehow. So, once you show it for B_i it is also true for any C_i, that is what we have here. So, this is property 6, which we call it as taking out what is known.

Z measurability means that we know Z, if Z is $\mathcal{G}$ measurable then we know value of Z. So, then you can take out that and then compute the remaining $E(X|\mathcal{G})$, so that is why this is called as taking out what is known property as far as this is concerned.

(Refer Slide Time: 62:51)



Now, we have two other properties and we will just complete with that. Now, if I take $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$, then what we claim is that

$$E(E(X|\mathcal{H})|\mathcal{G}) = E(X|\mathcal{H}) = E(E(X|\mathcal{G})\mathcal{H})$$

This is called as Tower property or iterated conditioning property.

You have $\mathcal{F}$ is the larger sigma field and $\mathcal{G}$ is the sub sigma field of $\mathcal{F}$ and $\mathcal{H}$ is sub sigma field of $\mathcal{G}$. Then what the left hand side is that first you compute the $E(X|\mathcal{H})$ and for that random variable then the compute condition of that random variable is given $\mathcal{G}$ then that will be equal to simply directly computing the $E(X|\mathcal{H})$.

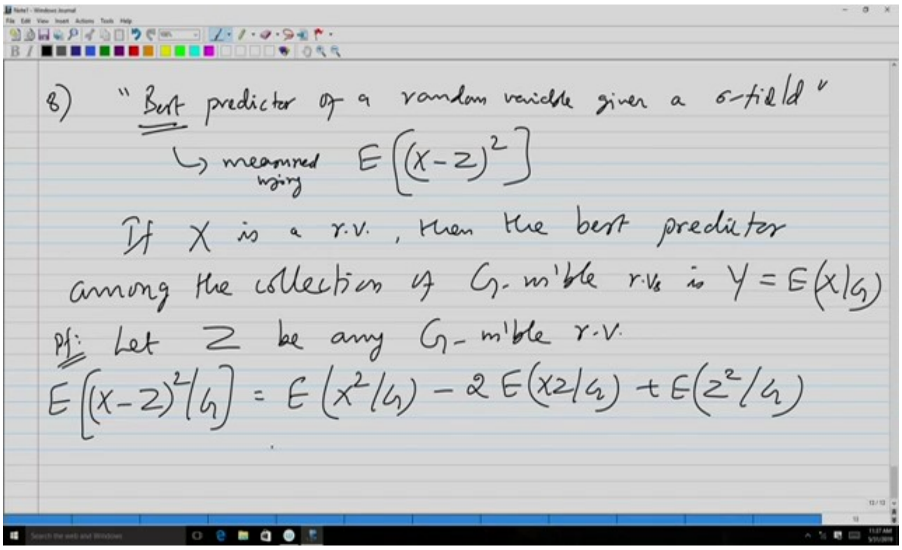
Similarly, the other way around, first with respect to $\mathcal{G}$ then with respect to $\mathcal{H}$, then it is simply with respect to the smaller sigma field is what will turn out to be is what the claim. Now, prove the first part is immediate as you can see $E(X|\mathcal{H})$ is $\mathcal{H}$ measurable and hence and $\mathcal{H}$ is a sub sigma field of $\mathcal{G}$ implies it is $\mathcal{G}$ measurable and whenever it is $\mathcal{G}$ measurable we have just seen one of the property if X is $\mathcal{G}$ measurable then $E(X|\mathcal{G}) = X$.

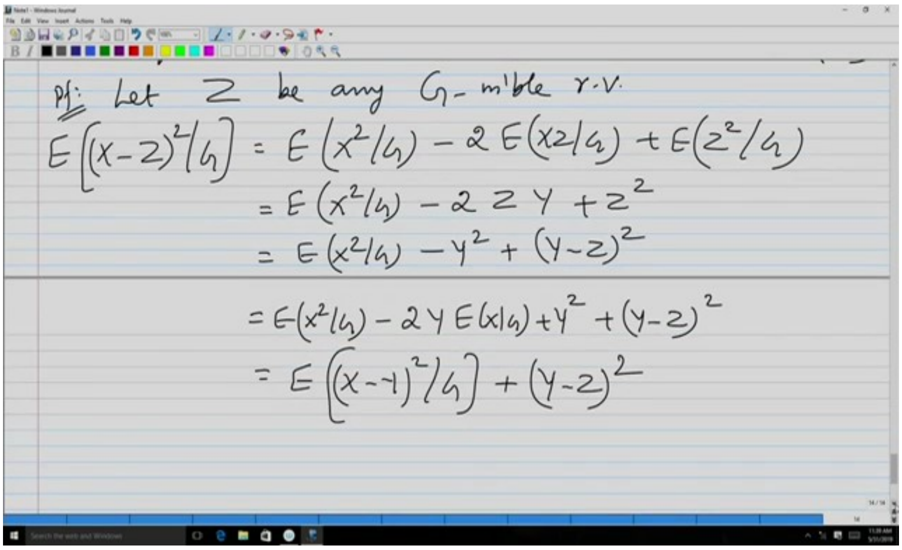
So, this particular quantity whenever any quantity is $\mathcal{G}$ measurable then expectation of X given whatever this quantity is here, is simply equal to that quantity, so that is immediate.

The second part which is the second in equality is what that you need to consider a bit. Now, let you know the whole thing be equal to say some W, is what then you have. Now this W, so that this B equal to some W. Now, this W is $\mathcal{H}$ measurable, and if you take C which is in $\mathcal{H}$, which is a sub sigma field of $\mathcal{G}$ then my $E(WI_C) = E(E(X|\mathcal{G})I_C) = E(XI_C)$. This is the definition of this W by definition for this W, this is equal to this but again by since C is in $\mathcal{H}$ which is sub sigma field of $\mathcal{G}$, so for C belongs to $\mathcal{G}$.

So, then again you can use the definition of $E(X|\mathcal{G})$ here further, you can see that $E(X|I_C)$, this is what we want to show, so this is also true. It means that for $C \in h$, expectation of W is indicator function of C is equal $E(X|I_C)$ and hence the second equality is also proved.

(Refer Slide Time: 67:24)





Now, we will go to the last property which will given interpretation to what this conditional expectation means and that interpretation help us understand the things even all other properties little bit more clear. Now, the interpretation is that this is in a way we said that $E(X|\mathcal{G})$, given $\mathcal{G}$ and X and whether there is unique and why we have to pick this particular quantity rather than over any other quantity, when you want to look at it.

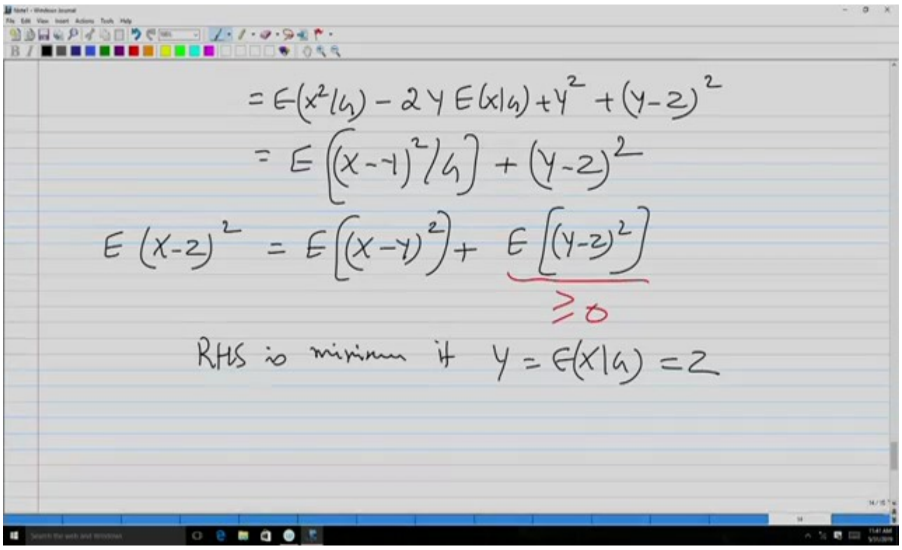
So, there the meaning comes from the fact that this may be so this may be viewed as the best predictor in some sense of random variable given sigma field. So, this is the meaning, so the best predictor of a random variable given a sigma field is what is this condition interpretation of X given $\mathcal{G}$. What is the best in the sense? So, this goodness of the best is measured by the quantity $E((X-Z)^2)$, so which is what is commonly referred to as the mean square error in terms of if you are aware about this.

So, what is the result, the result is the following: if X is a random variable then the best predictor among the collection of $\mathcal{G}$ measurable random variables is Y which is equal to $E(X|\mathcal{G})$ is this particular quantity, that is what is the claim.

Let us see the proof, so let Z you pick to be any $\mathcal{G}$ measurable random variable. Now, we could consider $E((X-Z)^2|\mathcal{G})$ to be $E(X^2|\mathcal{G}) - 2E(XZ|\mathcal{G}) + E(Z^2|\mathcal{G})$. is what where you have seen that already I have used the linearity property here, this could be further written as $E(X^2|\mathcal{G}) - 2ZY + Z^2$ as Z is G measurable random variable.

So, this is I can write this as this plus again this is equal to this quantity, which you can write it again as $E(X^2|\mathcal{G}) - Y^2 + (Y_Z)^2$. By adding and subtracting this and doing some more operations we got the final result as required.

(Refer Slide Time: 72:47)

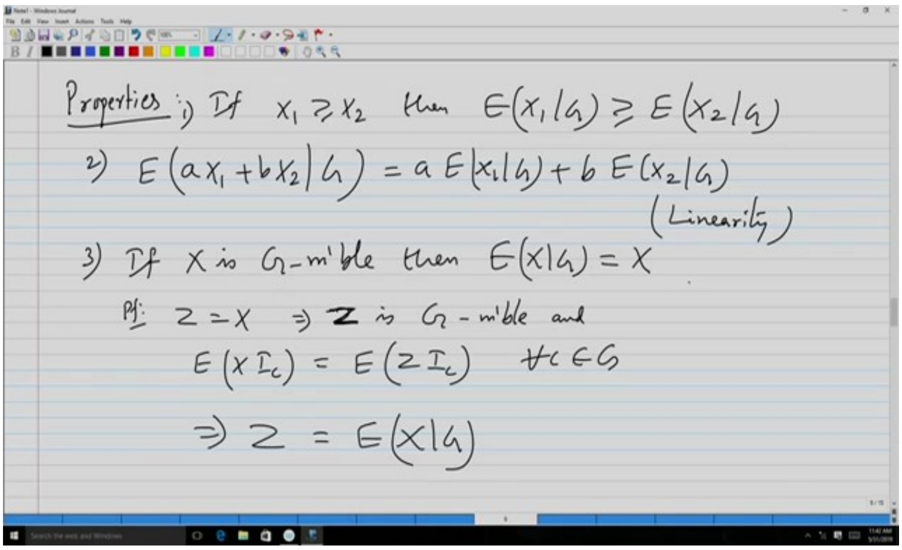


Now, we have used the fact that Y is $\mathcal{G}$ measurable and again if I take once more, the expectation on both sides, what you will end up with $E((X-Z)^2)$ is nothing but $E((X-Y)^2)$ plus $E((Y-Z)^2)$. You know that $(Y-Z)^2 \geq 0$, so expectation of a non-negative quantity is non negative.

So, this particular quantity is greater than 0, so the right hand side is bigger than or equal to $E((X-Y)^2)$ and the minimum for the right hand side would be achieved when Y is equal to Z. So, Z is nothing but a generic $\mathcal{G}$ measurable random variable, $Y = E(X|\mathcal{G})$. And the RHS is minimum if $Y = E(X|\mathcal{G}) = Z$.

Now, for all other random variables this quantity is going to be bigger than X, so this Z in a way is the best predictor among all $\mathcal{G}$ measurable random variables for predicting X given $\mathcal{G}$. So, this is the interpretation that you need to keep in mind.

(Refer Slide Time: 74:45)

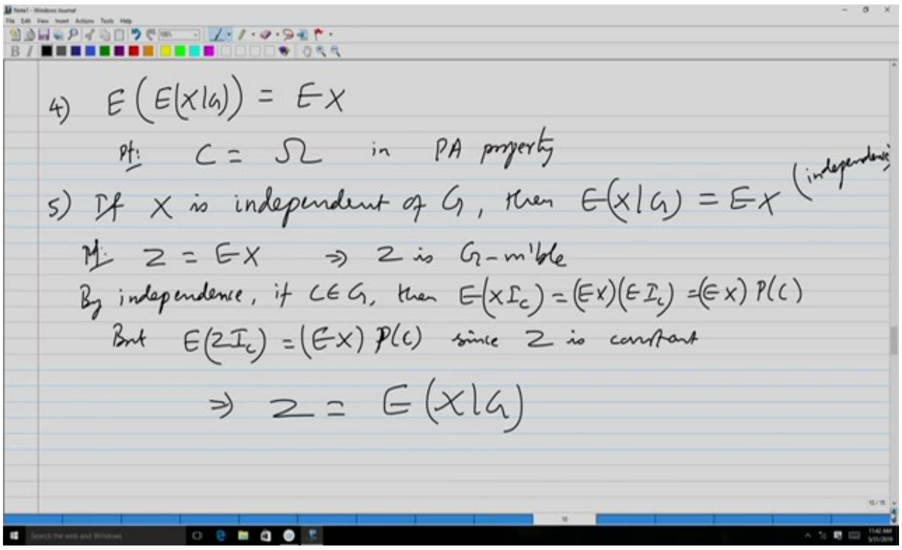


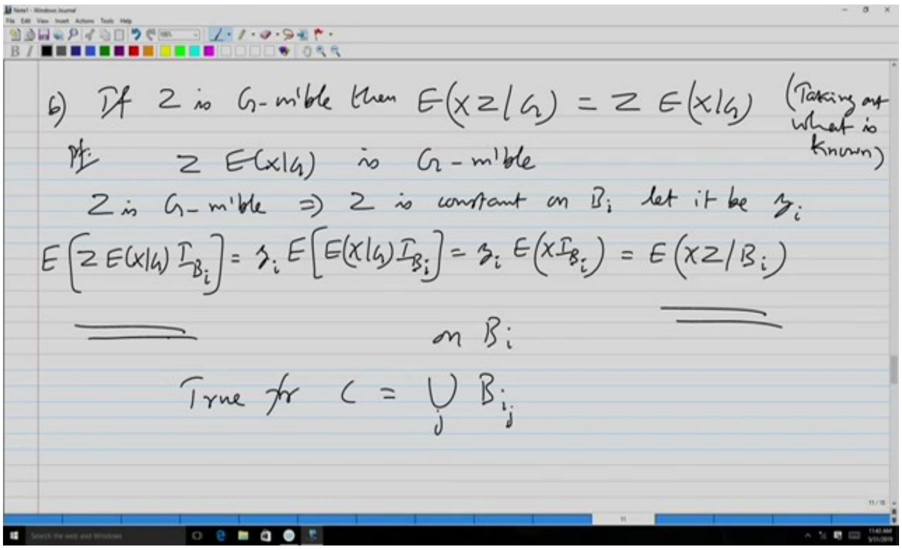
Now, with this interpretation, with this interpretation if I go back and look at each of these properties now you can give meaning to each of these properties that if $X_1 \geq X_2$ then the best predictor for X_1 given $\mathcal{G}$ will be greater than or equal to the best predictor for X_2 given $\mathcal{G}$.

Similarly, if I have to predict aX_1 plus bX_2, the best predictor can be, you can separately find out. $E(X_1|\mathcal{G})$ is the best predictor of X_1, the best predictor of X_2 and combine linearity in this form.

Now, if X is $\mathcal{G}$ measurable then the best predictor for X given G is X itself. Similarly, if I take the average of all the best predictors, the best predictor then it is $E(X)$ simply.

(Refer Slide Time: 75:32)





Now, if X is independent of $\mathcal{G}$, which means that $\mathcal{G}$ has no impact on X, then the best predictor of X given $\mathcal{G}$ reminds as in simple ordinary probability theory as $E(X)$. Similarly here the best predictor of XZ given $\mathcal{G}$ can be done by decomposing: Z taking out and then computing the best predictor of X given $\mathcal{G}$ separately and then multiplying it.

Now, here if I look at it, if the best predictor of X given two sigma fields one as a sub sigma field of the other is nothing but the best predictor of X given the smallest sigma field is what this tower property would mean. So, all this interpretation now you can give best upon this best predictor idea.

So, this is the reason why you know you pick the conditional expectation or this concept expectation of X given $\mathcal{G}$. The way we have define is the best predictor for X given $\mathcal{G}$, remember when we do not have any information about the random variable the best predictor is $E(X)$ which is equivalent in our case, we are saying that when we have no information on that the best predictor is given by $E(X)$.

Now, as information you accumulate, which is given in terms of $\mathcal{G}$ now then what is the best predictor is nothing but the conditional expectation of X given $\mathcal{G}$. So, this interpretation you keep it in mind because this best predictor idea is what that we are going to encounter, thus little bit more onto this conditional expectation which we will see in the following lectures, thank you.