**Introduction to Probabilistic Methods in PDE**

**Doctor Anindya Goswami**
**Department of Mathematics**
**Indian Institute of Science Education and Research, Pune**

**Lecture 06**

**Revision of Conditional Expectation (Part 02)**

(Refer Slide Time: 0:15)

**Conditional Expectation:**

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X$ a random variable on it with $E|X|$ finite. Assume that $\mathcal{G}$ is a sub $\sigma$ algebra of $\mathcal{F}$. Then $E(X|\mathcal{G})$ is a $\mathcal{G}$ measurable function such that for every $A \in \mathcal{G}$,

$$E[E(X|\mathcal{G})1_A] = E[X1_A].$$

So, now I come to another topic. So, this much, I just had to talk about the stochastic processes and their measurability issues and their equivalence in various different notions. Now, we recall what do we mean by conditional expectation. So, these are basically recollections because I have assumed that expertise  of major theory and basic probability theory.

So, let omega F, P be a probability space and X a random variable on it with expectation of mod X finite. And assume that script G is a sub-sigma algebra of F. Then we write down this manner expectation E of X given G, We read it in the following manner expected conditional expectation of X given G.

So, this conditional expectation X given G is nothing but a random variable, but G measurable function, G measurable function, such that for every A in G, we have this condition to be true. So, that is the definition of expectation X given G. What is its condition? That expectation of, conditional expectation of X given G multiplied with indicator function of A, that is equal to expectation of X multiplied with indicator function of A.

So here, Expectation is integration, correct? So, left hand side we are integrating expectation X given G on the domain A with respect to probability is P and right hand side, we are integrating X with respect to the probability is P on the domain A. However, on the left and right side both sides, we can, so here on the left hand side, since you know, this is G measurable function and A is also G measurable. So, left hand side, this integration, we can very well take as you know with respect to probability measure P which is restricted to the sigma algebra G. I can take that.

(Refer Slide Time: 2:38)

**Conditional Expectation:**

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X$ a random variable on it with $E|X|$ finite. Assume that $\mathcal{G}$ is a sub $\sigma$ algebra of $\mathcal{F}$. Then $E(X|\mathcal{G})$ is a $\mathcal{G}$ measurable function such that for every $A \in \mathcal{G}$,

$$E[E(X|\mathcal{G})1_A] = E[X1_A].$$

If $PX(A) := E[X1_A]$ for all $A \in \mathcal{G}$, then $PX$ is a signed measure on $(\Omega, \mathcal{G})$. Again $P|_{\mathcal{G}}(A) := P(A)$ for all $A \in \mathcal{G}$ is a measure on $(\Omega, \mathcal{G})$. Clearly $PX \ll P|_{\mathcal{G}}$, $PX$ is absolutely continuous w.r.t. $P|_{\mathcal{G}}$.
Then $E(X|\mathcal{G})$ is the Radon-Nikodym derivative of $PX$ w.r.t. $P|_{\mathcal{G}}$,

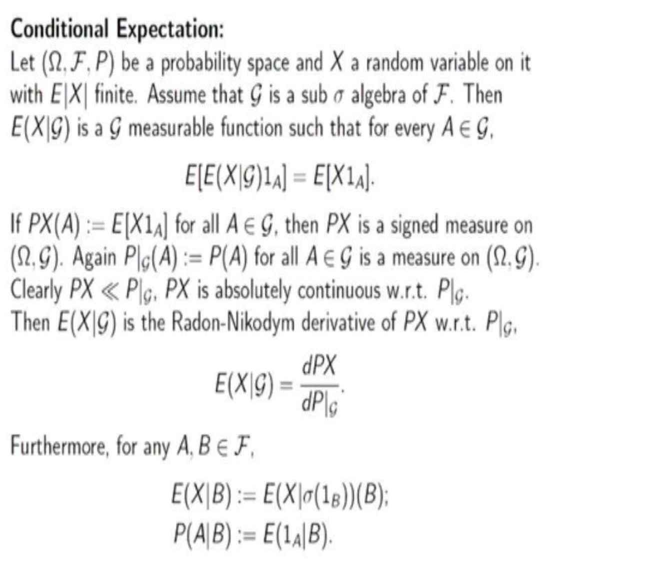
$$E(X|\mathcal{G}) = \frac{dPX}{dP|_{\mathcal{G}}}.$$

Furthermore, for any $A, B \in \mathcal{F}$,

$$E(X|B) := E(X|\sigma(1_B))(B);$$
$$P(A|B) := E(1_A|B).$$

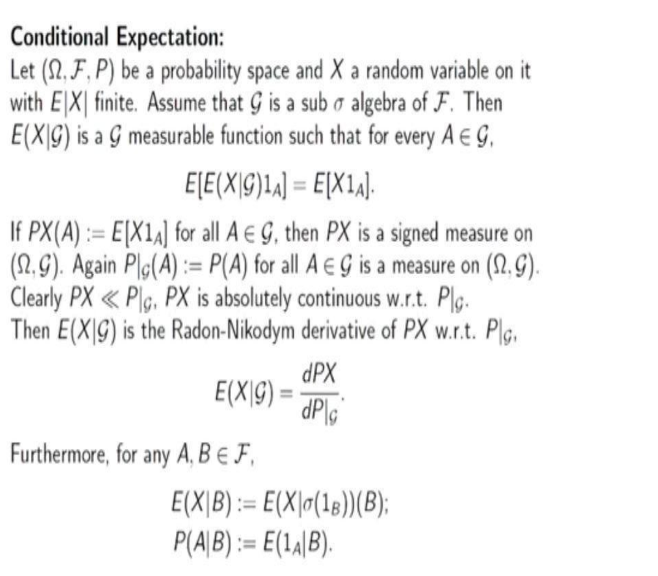
So, that is the definition. Let, us try to understand the definition in a little greater detail. So, we define another signed measure, basically this is not a measure. So, PX, PX is a single object, which if we evaluate on a set A where A is a G measurable set, so that evaluation, that definition is expectation of X times 1, indicator function of A. So, here this value could be positive or negative. However, this value, so if we take the positive part of the inside, so that would have the finite value, if I take the negative part of this function, I take the expression that will also be finite value. Why is it so? Because X has finite expectation, mod X has finite expectation.

(Refer Slide Time: 3:29)




So, I would get a signed measure here. So, PX is a signed measure for all A and G. Then PX, Then what we do is that, we also define P restricted on G as I was indicating earlier, which is nothing but P probability of A, but A is in G. So, this is just a restriction of probability measure, not on the full sigma algebra F but a sub-sigma algebra G. So this is also a measure.

(Refer Slide Time: 4:00)



Now, this PX measure and P, you know, restricted on G, so this has a relation. So, what is the relation? That if any set A has measure 0 with respect to this measure, then that set has also 0 value with respect to, actually I should have written mod here, with respect to the total variation of PX measure. So, what does it mean? That PX is a signed measure. Every signed measure, due to the Hanh Decomposition, has two parts, the positive measure part and the negative part. And if you add those things, you are going to get the total variation of the signed measure and the total variation measure is absolutely continuous with respect to this measure.

So, I mean, why is it so? Because you know, if any set A has 0 measure with respect to this PG measure, then that set over that set, this part would be 0 almost surely, so aspect this part would be 0. So, it is as easy as that. So, PX is absolutely continuous with respect to PG, then we can use the Radon-Nikodym theorem theorem . So, Radon-Nikodym theorem says that we can find out a Radon-Nikodym derivative of PX with respect to PG. And what is that? That would be exactly this. Why? Because this is retained in that manner. It is retained in exactly that manner. That, this part is integrated on the domain A over with respect to the probability measure. And here this PX appears here. So this expression is given G is a Radon-Nikodym derivative of PX with respect to PG.

So, this is another way of viewing a Conditional Expectation as Radon-Nikodym derivative. To define this way, we do not need to assume that the random variable has any finite variance. We just need it has finite expectation. But there are some other ways. So, if in addition to this, you know, you know that random variable has finite variance, then there is another geometrical way to define Conditional expectation as projection.

Now, we introduce a few more notations. So we write down expectation of X given B, B is just a measurable set with respect to F. So that we define as Expectation of X given the sigma algebra generated by indicator function of B. Why do we do that? Because, we have already introduced the meaning of expectation A is given as sigma algebra. So, that we find out and evaluate that on omega, any omega which is in B, does not matter what omega we choose, why, because this thing is a function, this Conditional Expectation, which is measurable with respect to the 1B, sigma algebra generated by 1B.

And as I have mentioned earlier, that if Y is measurable with respect to sigma X then Y is constant where X is constant, those parts. So, since 1B is constant on B, so this whole thing should be constant on B. So, this is meaningful, the right hand side is meaningful, whatever the value is, we are going to call that as expectation as X given B. Now, what do we mean by probability A given B? So, we are defining in this manner, in the following manner, expectation of indicator function of A, that is now a random variable, because A is an event indicator function of A is a random variable and given B.

So, whatever you know, we have defined earlier, using that we are defining these terms, which you have possibly seen earlier also in some earlier probability courses. And need to just cross verify whether this definition is consistent with those definitions. So, that we will do in the next slide.

(Refer Slide Time: 8:21)

- **Bayes' Theorem** For any $A, B \in \mathcal{F}$,

$$\begin{aligned} \Rightarrow P(A|B)P(B) &= E(1_A|\sigma(1_B))(B)P(B) = E[E(1_A|\sigma(1_B))1_B] \\ &= E[E(1_A 1_B|\sigma(1_B))] = E(1_A 1_B) \\ &= E(1_{A\cap B}) = P(A\cap B) \end{aligned}$$

$\therefore\ P(A|B)P(B) = P(A\cap B)$.

For the same reason $P(B|A)P(A) = P(A\cap B)$. Therefore, $P(A|B)P(B) = P(B|A)P(A)$.

**Conditional Expectation:**

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X$ a random variable on it with $E|X|$ finite. Assume that $\mathcal{G}$ is a sub $\sigma$ algebra of $\mathcal{F}$. Then $E(X|\mathcal{G})$ is a $\mathcal{G}$ measurable function such that for every $A \in \mathcal{G}$,

$$E[E(X|\mathcal{G})1_A] = E[X1_A].$$

If $PX(A) := E[X1_A]$ for all $A \in \mathcal{G}$, then $PX$ is a signed measure on $(\Omega, \mathcal{G})$. Again $P|_{\mathcal{G}}(A) := P(A)$ for all $A \in \mathcal{G}$ is a measure on $(\Omega, \mathcal{G})$. Clearly $PX \ll P|_{\mathcal{G}}$, $PX$ is absolutely continuous w.r.t. $P|_{\mathcal{G}}$.
Then $E(X|\mathcal{G})$ is the Radon-Nikodym derivative of $PX$ w.r.t. $P|_{\mathcal{G}}$,

$$E(X|\mathcal{G}) = \frac{dPX}{dP|_{\mathcal{G}}}.$$

Furthermore, for any $A, B \in \mathcal{F}$,

$$E(X|B) := E(X|\sigma(1_B))(B);$$
$$P(A|B) := E(1_A|B).$$

So, we also get the Bayes' theorem. So, what we want to show that Probability A given B multiplied with Probability B is equal to Probability of A intersection B. Correct? So, this is actually defining notion of Probability of A given B. So, we prove here that thing using the earlier definition.

So, here a probability A given B would be written using this notation. Now here B is the sigma algebra generated by indicator function of B. So, we write down everything here, explicitly. So, it is expectation of 1A given sigma of 1B, (())(09:06) evaluated at B and PB is retained. And then, what we do is that, we consider this is the value and this is the measure on which this value is achieved.

So, then we can think that this value is multiplied with the indicator function of B and that is a function and we are finding out area under the curve, like we are finding out the integration of that, if we do that then we are exactly going to get the value and the measure where the value was achieved. So we do that way.

So, we write down everything multiple 1B and take the expectation. There is a rationale behind getting this thing from here. One can actually cross verify the reverse direction. If you find out it, you are going to get this value multiplied with the probability of B. So, whatever the manner, you justify.

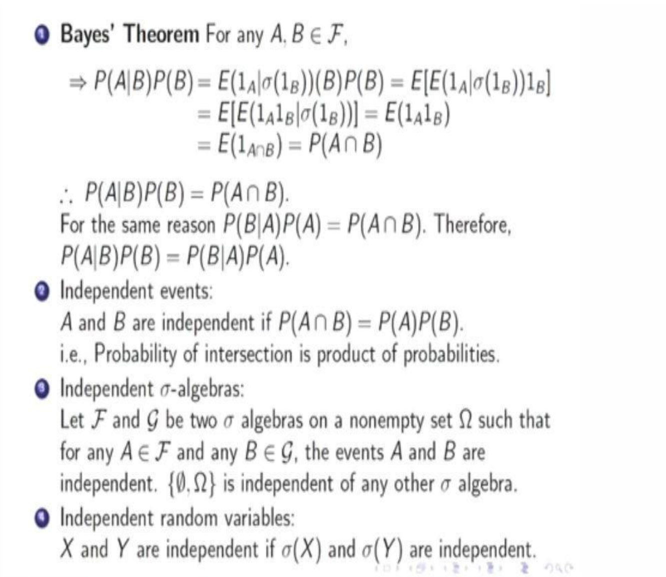
So, after that, what we do is that, we use this fact that since this is measurable with respect to the sigma algebra written here, so we can put it inside. So, this is a property which I am also going to quote separately later. So, what we are going to get is the expectation of this 1B is now with 1A, 1A into 1B. So, this product appears here and the conditional and the sigma algebra appears here.

And then we notice that there are two expectations. So, this first Expectation gives me a random variable having, which is sigma 1B measurable. But, another Expectation would give me just a real number. So, what is that? Expectation of Conditional Expectation is Expectation of the random variable itself.

So, we are going to get Expectation of 1A into 1B. But, 1A into 1 B is nothing but the product of the indicator function of the intersection of A and B. So, indicator function of again indicator function, expectation of indicator function is the probability of the set. So, we are going to get probability of A intersection B.

For the same manner, exactly, because A and B will not have specified any particular thing for symmetry, we are going to get, we can replace A by B and B by A, so we are going to get PB given A product P of A is equal to P of A intersection B. So, you are going to get the P of A given B into PB is equal to P of B given A into B, because A intersection B is same as B intersection A.

(Refer Slide Time: 12:00)

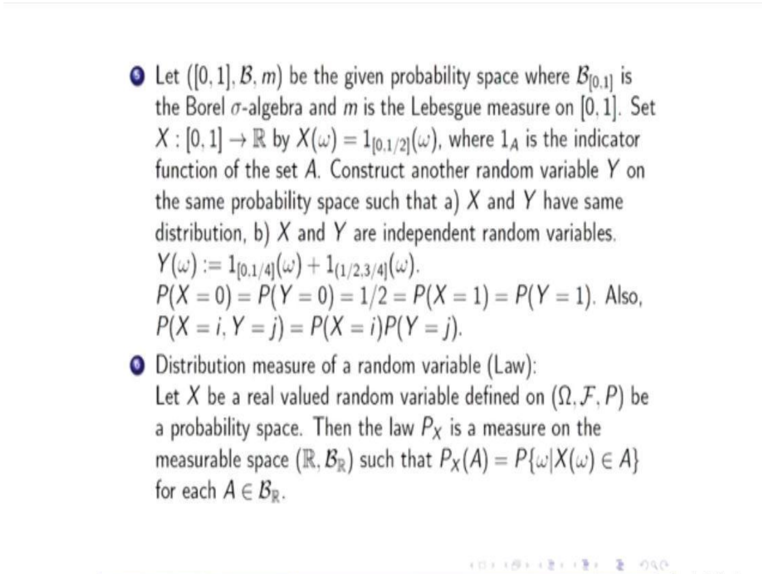



Next, we say, what do you mean by independent events. So ,the definition is that, if A and B are such that probability of A intersection B is product of probabilities P of A into P of B, then we say that A and B are independent. So, probability of intersection is product of probabilities.

Now, we define what we mean by independent sigma algebras. So, let F and G be two sub-sigma algebras on an empty set omega such that, for any A from F and any B from G, the events A and B are independent. So, does not matter what Ai choose and what Bi choose from F and G, but all the time this pair I am going to get independent. So, then we call F and G as independent sigma algebras. So, for example, if I take the trivial sigma algebra, (())(13:15) so like empty set and the full set, so this sigma algebra is independent of any other sigma algebra. So, it is actual independent sigma algebra (())(13:28)

Now, that would help us to define what do you mean by two independent random variables. X and Y are independent, if the sigma algebra generated by X and the sigma algebra is generated by Y this two are independent sigma algebras. So, that is how we introduce independence of random variables.

(Refer Slide Time: 13:57)




So, here we consider this example to illustrate that independence etc. So, what we do is that, so let, so it is a very concrete example, we take Omega as closed into 0, 1, B is the Borel sigma algebra on 0, 1, m is the Lebesgue measure. And then, (())(14:24), so this triplet is in probability space. So, now a set X is equal to indicator function of 0 to half. So, this is the same example time and again.

So, coming from closed 0 into R, so this indicator function we construct. Now, we construct another random variable Y on the same probability space such that X and Y have same distribution and but X and Y are independent random variables. Because, you know generally when we say that give me an example of two different random variables, two independent random variable, the straightforward answer is, let us consider outcomes of two independent random experiments, tossing coins twice. First time outcome would be X, second time would be Y. So, they are independent.

However, we are not considering the definition of a random variable as an outcome of a random experiment, but a measure of function. So, it is important to produce example of independence in this setting.

So, let us see what we do, we take Y as indicator function of 0 to one fourth and add with indicator function of half to three fourth. So, what we do is that, total measure where Y is 1 is still half, because for one fourth length, it is 1 and again here, for one fourth length it is 1. So, the Y thus constructed is surely a Bernoulli random variable having value 0 and 1 with half and half probabilities.

However, if we, so this line, I just explained, so they have the same distributions, they are identically distributed. However, if we consider the joint distribution, Probability X is equal to i and Y is equal to j, that turns out to be one fourth all the time, whatever i, j you choose, i is equal to 0 or 1, j is equal to 0 or 1, all time the left hand side is one fourth.

And right hand side, it is also half in to half one fourth, it is the way it is construct. So, here this is a construction of a pair of independent random variables. So, if you can define two independent random variables this way, you can actually define a countable collection of independent random variables this way in the similar manner.

So, next we introduce another term, distribution measure of a random variable or Law. What do you mean by that? So, we were talking mostly about the domain of this, Omega, is measurability etc. Now, here when we do not have that you know modeling setting Kolmogorov models setting (())(17:57) but we just consider a random variable as outcome of a random experiment, then we do not see that omega. We just see the realization of the random variable.

And then we ask how often the random variable took that particular value. So, that also gives you a measure. So, this measure is a distribution or Law. So, let us write this down scenario, let us clarify what we mean in this (())(18:23). So, let X be a real value random variable defined on an Omega F, P this probability space. Then the Law Px is a measure on the measurable space R here, that is the range space, such that Px of A is the probability of getting all this omega such that X of omega is inside A, is in A. So, for, so if you if Px satisfies this for all A in BR, then you call P subscript X as the Law of X or the distribution of X.

(Refer Slide Time: 19:08)

- Example: Consider the probability space $([0,1], \mathcal{B}_{[0,1]}, m)$ and $X(\omega) = F^{-1}(\omega)\ \forall\ \omega \in [0,1]$ where

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

Then $P_X(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$

Let us revisit some properties of conditional expectation and conditional probability

- Let $X$ be a random variable defined on $(\Omega, \mathcal{F}, P)$ with finite expectation.

  Then $E(X|\mathcal{G}) = \begin{cases} X, & \text{if } \mathcal{G} = \sigma(X) \\ EX, & \text{if } \mathcal{G} \text{ is independent of } \sigma(X). \end{cases}$
- $E(XY|\mathcal{G}) = XE(Y|\mathcal{G})$ if $X$ is $\mathcal{G}$ measurable.

So, let us look at the example, one example. So, this is also the same example what I have shown earlier for construction of a random variable having normal distribution. So, here we are recollecting that random variable. So, we have the measure probability space as a very you know standard choice, unit interval, Borel's sigma algebra on that and Lebesgue measure and X is the inverse of the CDF. Here CDF is a monotonic function. So it pointwise monotonic function, so there is no problem, the inverse exists and it is given by FX is equal to minus infinity to x, 1 over square of 2 pi, e to the minus t square by 2 dt. So, then what would be PX for this case? So, PX of A is here is the integration of this integral over A. That A is the PX for this case.

So, let us revisit some properties of Conditional Expectation and Conditional Probability. So, some of these properties I have used earlier, but I have never specified those explicitly. So, here I am doing that. So, let X be a random variable defined on the probability space Omega F, P with finite expectation. A finite expectation is important because you know that is the unlinking we are doing to define conditional expectation.

Then conditional expectation of X given G for, I mean it is not a defining criteria I am just saying that, for these special cases, it has very nice expression, When G is sigma algebra generated by X, or superset of that, super sigma algebra of that, for those cases, this conditional expectation X given G is exactly X.

On the other hand, if G is independent of sigma algebra generated by X, then expectation of X given G is just expectation of X. So, this also coincides with the basic intuition that the sigma algebra given here is represented information. So, what is that? So, if your sigma algebra is equal or more than the sigma generated by X, that means it has all the information of X. Then given all the information of X, and I am asking what is the average. So, you are not averaging at all, because you have the information already. So, you get the X itself.

On the other hand, if G is independent of X, sigma X that means G has no information of X. And then, you are trying to find out the expectation of X given this information. But that information is not helping you in any way. So, then you are going just get expectation of X. Say for example, if you choose G to be a trivial sigma algebra, which is of course independent of sigma X, then also you are going to be (())(22:27) that is the special case. So, these are actually, one can be actually proof step by step but for the sake of our purpose of the course, we are not going to visit the proofs, we are just recollecting these facts which we are going to use in the course of matter.

So, this is the second property, which I have used earlier in the proof of you know conditional probabilities. So, here expectation of the product of X and Y given G, it becomes X into expectation of Y given G if X is G measurable. If X is G measurable then you can take X outside of this expectation. So, this thing, you can actually see here in the manifestation of this property also.

Because, if you prove this, this part is trivial. why? Because if G is sigma X, so then X is G measurable, so you can take X outside of expectation then inside, only 1 remains. So, expectation of 1 is 1 always. So, it would be X. So, if one proves this is trivial.

(Refer Slide Time: 23:48)

- Let $\mathcal{G}_1$ and $\mathcal{G}_2$ are sub $\sigma$ algebras of $\mathcal{F}$ and $X$ is a random variable on $(\Omega, \mathcal{F}, P)$. If $X$ is conditionally independent to $\mathcal{G}_1$ given $\mathcal{G}_2$, then $E(X|\sigma(\mathcal{G}_1 \cup \mathcal{G}_2) = E(X|\mathcal{G}_2)$.
- If $X$ and $Z$ are independent. That does not imply that $E(X|Y, Z) = E(X|Y)$.
- If $Y$ is $\mathcal{G}$ measurable, then $E[(X - E(X|\mathcal{G}))(E(X|\mathcal{G}) - Y)] = 0$. If $X$ has finite variance, then the conditional expectation $E(X|\mathcal{G})$ is the projection of a $X$ in the subspace of variables measurable to $\mathcal{G}$.

A few more properties. So, let G1 and G2 are sub-sigma algebras of F and X is a random variable on the probability space (Omega ,F,P). If, X is conditionally independent to G1 given G2, what does it mean? That X belongs to A in that event and a member of G1. So, if I take probability of intersection of these two given G2, then that joint probability conditional probability of the joint event is the product of the conditional probabilities. The probability of X belongs to A given G2 multiplied with probability of that event from G1 given G2.

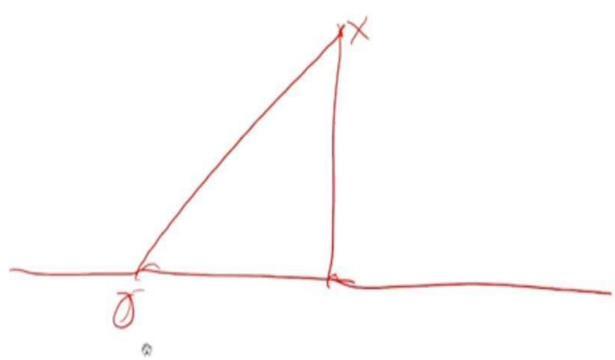
So, if that is the case, then why not get expectation conditional expectation of X given, here you have G1 and G2 together. So, what do I have? I have just X and G1 are independent and we know that X and G1 are independent given G2 and then we are asking that what is the conditional expectation of X given sigma algebra generated by G1 and G2 together. So that turns out to be the expectation of X given G2. So, here as if that since X you know is independent with G1, so you can ignore G1 as if that thing happens here.

However, if you have X and Z are independent, X and Z are independent, but you do not say that, whether they are conditionally independent with respect toY, do not put that condition. So, then it you do not get, that does not imply that expectation X given Y, Z is equal to expectation of X given Y, it does not imply this. So, you need that independence, that conditional independence, you need conditional independence then only you can assure this.

So, now we see that what is the implication of conditional expectation of X given a sigma algebra G. So, this implication is useful when X has finite variance. So, we consider here, say Y is a random variable, which is G measurable, any arbitrary. And then, we consider this I mean the product of the expectation. The product is taking the following manner. We take conditional expectation X given G and we take, we subtract conditional expectation of X given G from X. So we get a random variable there.

And again, we consider conditional expectation of X given G and minus Y. So, here Y is G measurable, conditional expectation of X given G is also G measurable, so their subtraction is also G measurable. So, this random variable is in the sub-space of G measurable functions. And then, if we take our expectation of this product, which is nothing but inner product in the L2 space, so this turns out to be 0. I am not proving this, but I am just quoting this result, this is 0. An I also I am highlighting the implication of this fact. So, conditional expectation of this product is 0. So, if X is finite variance, then the conditional expectation of X given G, these now can be considered as projection of X in the subspace of variables measurable to G.

(Refer Slide Time: 28:22)



3. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ are sub $\sigma$ algebras of $\mathcal{F}$ and $X$ is a random variable on $(\Omega, \mathcal{F}, P)$. If $X$ is conditionally independent to $\mathcal{G}_1$ given $\mathcal{G}_2$, then $E(X|\sigma(\mathcal{G}_1 \cup \mathcal{G}_2) = E(X|\mathcal{G}_2)$.
4. If $X$ and $Z$ are independent. That does not imply that $E(X|Y, Z) = E(X|Y)$.
5. If $Y$ is $\mathcal{G}$ measurable, then $E[(X - E(X|\mathcal{G}))(E(X|\mathcal{G}) - Y)] = 0$. If $X$ has finite variance, then the conditional expectation $E(X|\mathcal{G})$ is the projection of a $X$ in the subspace of variables measurable to $\mathcal{G}$.

So, let us explain this with this drawing. Consider that this whole plane is the L2 space of random variables, which have finite variance and imagine this line is signifies or you know corresponding to the sub-space of G measurable random variables. So, G measurable L2 random variables. And then consider Y is one such random variable. And here expectation of X given G is on the sub-space, expectation X given G is on the sub-space. We consider this as the point. This point corresponds to expectation X given G, then difference X minus expectation X given G is this part and expectation X given G minus Y is this part.

And this equality says that expectation of this product is 0 or saying that the inner product of this is 0, that means they are orthogonal. So, that means they are orthogonal, that means that this is projection of X on this sub-space of G measurable L2 of random variables. So, by this I stop now. Thank you.