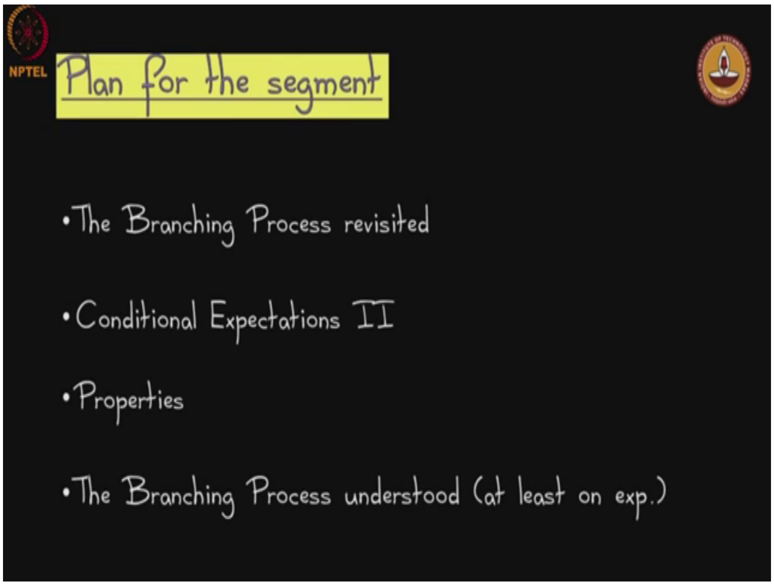
Probability & Computing
Prof. John Augustine
Department of Science and Engineering
Indian Institute of Science, Madras

Module – 02
Discrete Random Variables
Lecture - 12
Conditional Expectation II

So, let us start with the segment 4 of module 2. Here we will be talking about conditional expectation as a variation of what we have already looked at. So, I call it conditional expectation II.

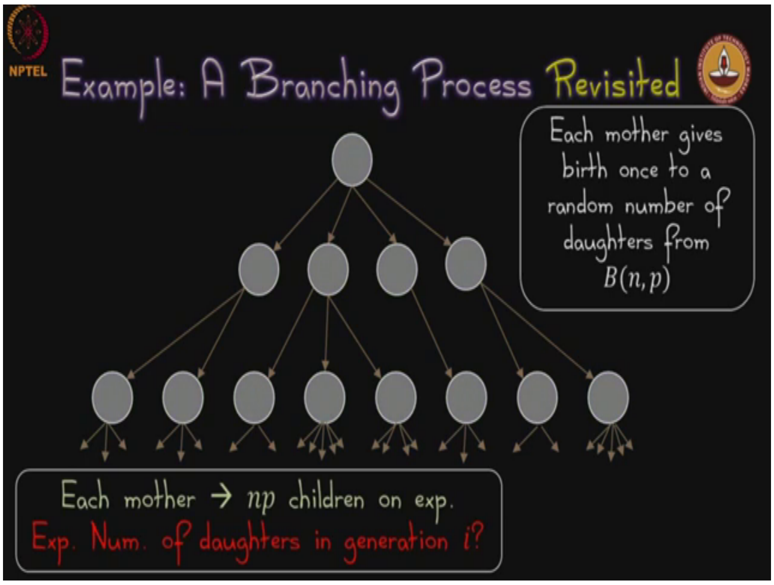
(Refer Slide Time: 00:30)



What is the plan for the segment? We will just revisit this branching process that we talked about in the last segment.

And then in order to understand this branching process we particularly want to know the expected number of children at an arbitrary generation I and the to understand that question we will introduce this notion of conditional expectation version 2 if you will, will describe some properties of that notion. And then are get back to the branching process and try to understand that question.

(Refer Slide Time: 01:03)

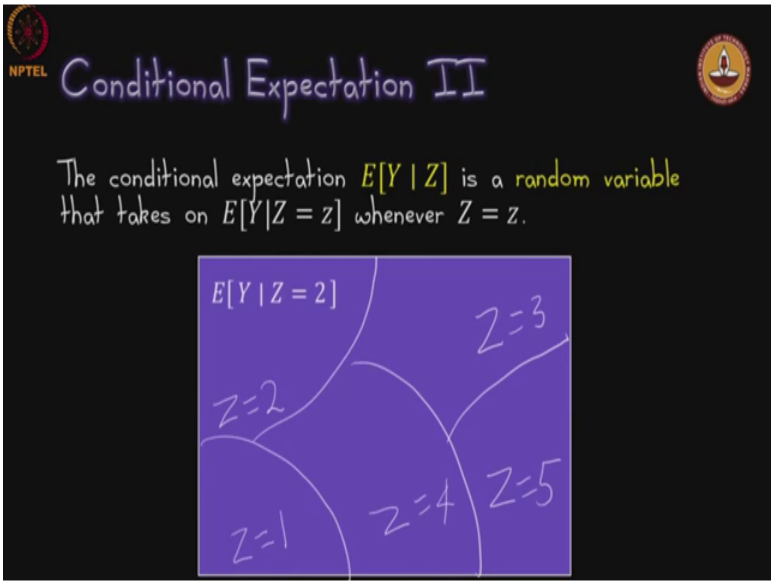


So, let us a quick reminder of the branching process. So, we start with one mother if you will and that mother gives birth to some number of daughters and those daughters become mothers and give birth to their own daughters and so on and so forth. And the way we are modelling it each mother gives birth once to a random number of daughters, and the random number is chosen from the binomial distribution with parameters n and p. Each mother we know one expectation gives birth to n times p children. So, even you also know for example, if there are k mothers in 1 generation in the next generation the expected number is k times n times p.

But what we do not know is how to generalize it to an arbitrary generation I ok, intuitively the first generation there will be n p daughters on expectation and those n p in turn will produce n p, each will produce n p daughters. So, it will be n p raised to the power two in the second generation and so on and so forth. So, an arbitrary I th the generation intuitively it n p to the I.

We want to make sure that we can we have a clear formal basis for this intuition ok.

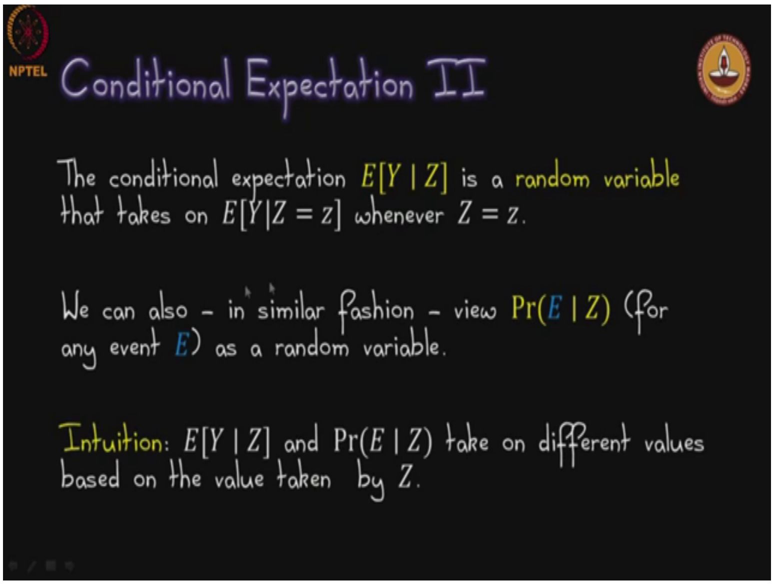
(Refer Slide Time: 02:19)



So, would, that brings us to this notion of conditional expectation and the second version in which we defined this expectation of Y given a random variable Z. So, this expectation of Y given a random variable Z itself is a random variable and takes the value expectation of Y, given a specific value of Z this lowercase Z whenever the random variable uppercase Z takes the value lowercase z.

So, if you think of the sample space as being divided into regions. So, with various regions corresponding to various values of a Z each region if you condition on that region you would get an expectation of the random variable Y and that expectation itself is now a random variable over the entire sample space right and that is this expectation of Y given Z.

(Refer Slide Time: 03:17)

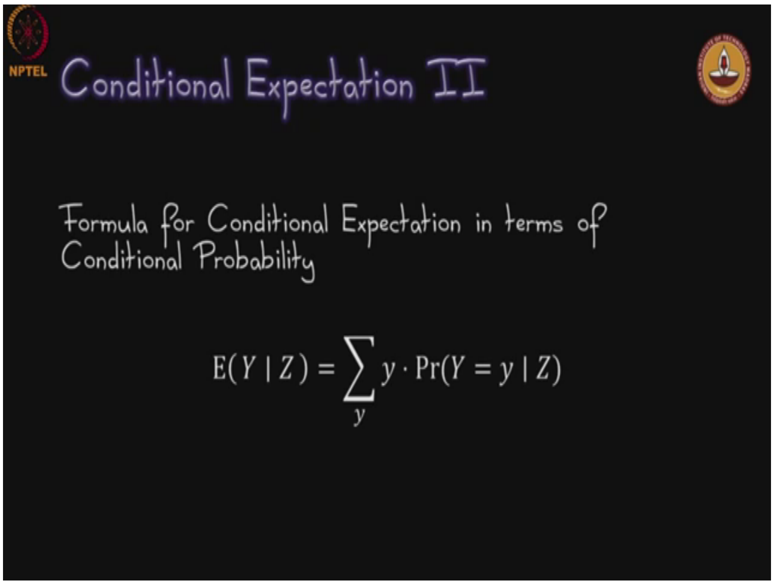


In a similar fashion we can also view the probability of some event E over this random variable Z.

So, going back to this picture again Z let us say it divides up the sample space into regions and in each region you have a certain conditional probability of some even E and that now becomes a random variable and that is what we will denote by probability of E given this random variable set.

So, as we talked about this E of Y and given Z and probability of E given Z they take on different values based on the a value is taken by Z, in this that is their random variables.

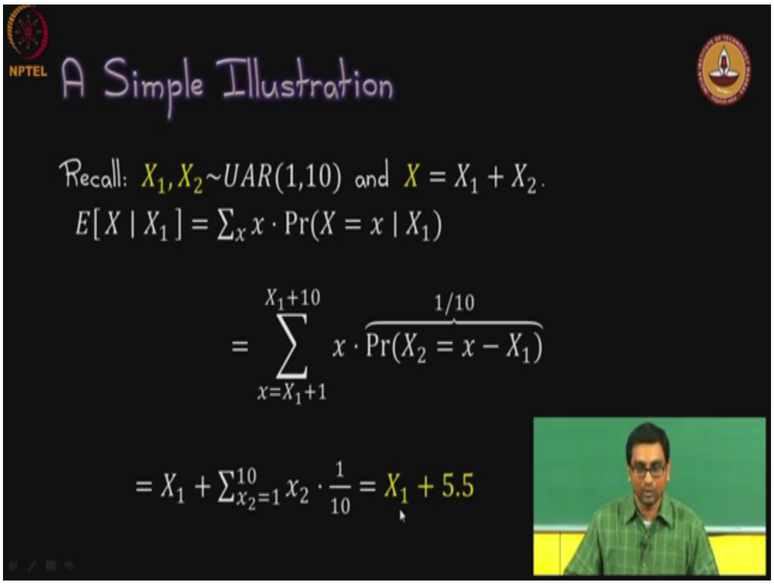
(Refer Slide Time: 04:04)



Let us work out a formula for the conditional expectation in terms of this conditional probability, we are asking what is the expectation of Y given Z.

Now, it is an expectation, so we will try to work through some through all possible values of Y. And in each case its very simple we just use the conditional probability of Y taking the specific value of lowercase y conditioned on this z. So, it is a very natural formula that we can have for expectation condition expectation Y conditioned on Z.

(Refer Slide Time: 04:44)



So, let us work out a simple example hopefully this will be helpful in making the idea concrete. So, recall we have already talked about this example where we have to run two random numbers generated uniformly at random between 1 and 10 ok. So, we call them X 1 and X 2 and capital X is the sum of X 1 and X 2.

So, now we ask what is the expectation of this capital X given this lowercase X 1 and we can apply the formula. So, this is simply the formula that I had mentioned in the previous slide you are looking for the expectation of X. So, you have to run through all the values that X can take and then we weighted by their individual probabilities of course, each conditioned on X 1, whatever; that means, let us see what that works out.

So, now what are the values that this uppercase X can take? Well, it is hinging on X 1. So, if X whatever X 1 takes it can take a value X 1 plus 1 all the way up to X 1 plus 10. So, this lowercase x therefore, runs from X 1 plus 1 to X 1 plus 10 and of course, here you have to sum the individual elements weighted by their probabilities. Now, what is this probability that x equals this lowercase x conditioned on some X 1? What is this basically talking about?

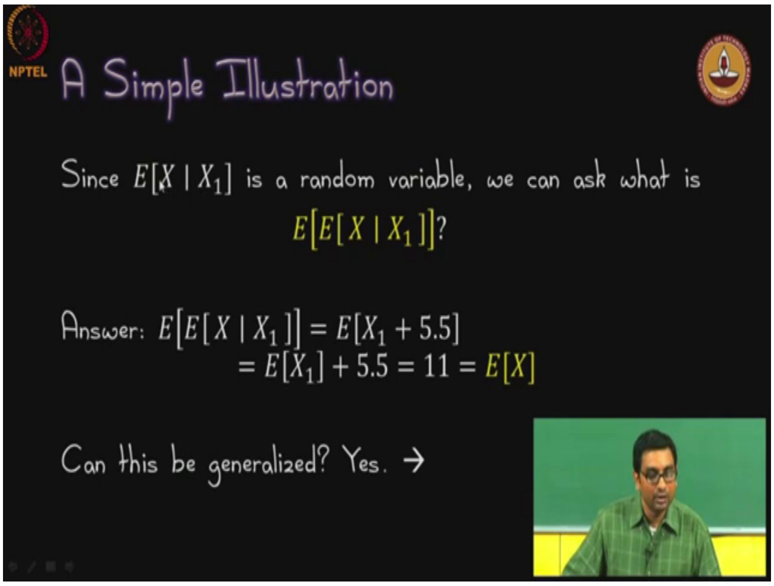
So, this uppercase X has to take on a specific value of x a conditioned on the value of X 1 ok. In other words when will that happen when X 2 takes the value of, so basically what should happen is x another way to write this is x should be equal to this lowercase x should be equal to X 1 plus X 2 right and for that specific value of X 2 the for the event that x that happens when this X 2 takes this value little lowercase x minus X 1. What is the probability of that specific event happening it is going to be 1 over 10.

So, this whole probability is basically 1 over 10. So, this is. So, now, you can expand this out. So, if you work out this sum what is going to happen over here? You are going to run from values of x is going to range from X 1 plus 1 to X 1 plus 10. So, you are going to have X 1 added 10 times, but each of those times it is going to be weighted by 1 over 10. So, you are going to have and it is going to average out to just X 1 plus this second you are adding a quantity here X 1 plus 1 all the way to the X 1 plus 10 right, you can think of it as another I, if you can think of it as an value I running from 1 to 10 or rather X 2, X 2 the running from 1 to 10 that also will take the values is weighted by the probability is 1 by 10.

So, this the second term if you think about it, it is this basically summation 1 through 10 X 2 times 1 over 10 that is going to work out to 5.5. So, what is it ultimately going to be? This expectation is going to be X 1 plus 5.5. You can work out from the details, but look let us look at the intuition here we are asking what is the expectation of uppercase X given some lowercase X 1 and it is this is basically remember this is a random variable.

So, you on the right hand side you cannot have a constant, you will have to again have a random variable on the right hand side the random variable is X 1 and over and above that random variable you going to add a 5.5 and that makes a lot of sense because now whatever your X 1 value is your x is going to be on expectation 5.5 plus that original value of X 1 ok. So, this hopefully the intuition at least is very clear. I would recommend that you go through the details to convince yourself ok.

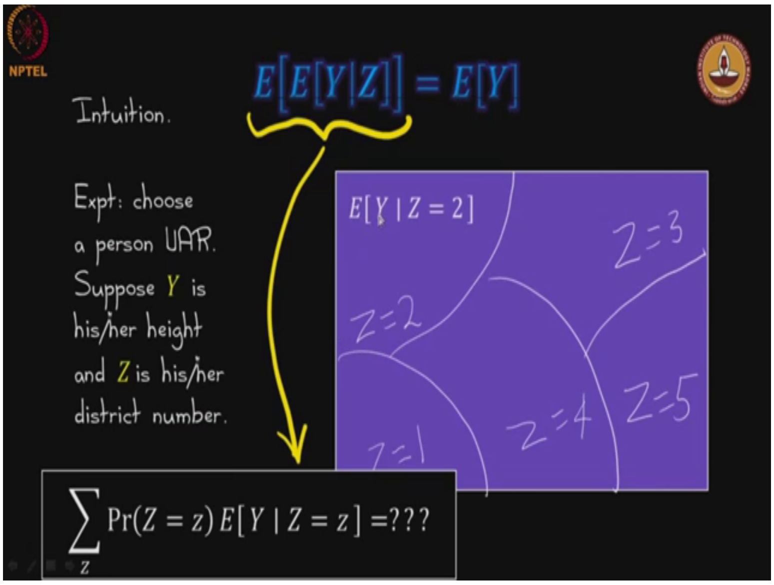
(Refer Slide Time: 09:09)



Here we notice that this expectation of X given X 1 or any of these sort of conditional expectations is itself a random variable. So, it is a meaningful question to ask is what is the expectation of the expectation of X given X 1. So, what will that be? Well, let us work it out in this example. What is the expectation? Well, we already worked out that expectation of X given X 1 is X 1 plus 5.5. So, we can apply plug that into the interior of this expectation.

And now we get two terms. So, now we can apply linearity of expectation. So, it is going to be expectation of X 1 plus the expectation of 5.5 which is going to be just 5.5. What is expectation of X 1? It is going to be again 5.5, remember it is a uniform random number between 1 and 10, so it is going to be 5.5, its going to work out to 11. And if you think about it, it is also the expectation of uppercase x why because uppercase X is X 1 plus X 2. So, expectation of uppercase X is expectation of X 1 which is 5.5 plus expectation of X 2 which is 5.5. So, it is that is also going to be 1 and so there is this when you take the expectation of the expectation what you end up getting is just the expectation of the original random variable X, ok.

(Refer Slide Time: 10:39)



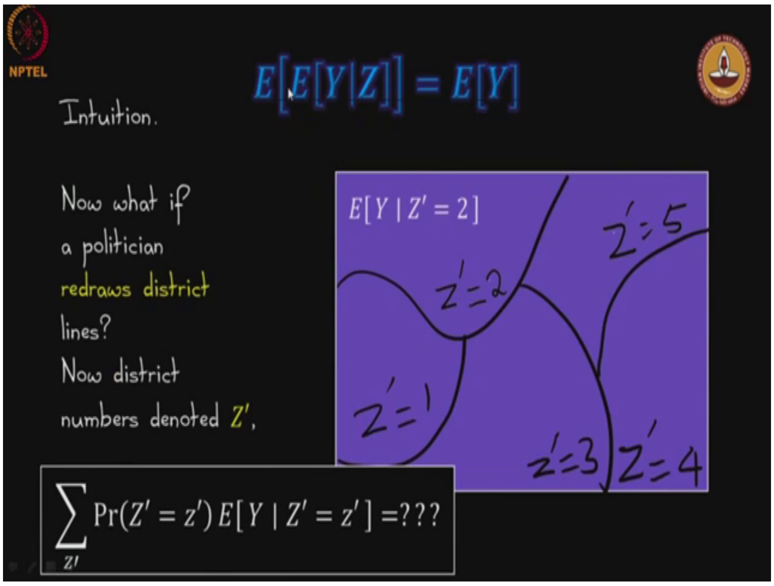
So, this is true from the example. The question is can this be generalized and as you would expect, yes you can generalize it.

So, in general what we are claiming here is if you remember expectation of Y given Z is a random variable, if you take the expectation of that random variable you get back expectation of Y. So, I am going to give you an illustrative way to look at it at least this helped me to look at it and make some sense out of it. So, maybe this will help you use well and then we will formalize it as well, ok.

So, what I am going to think of is this experiment. So, we have a geographic area and we are looking at people in this geography. So, this is think of this sample space as people in some region and the experiment is to choose a person uniformly at random and each person the that you choose Y corresponds to the height of that person. And Z corresponds to this district number. So, you have various districts at say in this region. So, now, what we what you can think of is expectation of Y given Z equal to 2 is basically restrict to yourself to the second district and you ask what is the expected height of the people in the second district.

So, now we ask in this context we ask what is expectation of Y given Z the expectation of that, ok. If this is this is an expectation of a random variable right. Let us look at what this is about. Each of these regions has a certain Z value associated with it and each of them has a value of the expectation of Y given that Z value. So, when we want to take the expectation what do we do? For each of these regions we take the expectation of Y given Z equal to that particular Z value and then we wait that by the probability that Z equals that value and we sum it over all possible Z values that is that will give us the left hand side. The way to think about it is you look at each one of these is a district, you look at the average height per for each district, but then you weight it by the size of those districts that is what you have on the left hand side.

(Refer Slide Time: 13:01)



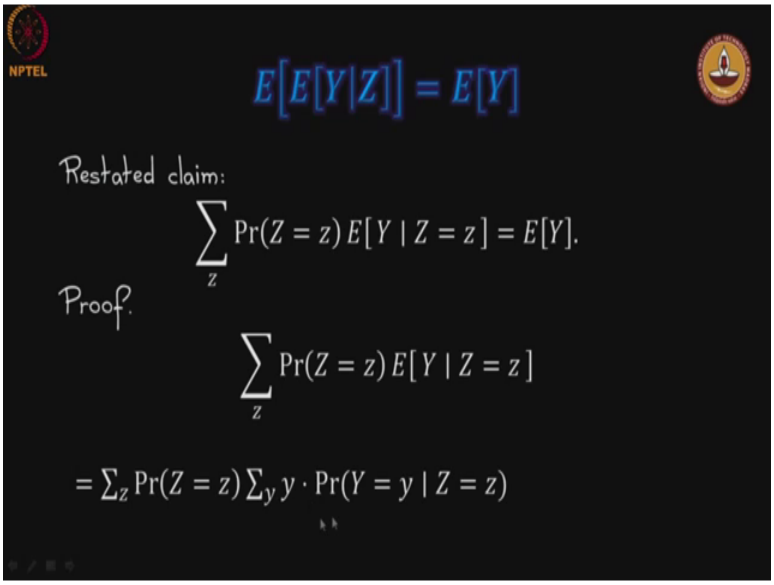
Now, let us play this intuitive thing. What happens if someone comes and redraws the districts? Some politician comes redraw districts, so now, (Refer Time: 13:08) Z it is Z prime. So, now, again the same you can still ask what is the expectation of Y, now let us say given Z prime ok. So, that should be a Z prime over there. And intuitively again now what you are doing? You are doing this weighted height of people across this new districting.

And in the must there is some intuition that should tell you that look these are is essentially the same ok. These are just weighted slightly differently based on two different politicians and you know or whatnot, but it is essentially there referring to the same quantity and that is what is on the right hand side expectation of Y.

Student: Only have (Refer Time: 13:48) because expectation of (Refer Time: 13:50) the intuition.

Yes, intuition, yes that is correct. So, let us, but let us work it out more formally now.

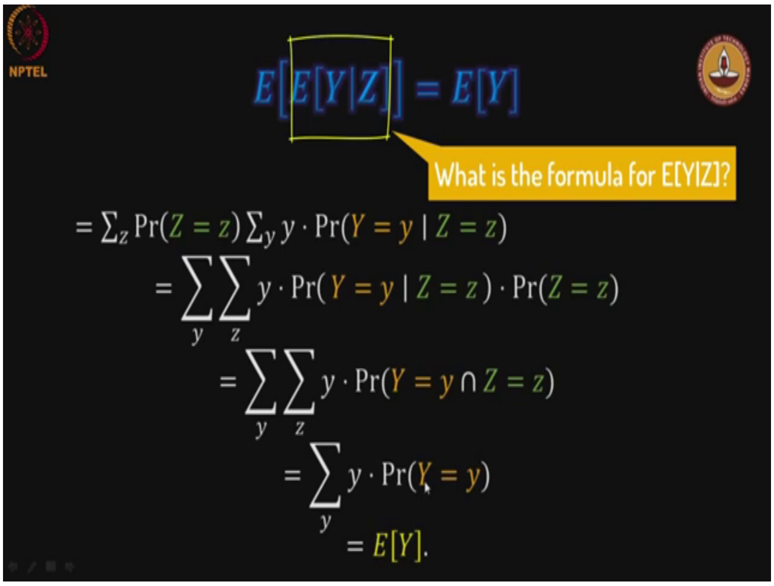
(Refer Slide Time: 13:58)



Let us try to work out this intuition ok. So, let us restate the claim. So, what we are claiming is that this left hand side remember we worked it out as this formula expectation of Y given specific values of Z weighted by the probabilities of Z taking various values ok. And on the right hand side we want to claim that that is equal to this expectation of Y, all right.

So, we will take the left hand side we have this expectation of Y given a specific value of Z. So, we can expand that out we have a formula for that. So, that is basically expectation of Y. So, you have to run through all possible values that Y can take and weighted by these conditional probabilities, sum them up and weight them by the conditional probabilities I am just restating that here. So, now what we are going to do is, so I am going to interchange the summations I am going to bring the summation Y outside and I am taking this summation of I mean this probability of various values of Z, inside the inner summation. So, I get this expression.

(Refer Slide Time: 14:46)



Here what are we doing? So, this is basically let us go back here what we have here is a conditional probability, probability of Y equal to the specific value of lowercase y conditioned on this value of Z times the probability that Z will take that value. This is simply nothing, but the probability of Y equal to y intersected with Z equal to z, this is just a conditional probability formula. Now, how do we get this term?

Student: Summation over this.

Yes. So, this is summation over all Z. So, this is what property.

Student: Law of total property.

Law of total probability, right. So, you are considering all values have Z. So, this is going to cover the entire sample space. So, this is just the law of total probability you will get summation over y basically weighted by the probabilities of various values of y and this of course, is nothing, but the expectation of Y.

Student: can you just what is the initial expression for E of Y given Z bigger.

E of, E of Y given.

Student: (Refer Time: 16:12) you cannot (Refer Time: 16:15) Z equal Z in that way like when you do like you cannot say that sigma E of Y given Z is given sigma Y (Refer Time: 16:18)

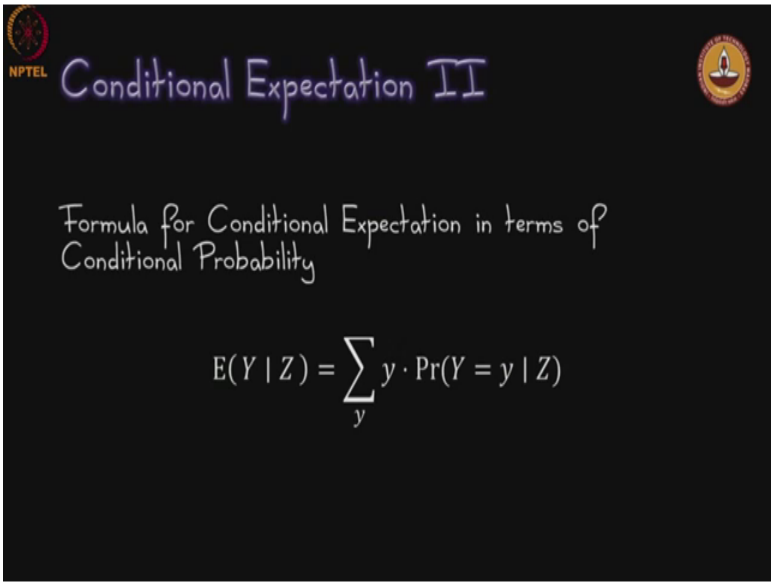
(Refer Time: 16:18), let us let us go back are you talking about this entire.

Student: I am talking about the guy inside the first expectation.

So, for that we have this other expression.

Student: Yeah. So, how did you how did we convert that into Z equal to z expression in the (Refer Time: 16:34).

(Refer Slide Time: 16:27)

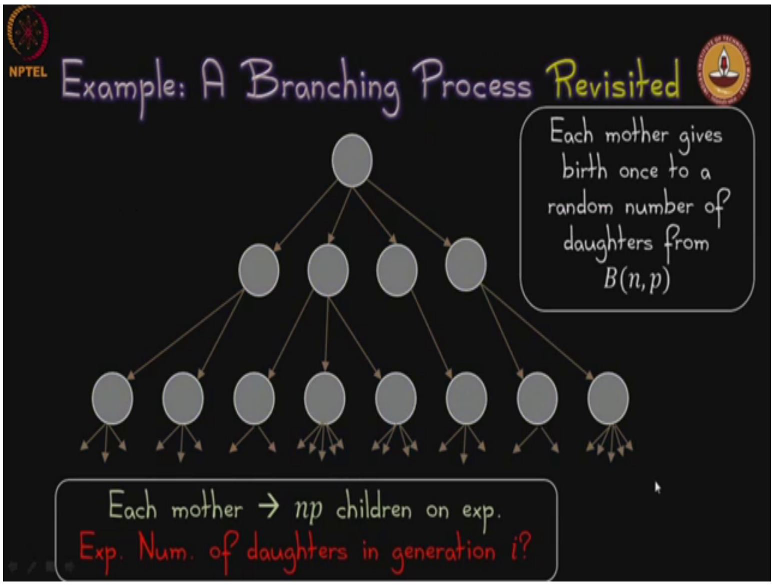


Let us look at that so, but now we are looking at this expression expectation of this random variable, right. So, now think of it this way. Look there are these what is this is an expectation of a random variable. So, what is inside is a random variable ok. What is that random variable going to take on, for each district if you will each event if you will it is going to take on a value ok. What is that value? It is going to take on expectation of Y given that specific value of Z

So, in this district the its going to take on the value of the height of people in that district that is expectation of Y given that that district number is 2. For each district it is going to take on a different value ok. So, now, that is this random variable inside E of Y given Z.

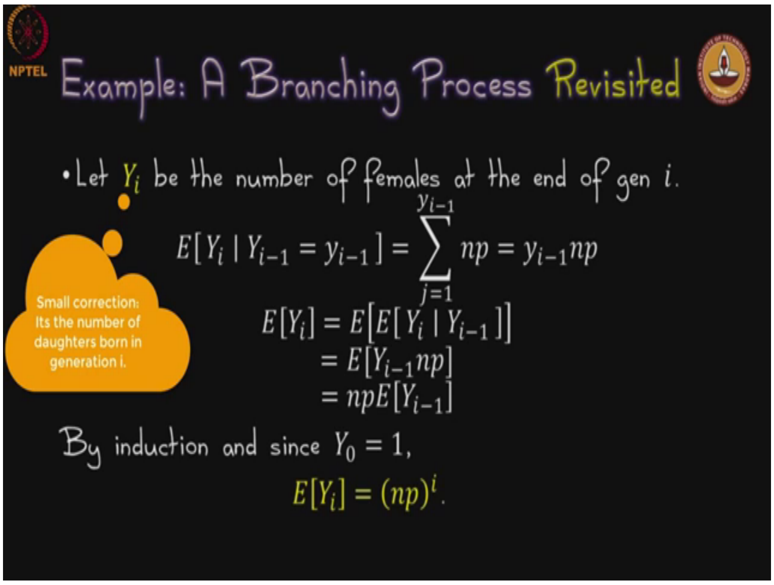
So, when we take the expectation of that we sum up all those values. So, do not for a minute forget that this is expectation it is you are summing up over all these values, but then you have to wait them by the district probabilities.

(Refer Slide Time: 17:47)



Now, let us go back to the branching process this is interesting because now what we have done is we built the machinery to address this question with this branching process. So, now, we want to know at an arbitrary generation I, what is the expected number without conditioning on the previous generation. If we know the previous generation we know, we already had the tools and techniques to answer that question.

(Refer Slide Time: 18:15)



If it if we cannot condition on the previous generation alone what do we do let us see how we can work this out. So, let us use the random variable Y i to denote the number of females at the end of generation i. So, now we can ask what is this is something we already know how to do, it should not be is to surprising here. What is the expectation of some Y i? As that is the number of females at the end of some generation given in the previous generation you had some certain specific number.

So, remember this is a very specific number, this is a random variable, but this is a specific number. So, then if you know that in the previous generation you had a specific number of females then you know that each one of those females is going to have on expectation n p daughters. So, what is it the total number of daughters at the end of the ith generation? It is the number of females in the previous generation times n times p ok. This should be straightforward. So, now, let us let us hang on to that.

What we want is expectation of Y i and notice there is no conditioning her, we just want to look at the ith generation and ask how many what is the expected number of daughters at the end of the ith generation. And this is where we go back to this claim that we have here ok. What we are looking at is the right hand side and I am going to apply the left hand side.

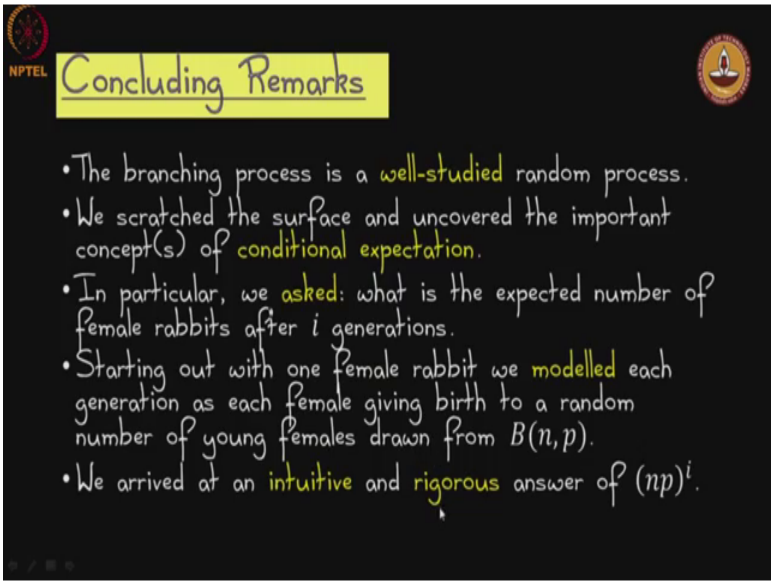
So, now we know that expectation of Y i is nothing, but the expectation of the expectation of Y i given Y i minus 1 ok. This is just applying that claim that we worked up just now just in the opposite direction. And this inner part expectation of Y i given Y i minus 1 intuitively is nothing, but Y i whatever the previous number was this let us leave that as a random variable ok. In the previous generation it was Y i minus 1 times n p ok. I am not making it specific times n p. But now the n p can come out because that is just those are just parameters numbers. So, they can come out by linearity of expectation. So, n times p times expectation of Y i minus 1 ok.

So, what we have done is made an interesting jump in the sense that we made one step of an inductive argument, basically what we are saying is if we can represent the number of females in the in the previous generation by a random variable Y i minus 1, the number of females in the current generation is going to be n p times the expectation of Y i minus 1.

Now, what does this mean? And this we can remember this we could do only because we had we understood we took the time to understand this knows the second notion of conditional expectation and we claim that that notion of expectation of that notion has this left hand side idea. So, now, if you let us look at what you have done in the ith gen the expected number of females in the ith generation is n p times the expected number of females in the previous generation ok.

So, then it will be n p the whole squared times the number of females in the, in the grandmother generation and so on. So, at and what we know is at the very start of this branching process the we started with one female and so Y 0 if you will actually just has to be a Y 1, Y 1 if you will is 1. At the end of the first; I guess it depends on how you start counting the generation, but at the very at the very beginning there is just one female. So, you can think of this as basically an inductive argument in which the base case is that initially you start off with one female. So, if you substitute you are going to get the expectation of Y i to be n p the whole to power i. Basically understood at least on expectation this branching process.

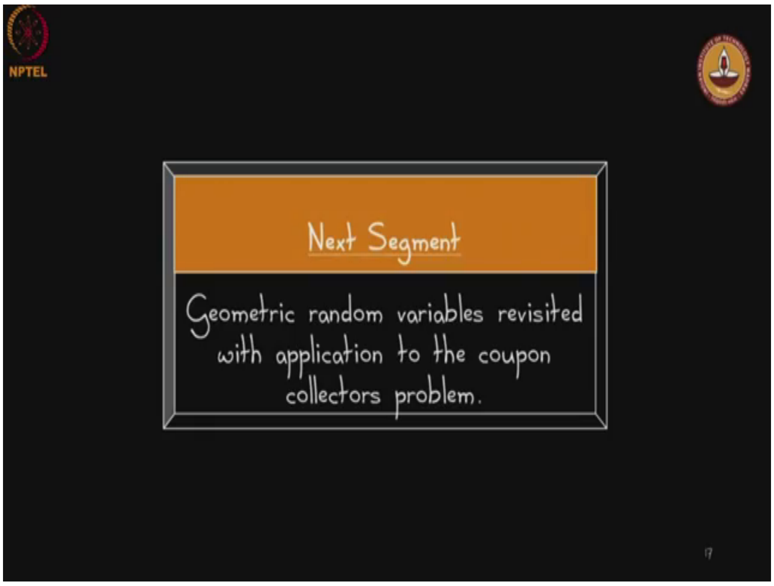
(Refer Slide Time: 22:18)



We as a in this segment we studied the second notion of conditional expectation we applied that to this branching process to understand this discussion of the expected number of female rabbits after ith generations. What we did importantly is we confirmed an intuition. We have an intuitive sense of how many accepted number of females we will have in the ith generation, now we have a way to dot the i's and dash the t's and claim that our intuition is in fact, rigorously correct.

So, with that we conclude this segment.

(Refer Slide Time: 22:55)



In the next segment we will be talking about geometric random variables which we have already briefly talked about. And we will apply that to a very very important context called the coupon collectors problem, and with that we will end this segment.