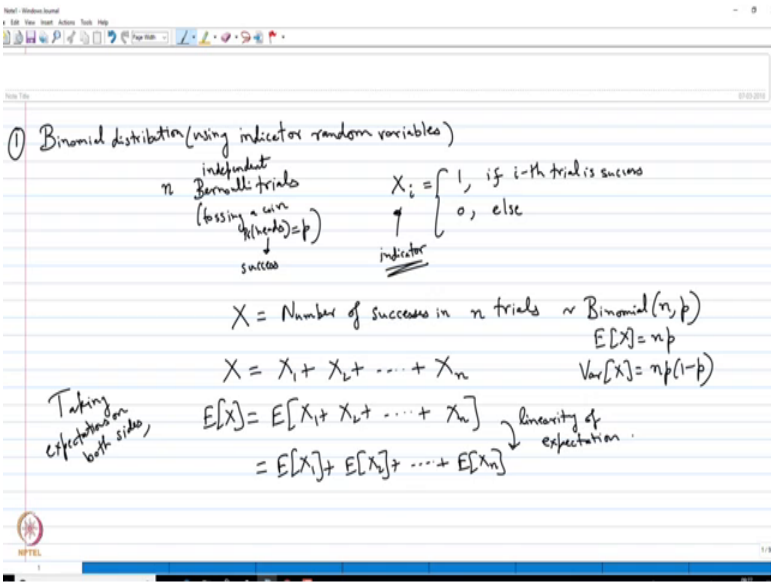
Probability Foundations for Electrical Engineers
Prof. Andrew Thangaraj
Department of Electrical Engineering
Indian Institute of Technology, Madras

Lecture – 72
Examples: Discrete Distributions

Hello and welcome to this lecture on examples with involving expectations for discrete random variables. I will primarily use indicator random variables and show you two examples where the property of linearity of expectation; along with the use of indicator random variables simplify some of the computations ok.

So, this is a powerful sort of method it is used across the board in various problems. Hopefully it will be of interest to you and you will enjoy. Ok, let us get started ok.

(Refer Slide Time: 00:52)



So, the first example we look at is actually a very simple example. So, let us say, it is like it is its actually binomial distribution using indicator random variables ok. Let us we will see it is a very simple sort of idea ok. So, put this in brackets.

So, basically what I will do is I will define, I will do n Bernoulli trials. What is a Bernoulli trial? You can think of it as tossing a coin probability of heads equals let us say p n Bernoulli trials and heads is the ok. So, I will say a probability of heads, heads is basically success ok. So, I am doing n independent Bernoulli trials. So, those are all things very important to write down, independent Bernoulli trials and each trial has a success probability of p ok.

So, let us say, so X i is 1 if ith trial is success that is 0 otherwise ok. So, what do I mean by this? I am doing n coin tosses if the ith toss is a head I am going to say this X i random variable is 1 otherwise I am going to say its 0 ok. So, this is an indicator random variable right, ok. So, it indicates whether or not the ith trial was successful ok. So, this X i is an indicator random variable it indicates whether or not the ith trial was the success ok.

So, now, what is my binomial random variable? So, I am going to define X as the number of successes in n trials ok. So, in these n Bernoulli trials how many successes I had that I denote as X. We know very well that this is a binomial with parameters n and p ok. So, we have studied this several times before and you might also remember or know that this X can be written as X 1 plus X 2 plus so on till X n ok.

So, if you add up these n indicator random variables you end up getting the binomial random variable which is the number of successes. You can see clearly right. So, X 1 is 0 means it was not a success X 1 is 1 means it was a success. So, every time you have a 1 in the sum it indicates the success in that trial. So, you add up all these random variables you will get the number of successes ok. So, the number of successes can also be written like this ok. It is a very simple idea.

So, now, we know very well that you know the expectation of X for a Bernoulli random variable is n times p.

So, these are formulae that you might remember and variance of X is n times p times 1 minus p ok. So, this is something you might remember from the time you studied binomial distribution. So, one can use these indicator random variables this writing of X in terms of the sum of n independent indicator random variables to quickly establish these kind of relationships, ok.

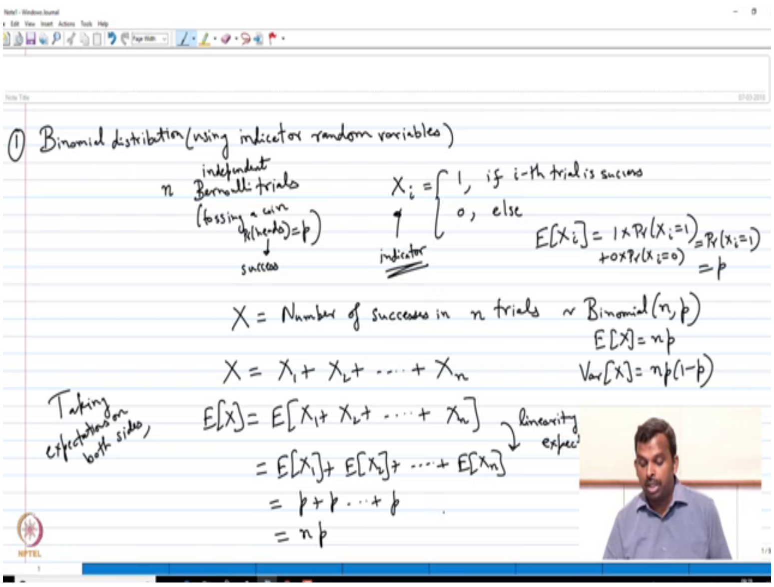
So, I am going to start by showing you how that is done. So, if you want to do expected value of X ok. So, expectation you can think of it as like some sort of an operator ok. So, I have a equation involving random variables I can take expectations on both sides ok, so taking expectations on both sides ok.

So, you can see expected value of X is written as expected value of X 1 plus X 2 plus X n ok. So, now, this easy enough to write, but it turns out expected value is a linear operator ok. So, it distributes over addition. So, if you have expected value of X 1 plus X 2 it is the same as expected value of X 1 plus expected value of X 2. So, you use linearity of expectation to get this ok.

So, here I have used linearity of expectation ok. So, now, in this case X 1 and X 2 and all that are additionally independent, but that is not really needed even if the X i's are not independent you can do this in the next example we will see will exploit that situation, but right now anyway linearity works either way. So, we do that.

So, now, what is the big advantage with this kind of thing?

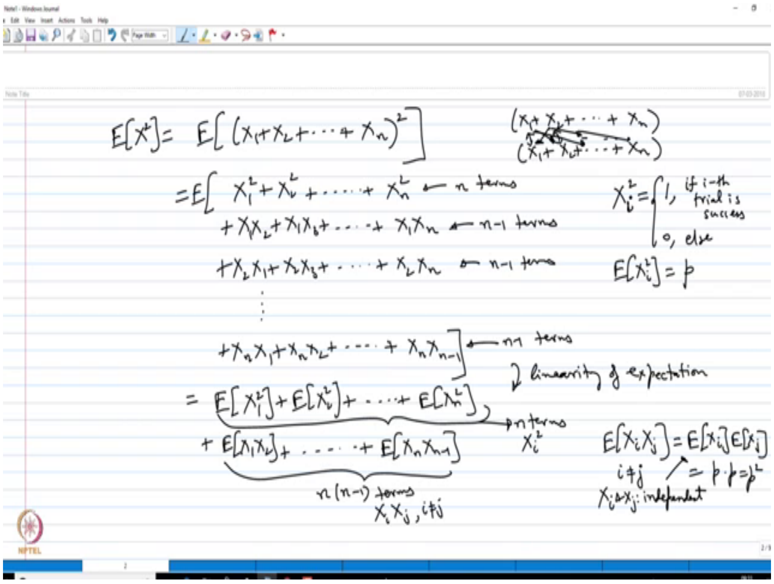
(Refer Slide Time: 06:05)



The reason is if you if you want to find expected value of X i it is very very easy it is a random variable that takes just two values 0 and 1. So, it is 1 times probability that X i is 1 plus 0 times probability that X i is 0 ok.

So, this simply evaluates to probability that X i is 1 what is the probability of success and that is just p ok. So, expected value of X i is p for every i, ok. So, this is just simply p plus p plus p n times and that is just n times p ok. So, very quick definitely derivation for the mean value or expected value of binomial distribution.

(Refer Slide Time: 06:50)



So, what about variance? For variance you have to do a little bit more of work, ok. So, if you want to do expected value of X squared, you will get expected value of X 1 plus X 2 plus so on till X n whole squared ok.

So, now the squaring is not a linear operation I cannot really distribute over the multiplication, but I can expand the squaring. How do you expand the squaring? So, you write it as X 1 squared plus X 2 squared plus so on till X n squared ok. So, you just multiply these terms out. Then what are the additional terms you will get? You will get X 1 times X 2 plus X 1 times X 3 plus so on till X 1 times X n remember I am multiplying these two things X 1 plus X 2 plus X n times X 1 plus X 2 plus X n ok.

So, you will get these terms that gives you the square terms X 1 square X 2 square and then X 1 is multiplied X 2 X 1 will multiply X 3 so on till X n that will give you the next term and then X 2 will multiply X 1 X 2 will multiply X 3 etcetera. So, you will get X 2 X 1 plus X 2 X 3 plus so on till X 2 X n ok.

So, the point is there are n terms here how many terms are here n minus 1 terms ok. So, X 1 multiplies everything other than X 1 itself. So, likewise here there are n minus 1 terms. So, this will go on and on till you have X n X n, X 1 X n X 2 plus dot dot dot X n X n minus 1. So, this is again n minus 1 terms and then you can put a closed bracket here to indicate that this is expectation ok.

So, when you multiply it out you get something like this. So, now, we distribute the expectation ok. So, this becomes expected value of X 1 squared plus the expected value of X 2 squared plus so on till expected value of X n squared. So, here I use linearity of expectation and then you have all these terms X 1 X 2 plus dot dot dot I will write all the way down to the last term which is X n X n minus 1.

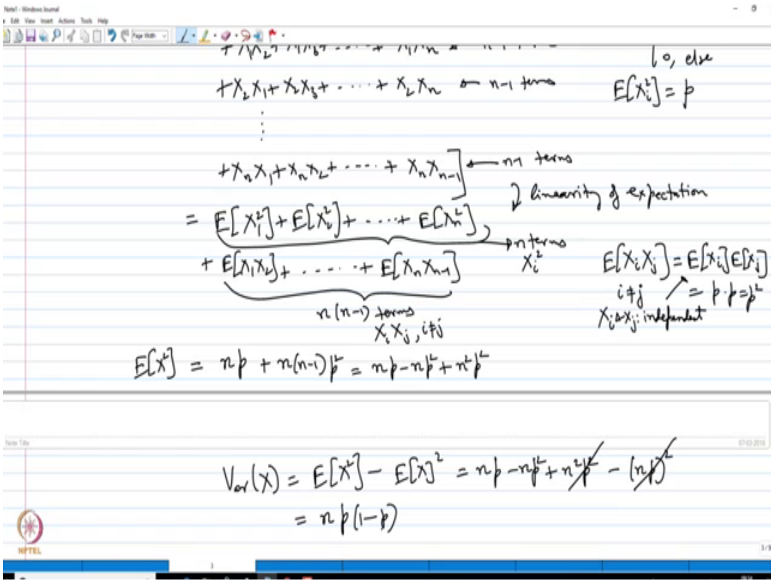
How many such terms do I have here? So, here I have n terms here I have n times n minus 1 terms right all of them together I can go and they used linearity of expectation. So, I already get n times n minus 1. So, you see there are n minus 1 terms here and then n minus 1 terms here so on till n minus 1 terms here. How many such things are there? m from X 1 X 2 to X n. So, we have n of those. So, n times n minus 1 ok.

So, if you come if you look at X 1 squared remember X 1 is just an indicator random variable it takes value 0 with probability something one with probability something. Now, if you square it you will again get a the same sort of random variable right. So, it is going to be one. So, if I can cut actually X i squared it is going to be 1, 1 1 squared is just 1 1 if ith trial is success, and 0 else. It is the same thing I mean X i squared the next day are exactly the same because X i just take 0 and 1. So, that is very useful. So, if you put that if you if you actually compute expected value of X i squared you will simply get p again the same calculation as before.

Now, what about X 1 X 2 or any other X i X j? So, if you want to find expected value of X i X j for i not equal to j i and j are not equal say for instance X 1 X 2 or X 1 X 3. So, in this case I know X 1 and X, X i and X j are independent ok. So, if they are independent. So, if i not equal to j, X i and X j are independent; Remember in the previous case here expected value of X i squared you will get, if i is equal to j, if i is equal to j X i and X a are not independent. You have to calculate the X i squared carefully, but if i is not equal to j expected value of xi and X j you know X i and X j are independent. So, when they are independent you can do the calculation very easily. So, this becomes expected value of X i times expected value of X j.

Now, what is expected value of X i? It is p times p that is just p squared ok. So, fully you see the calculation here. So, expected value of X i square is p expected value of X i X j is p squared for i is not equal to j ok. So, these n terms correspond to X i squared right these n terms correspond to X i squared these n into X n minus 1 terms are X i, X j for i not equal to j ok. So, now, you go in and use the same formula here. So, this becomes equal to n times p right expected value of X i squared is p we had that right there and then plus n times n minus 1 times expected value of X i X j which is p squared ok.

(Refer Slide Time: 12:03)

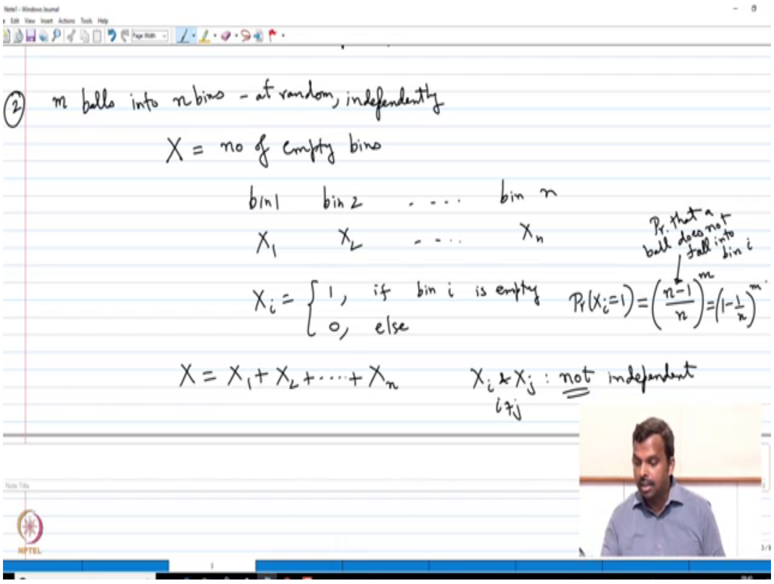


So, you get n p minus n p square plus n squared p squared ok. So, I have just expanded this written it in a different way. So, that is expected value of X, X squared ok. What is variance of X? Variance of X is the expected value of X squared minus expected value of X the whole square. So, now, expected value of X squared I calculated as this from about ok. And then what is expected value of X the whole squared n times p? The whole square ok. So, these two will cancel and you simply get variance to be n p times 1 minus p pull n p common out of these two things and you get n p into 1 minus p ok.

So, this is the first example I wanted to do and hopefully this illustrates for you the ways in which you can use indicator random variables in a simple way. In this particular case you also ended up having the independent trials. So, maybe some of it was obvious, but nevertheless I think if you directly calculate with the binomial distribution you need to do some summations and all that very carefully for deriving these formulae. So, getting to do it through the indicator random variable route is also very very important ok.

So, the next example we will see will be slightly more complicated than this one will once again use the indicator random variables, but we will not have independence ok.

(Refer Slide Time: 13:48)



So, that example is balls on bins our favorite example from the discrete world is resurrected again. So, let us say you throw m balls into n bins so once again at random independently ok.

So, every ball is thrown uniformly at random into any of these n bins and two different balls what happens is independent of each other ok. So, I am going to define the random variable X as the number of empty bins ok. So, you throw all these balls into these bins and finally, you count up how many bins are empty and that is your random variable X ok. So, it turns out this is not quite easy to study, this is not a random variable that is very easy to study. For instance you could I mean maybe some extreme cases you can study a bit more easily, but nevertheless is that slightly difficult random variable to get your hands around ok.

So, one of the things is when you have a random variable which may be the distribution is difficult to write down or difficult to get some intuition about. One of the things to do is to compute its expected value ok, you computed this expected value may be the expected value of X squared you compute for the variance of X. You compute if you can compute those things you get a rough idea of what the random variable is going to look like ok. Particularly if you can find the expected value in variance it turns out you can say a lot of things about the range at which the random variable will take values with high probability and all that ok.

So, all of that is useful in various applications and let us see how to do expected value for this ok. So, it turns out finding expected value of this is quite easy ok. So, what do you do you use indicator random variables ok. So, I have bin 1, bin 2 so on till bin n and I will have X 1, X 2 so on till X n be indicator random variables indicating whether or not the ith bin is empty. So, X i it is going to be one if bin is empty it is going to be 0 otherwise ok. So, X i is an indicator random variable it takes the value 1, if the ith bin is empty after you have thrown these m balls into n bins you go look at the first bin. If the first bin is empty you say that the X 1 random variable is 1, otherwise it is 0 ok. So, that is what happens in this experiment that is how this random variable xi is defined.

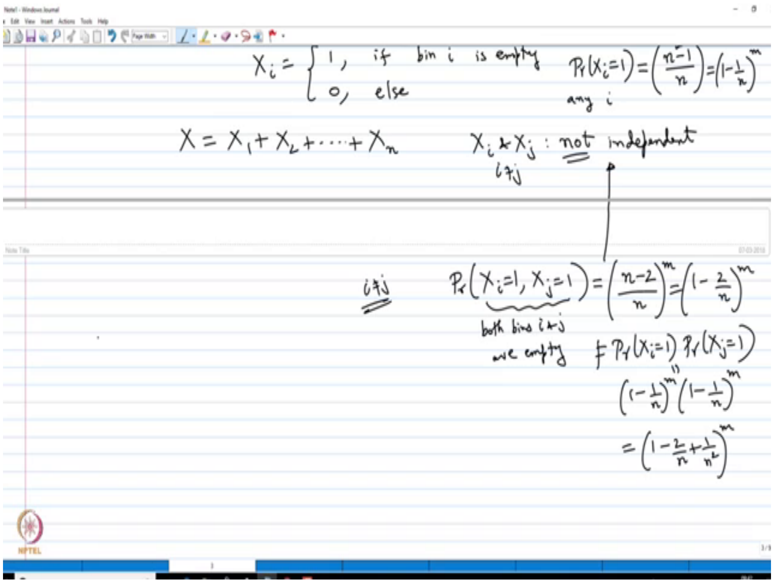
Now, in terms of these X i's these indicator random variables one can write down X quite easily. So, what is X X is X 1 plus X 2 plus so on till X n ok. So, I mean this situation at this point looks. So, similar to the binomial Bernoulli example we had previously, but the crucial differences these X is are not independent ok. So, remember that xi is X i and X j for i not equal to j are not independent.

So, we know this, we know this you can calculate this and convince yourself that this is true in so many different ways it is see, see you also intuitively you can see. So, supposing I tell you that bin 1 was empty the probability that bin 2 is empty is going to change, right. So, because you know when 1 is already empty there are m balls that are going to be thrown only into n minus 1 bins. So, there is going to be a higher probability that bin 2 I mean lower probability that bin 2 is going to be empty than otherwise ok. So, things are going to change you can you can also establish this very precisely you can calculate the probabilities etcetera etcetera ok.

So, they are not independent, but nevertheless one can find probability that X i is equal to 1 ok. So, what is the probability that the ith bin is empty? So, that is not too difficult to do. So, the ith; ith what is the probability that a particular bin the first bin let us say is empty is that every ball should land in any of the other n minus 1 bins. It should never land in the first bin and every ball is independent, so for every ball the probability that it lands in any of the other bins is n minus 1 by n this raised to the power m is your answer ok.

So, what is this? This is the probability that a ball does not fall in bin i right. So, that is the probability that a ball does not fall into bin i and I want every single ball of the m m balls to not fall in bin i. So, probability that X i equal to 1 the bin i is empty is n minus 1 by n raised to the power m you can also write it like this 1 minus 1 by n to the power m if you like ok. So, that is the answer ok. So, hopefully you understood how that works and so this is the same for any bin for any i this is true ok, any i this is true. So, probability that X is 1 is 1 minus 1 by n power m probability that X j is 1 is also 1 minus 1 by n power m ok.

(Refer Slide Time: 19:00)



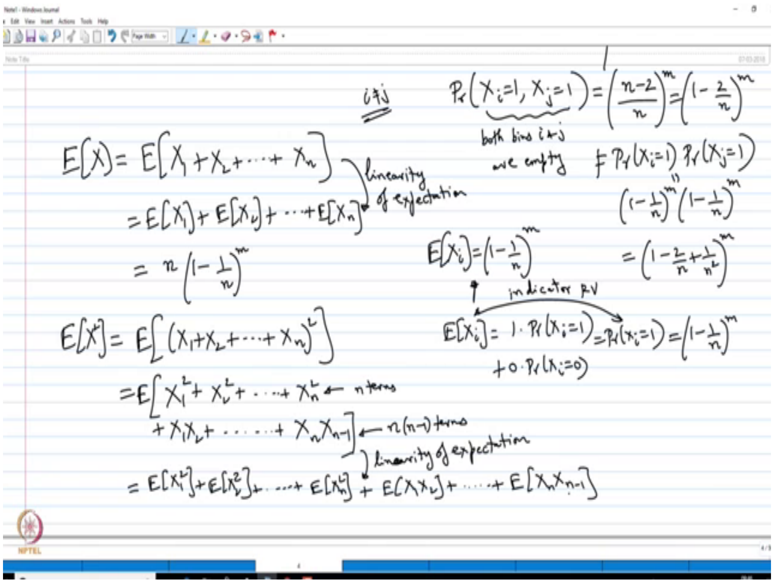
So, it turns out one can also find other probability. So, for instance for i not equal to j what is the probability? That X i is 1 and X j is 1 ok. So, what does this tell you both bins i and j are empty ok. So, turn this is also not too bad to calculate. What is the probability that both bin i and bin j are empty? Every ball ok, every ball that you throw should not fall into either bin i or bin j its allowed to fall only into the other n minus 2 bins.

So, using the same sort of logic you see this works out as n minus 2 by n raised to the power m. So, 1 minus 2 by n raised to the power ok. So, right off the bat you can see that this is not equal to probability of X i equals 1 times probability that X j is equal to 1 ok.

So, what is this guy? This guy is 1 minus 1 by n power m times 1 minus 1 by n power m if you want to write it as power m this will be equal to 1 minus 2 by n plus 1 by n squared power m which is clearly not equal to 1 minus 2 by n power m ok. So, it turns out that is one over counting here which gets rid of this problem. So, X i and X j are not independent. So, this is a proof for why X i X j are not independent. They were independent this would be true, but they are not independent, but even though they are not independent you can calculate just with X i and X j you can calculate their probabilities very very easily ok. So, this is something important to remember.

So, now let us come back to our original problem of number of empty bins ok. So, when you want to do number of empty bins like I said probability distribution of X is a little bit more complicated to write down. But expected value of X is not too bad ok.

(Refer Slide Time: 21:05)



So, expected value of X is expected value of X 1 plus X 2 plus so on till X n. Now we use linearity of expectation ok. So, a lot of people who see this for the first time always ask this question, but X 1 and X 2 and all that are not independent how can you use linearity of expectation. It is precisely be convinced it is true linearity of expectation is always true ok.

Linearity of expectation is true for dependent random variables also arbitrary dependence is possible as long as you have addition expectation will distribute over addition ok. So, it is a linear operator does the dependence or independence of the X is do not matter at all for the linearity of expectation to work remember that that is a very important thing to remember. So once you know this you are done ok. So, why is that? Because I know probability of X i is 1. So, what is expected value of X i? It is easy to calculate it is going to be 1 minus 1 by n power m ok. So, why is that true? Why is the expected value this?

So, what is expected value of X i? There is one times probability that X i is 1 plus 0 times probability that X is 0, which is nothing, but probability that X is 1. So, this is another random another fact which you should remember for indicator random variables for any indicator random variable its expectation simply equal to the probability that it is equal to 1 ok. So, it is very easy to see why that is true. It takes just two values 0 and 1 and probability that it takes the value 1 is simply the expectation ok.

So, this we saw already there is 1 minus 1 by n power m. So, that is true ok. So, what is each I the expected value is simply 1 minus 1 by n power i power m. So, here you have n such terms. So, it is n times 1 minus 1 by n power m ok. So, even though the actual distribution of X might be difficult to write down you can find its expectation ok. So, this is also an interesting fact to remember as you go forward studying more probability a random variable might be defined in a very complicated way quite often its expectation is easy to find ok. So, as long as it is composed of some smaller random variables which you can handle in some sense you can find this expectation maybe not very easily, but in quite often it ends up being an easy expression to write down ok.

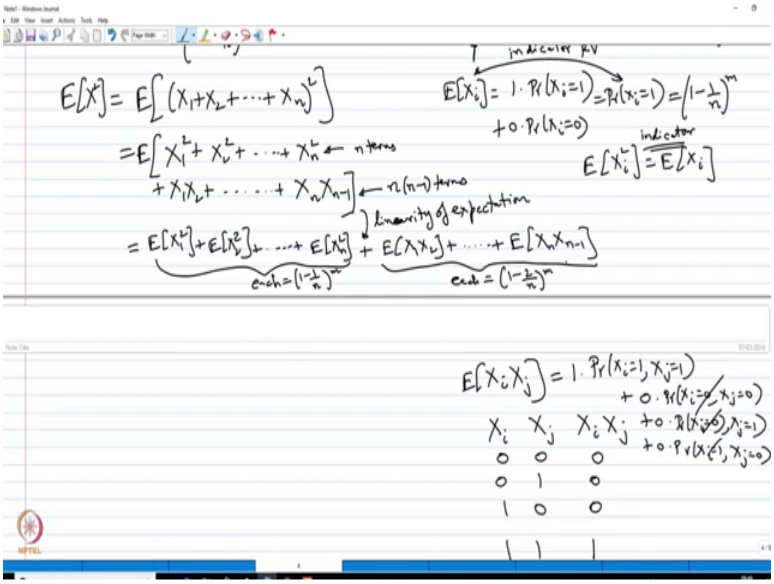
So, the next thing I am going to show you is what to do with say variance of X or the expected value of X square ok. So, expected value of X square and variance are I mean more or less equivalent in some sense I know expected value of X if I know expected value of X squared I also know its variance ok.

So, how do I find the expected value of X square? We can use the same trick once again ok. So, this is expected value of X 1 plus X 2 plus so on till X n squared and now you again expand this guy right I am going to use linearity of expectation even though there is also dependency I can use linearity of expectation. We have X 1 squared plus X 2 squared plus so on till X n squared plus you know X 1 X 2 plus so on till you solve for X n, X n minus 1 see these will be n terms here there will be n times n minus 1 terms here ok.

Hopefully, see that this is the same thing as what I did before exact same simplification. So, now, we use linearity of expectation. If we do that this will simply become expected value of X 1 squared plus the expected value of X 2 squared plus so on till expected value of X n squared and then you will have all these other terms expected value of X 1

X 2 plus dot dot dot dot dot expected value of X n X n minus 1 ok. Hopefully you can, hopefully this is easy ok.

(Refer Slide Time: 25:30)



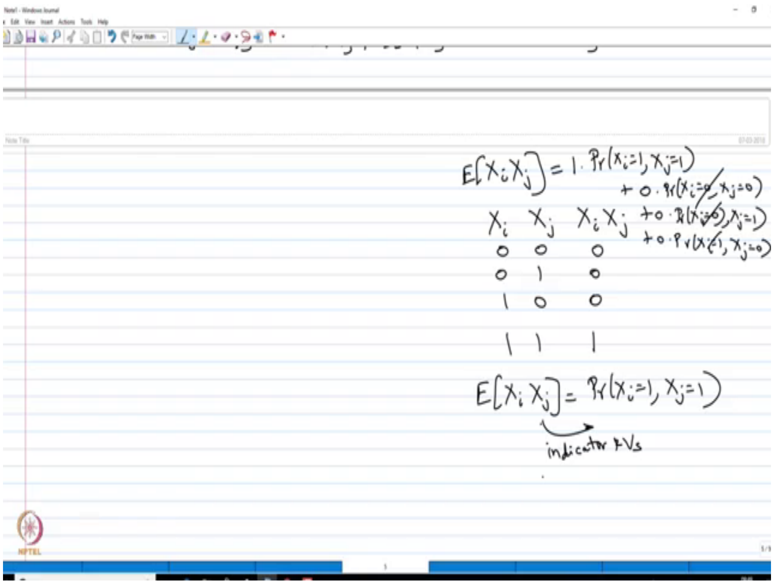
So, now what is expected value of X i squared? Remember expected value of X i and expected value of X i squared are the same right it is the same as expected value of X i. So, again this is true for any indicator random variable ok, because X i takes just values 0 and 1, if you square it or cube it or anything here nothing is going to change you will get the same probability. So, that is this is easy to handle. And what is expected value of X i times X j? Remember; so here what is the expected value of X i X j? So, I am multiplying xi and X j right.

So, remember X i and X j take two values 0 or 1 ok. So, if you write down X i X j X i X j they take 0 0 0 1 1 0 1 1. What will happen when you multiply X i and X j? All these guys will become 0 this will be the only thing that is left ok. So, if you do expected value of X i X j you will get one times probability that X i equals 1 comma X j equals 1 both xi and X j are 1 plus 0 times various other things ok, so probability of X i equal to 0 all the other things right.

So, X j equal to 0 plus 0 times probability of X i equal to 0 X j equal to 1, plus another 0 times probability of X i equal to 1, X j equal to 0 ok. So, all these terms will go away and we will be just left with probability of X equal to 1 X j equal to 1 ok.

So, here is another interesting thing about indicator random variables. The only thing we used here is indicator random variable I do not care about dependence or anything like that.

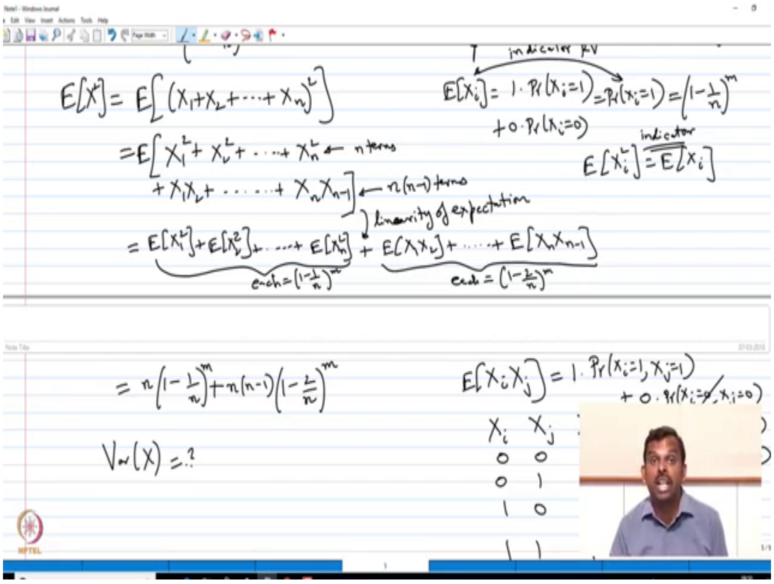
(Refer Slide Time: 27:10)



So, expected value of X i X j is equal to probability that X i equal to 1 comma X j equal to 1. So, this is again a fact which is true for any indicator random variable and you can also extend this X i, X j, X k etcetera keep on extending this simply the probability that all of them take together take a value 1. All of them individually I mean jointly take the value one ok. So, this we have already calculated right. So, we just went and did this calculation here X i X j equal to 1 and that is simply 1 minus 2 by n power m ok.

So, that is now easy to write down. So, here you see that there are n terms here and all of them are equal to each equals 1 minus 1 by n power m there are n into n minus 1 terms here and each equals 1 minus 2 by n power m ok.

(Refer Slide Time: 28:06)



So, you see I can write this as n times 1 minus 1 by n power m minus plus n into n minus 1 times 1 minus 2 by n raised to the power m ok.

So, that is the little trick involved in this in this calculation. Now, main trick is linearity of expectation and the fact that indicator random variables pretty much tell you the expectation is the same as joint probability calculation for both of them taking the value 1, ok. So, for this problem which was I mean like, once again I want to reaffirm that we never found the distribution of X right, for the number of empty bins finding the distribution of X is a little bit more painful and if you want to calculate the expected value of X square through that root find the distribution of X first and then use the formula for expectation it is not that easy, ok.

On the other hand using indicator random variables and using the linearity of expectation, and using properties of indicators one can very quickly get to the answer ok. So, I will leave the computation of variance as an exercise. It is its quite easy to do, I will leave it as an exercise to you to finish up this is the end of the lecture on expectations for calculation of expectations for discrete random variables, slightly more complicated cases involving indicator random variables.

Hopefully you like this lecture. And this will also be the last lecture as far as examples go in this course and this is the last week of the course, and if you have managed to come along this far in this course I am really happy that you managed to do so. Hope you enjoyed this course.

Thank you very much for being a part of it. Bye-Bye.