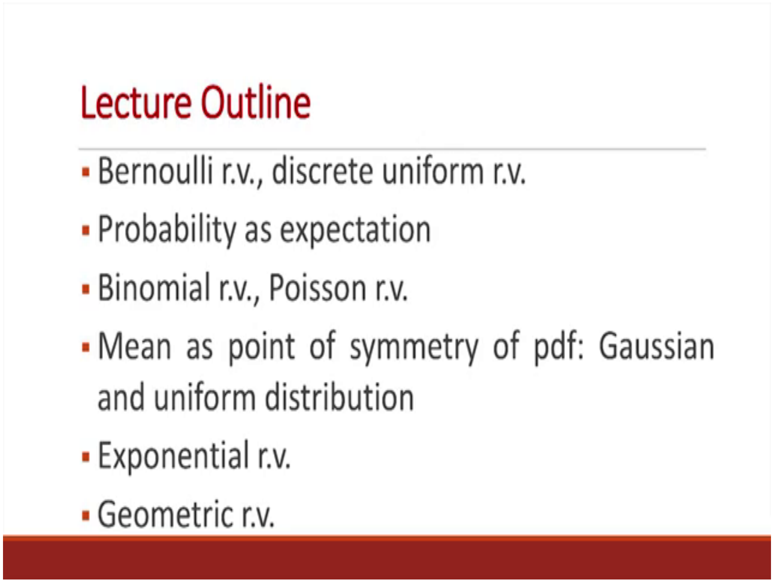
Probability Foundations for Electrical Engineers
Prof. Aravind R
Department of Electrical Engineering
Indian Institute of Technology, Madras

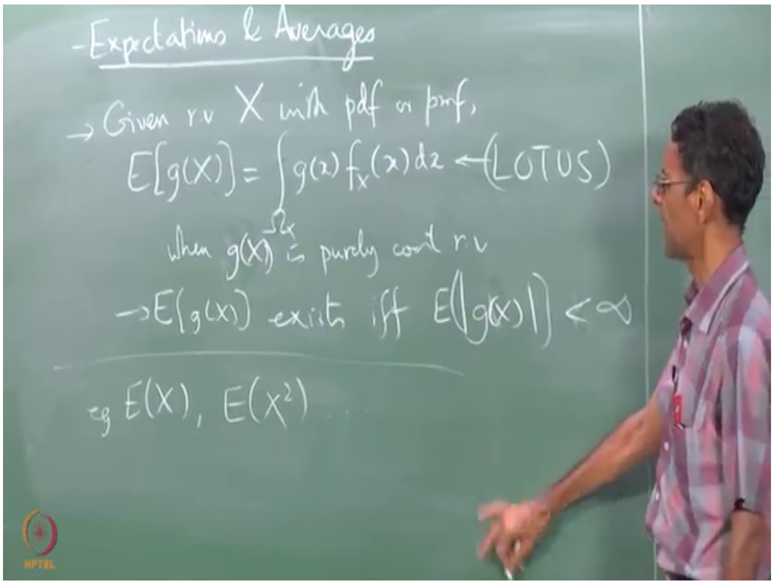
Lecture - 65
Expectation Computations for Important Distributions

(Refer Slide Time: 00:15)



Yesterday we started the this very very important topic on what you should view this you should view this is some yet another type of probability calculation, or calculation using a pdf and pmf is not it.

(Refer Slide Time: 00:40).



And basically so given some random variable with the some pdf or pmf does not matter, we are doing let me write down the most general thing, the E of g of X for it turns out right this is the kind of the more most general formula, we can write assuming for continuous g right.

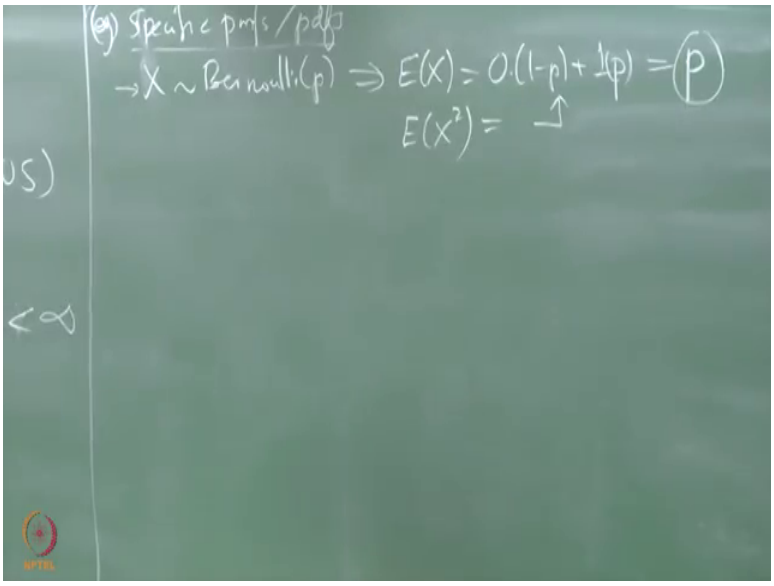
So, for g of X being a purely continuous random variable right, X is a special case of g of X is it clearly right, and this I am and a summation I do not want to repeat. So, this is basically referred to as a law of the unconscious statistician, I hope some of you went and Googled that phrase and hm.

Student: (Refer Time: 01:48). It has a Wikipedia page.

In a whole page on it so, we have been doing I have been doing it for thirty years, and I do not know I stumbled on this term only la in the last few years, before that I was not myself aware of it earlier we used to call it the fundamental theorem of expectations, when I this sounds a lot nicer. So, anyway, so this is so the condition of course, is that right for that is E of g of X exists the existence of this integral, if and only, if E of mod g of X is less than right exists or is less than infinity, and this is just to ensure that I said yesterday you do not write infinity minus infinity as some finite value ok. So, where do we go from here.

So, we so examples of course, E of X itself and then E of X squared and so on, these are the important individual applications of and we will do many more the dot dot really means there are more instances of e coming our way, but for now if we focus on E E of X andE of X squared for particular pdfs for particular distributions, we get I am now going to do a and do a big calculation on the board, but just 1 or 2 I will just do for illustration case illustration purposes. So, let me start it up here keep this blank right.

(Refer Slide Time: 03:26)



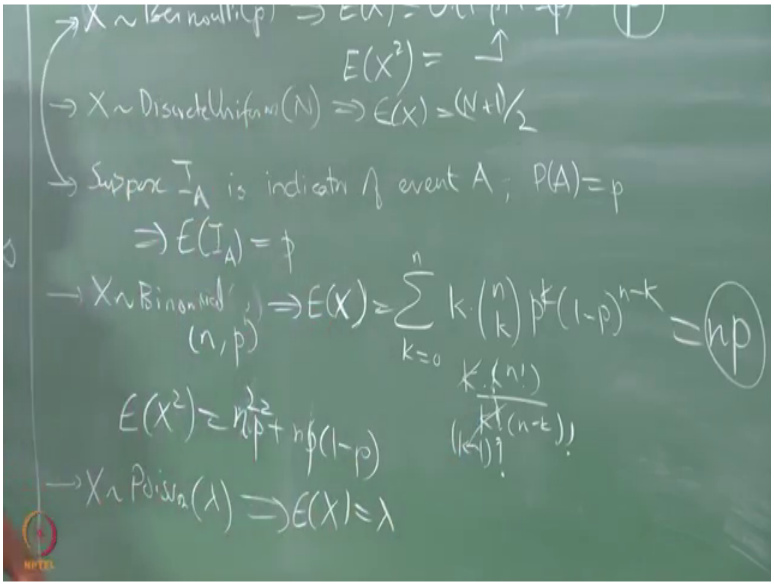
These are examples again of specific pmfs pdfs right, we start with a discreet well the simplest is Bernoulli.

So, it turns out for this very simple case which is what just 0 1, and probability of x being equal to 1 being p. So, what is the E of X it is very simple as just 0 into 1 minus p plus, 1 into p that is all just following the definition, it is basically p what about E of X squared, if you apply just simply this LOTUS turns out that X squared and also takes values only 0, and 0 and 1. So, it is very easy, so again exactly it is exactly the same thing 0 squared is 0, and 1 squared is 1.

So, you get once again you get back exactly the same number this is the very unique thing about this Bernoulli distribution that for which E of X equals E of X squared right ok. So, yesterday I also pointed out that E of X squared is always non negative right, there is no way that E of X squared can be negative, and that is a general theorem relating to the non the average of some non negative function not necessarily just X squared, if you try it X power four also you will get the same result.

So then the discrete uniform I just want to mention that although, I am not going to calculate E of X squared, because it is little more involved, but E of X can be straight T away written down.

(Refer Slide Time: 05:20)



So, what is a discrete uniform so, this is a new this new notation we are using. So, if I say discrete uniform, and it means you will it is like a die roll from 1 to 6 except that I am taking 1 to n instead of 1 to 6, what is going to be them E of X for this it is always n plus 1 by 2, which is how we got the 3.6 yesterday for the and for the case of n equal to 6, this just this one thing is worth remembering right, and not n plus half, but n plus 1 right n plus 1 by 2 it is again very easy to derive.

All you have to do is do that arithmetic you know some of a first n natural numbers is n into n plus 1 by 2, you know that if you do that n is immediate ok. Let us move on to the most somewhat more theoretically interesting things, what if I had an indicator random variable, right which is basically nothing but Bernoulli. Suppose I A is indicator of a of some event A with probability P A let us say is equal to P, this ties into the Bernoulli, what so you can say well definitely what is E of I A are going to be it just going to be equal to p itself.

But normally we do not think of when we you know we do not you know we somehow I mean I may have given the impression yesterday for example, that probabilities and expectations are I mean expectation is something built on top of a pdf, but you can think of probability itself as some special case of an expectation right. The idea of expectation actually it what does it do it enlarges our concepts of probability right, every probability can be viewed as an expectation that is what this is saying. So, there in other words the probability of an event is actually in itself an expected value.

So, that these two are very I mean maybe I should have put this ahead of this, but anyway it does not matter ok. So, so then let us go on to some other discrete binomial let us say, now for other pmfs it turns out that the actual calculation if you want to do is a little non trivial. So, basically it means that you have to do some of this kind k times n choose k p power k one minus p power n minus k, there are several binomial sorry it should be binomial let us say n small n comma p itself hm.

So, E of X equals this where did this k come from this is value taken by x ok. So, if you did not have this k you just be summing all the probability. So, the pmf and it will be equal to 1, with this k what happens what do you think the answer is it is n times small n times p, n p how do I get that there are several ways of getting it 1 is to as we will see we can explain the concept of or we can exploit the concept of the binomial being a sum of IID Bernoullis and get that n p, or for now since we are not that far into expectation.

Just apply the from first principles you apply the summation, or you have to do is cancel this you know write this what is a k into n choose k is what, k into n choose k is it is basically k times n factorial by k factorial to n minus k factorial. So, this k will this k will cancel this k minus 1 factorial, and give you k minus 1 factorial, but then the summation I have to go from 1 to n naught 0 to n, because you have right, you if you go from 0 to n you get 0 minus 1 factorial which is not defined here.

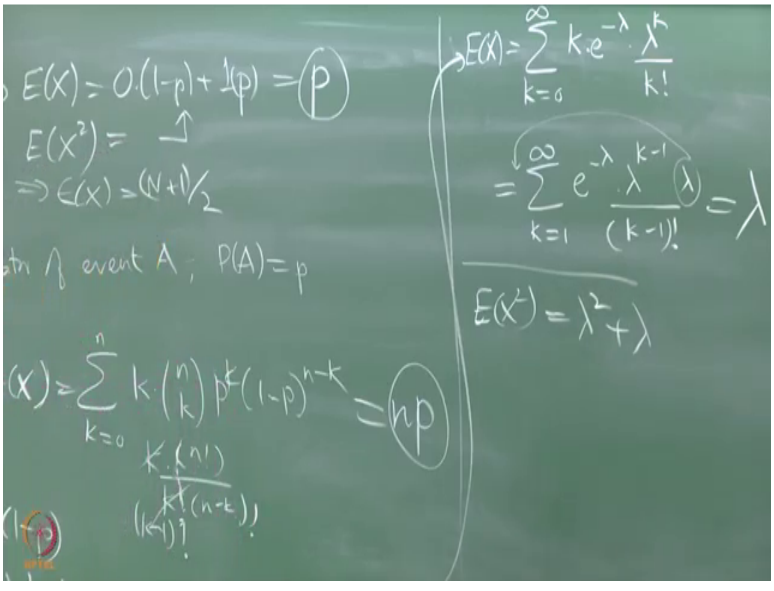
So, right so and you know that this sum anyway is 0 for k equal to 0 the contribution. So, you can manipulate it, and I because I have to go a little fast here I do not want to spend too much time on all these manipulations right, and this manipulation gives you the answer n P it is very easy to manipulate all you have to do is write this as p power k minus 1 into p and take that p out and you will get.

And this as n into n minus 1 factorial, and the n and the p will come out, and the rest of the sum will add up to 1. So, verbally I will dispose of it right E of X squared is more elaborate right it is not. So, simple but it is it is just a couple of more steps you put k squared here instead of k right, and you can still do the same type of simplification, and what you get is let me make sure I do not mess up here, it is it is a n it is a n squared sorry n squared p squared plus n into n p into 1 minus p this is what you should get. It is a if you just do it directly it is a bit of a mess it is not as simple, but I have worked it out here, but I you know if I keep writing all that down it becomes too much time is taken up.

So, you have an n squared p squared here let me write it very clearly it is n squared p square plus, and anyway all these results are available in lots of places. So, I do not think we should worry too much about exact formula right ok. So, then I just want to summarize them here because these are things that we will be drawing on a lot. So, please commit some of these 20 memory so, that when it is needed you can pull it out. What about Poisson.

If you have do poison lambda you can show that E of X is exactly lambda, again this is not too difficult as follows exactly the summation here, because k times there is a k factorial here, maybe I should maybe I will do it here, I do not want to.

(Refer Slide Time: 12:12)



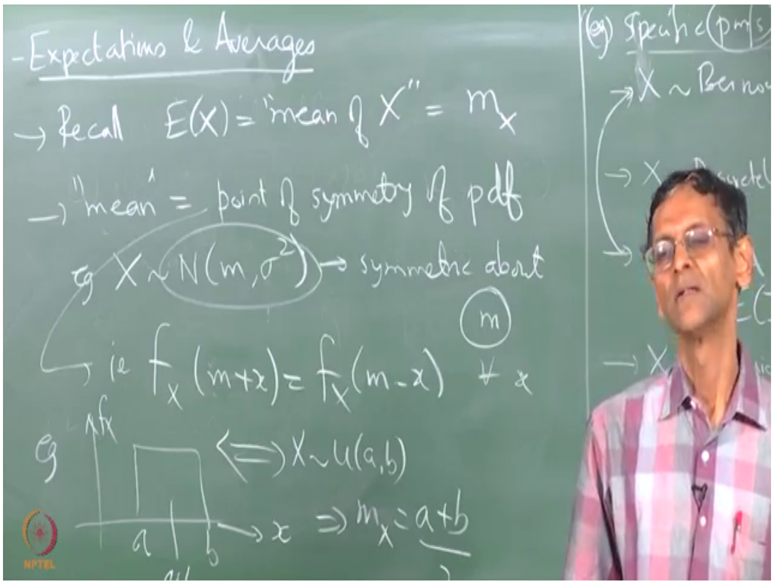
So, for the Poisson case I get E of X as sigma k equal to 0 to infinity right what are the product terms here, we will try to write it out fully. So, that we do not mistake anything or miss anything. So, k into e power minus lambda to the power k divided by k factorial, and all you have to do is cancel.

So, this becomes k equal to one to infinity e power minus lambda lambda power k minus 1 into lambda divided by k minus 1 factorial, I have to deliberately write it like that, and then I will put k k k prime equal to k minus one or something and you will find that this lambda will just peacefully come out here the rest of it will add to 1,exactly like the binomial.

Similarly you can show that E of X squared for the poison case is lambda squared plus lambda. So, E of X is lambda and E of X square is lambda squared plus lambda manipulations of this kind I am going to leave it to you, because it is a total waste of time, if I you know for me to do it on the board, I hope that you people have no difficulty in these types of manipulations right.

Because this is not the kind of thing that I mean we are to be spending time on here, if you have any difficulty please contact me offline I can guide you it is it is really not, all you have to do is put k square here for example, and then do the same kinds of manipulations on that I have done it here. In fact, it is not very complicated at all it comes in three or four steps ok. What about so the continuous case, for the continuous case and also for the discrete case we have very important result regarding them the E of X or the mean, remember what did I say yesterday E of X is also called the mean of X.

(Refer Slide Time: 14:30)



Also write it as m X, with all these subscripts or all with uppercase right, you are referring to the random variable, and so it has to be uppercase. The important what result regarding the mean is that especially for a continuous pdf, the continuous case is that if there is a point of symmetry in the pdf, and if the mean exists then that point of symmetry must be the mean.

So, mean equals point of symmetry actually it is true also on the pmf case, but pmfs such typically not very not particularly node to be symmetric, except the case of half half Bernoulli half half right, Bernoulli half half you can sort of says probability half is sort of symmetric about the value half in the middle 0, or 1 or but that is a very special case. So, we would not think of generally we do not use a pdfs pmfs is being symmetric, but pdfs definitely.

So, if the mean exists of course, right in the Cauchy case we mean does not exist. So, we even though there is a point of symmetry it is not regarded as a mean, but most importantly in the Gaussian case, if I say if I pick the arbitrary Gaussian pdf with mean para parameter m, now we can call it the mean.

So, this is this pdf symmetric about x right about m the value of m, what do I mean by symmetric is it clear about what I mean by symmetric means that if you take that is f of there is I e I can write mathematically what is the symmetry condition f X of X plus X plus sorry, m plus X will be equal to f X of m minus X, actually for all small X, that is the meaning of symmetry. Even if the pdf is limited it is not of infinite extent you still have 0 equal to 0. So, if I have for example, a uniform pdf from a to b what is the point of symmetry of the uniform pdf, what is the point of symmetry of the uniform pdf.

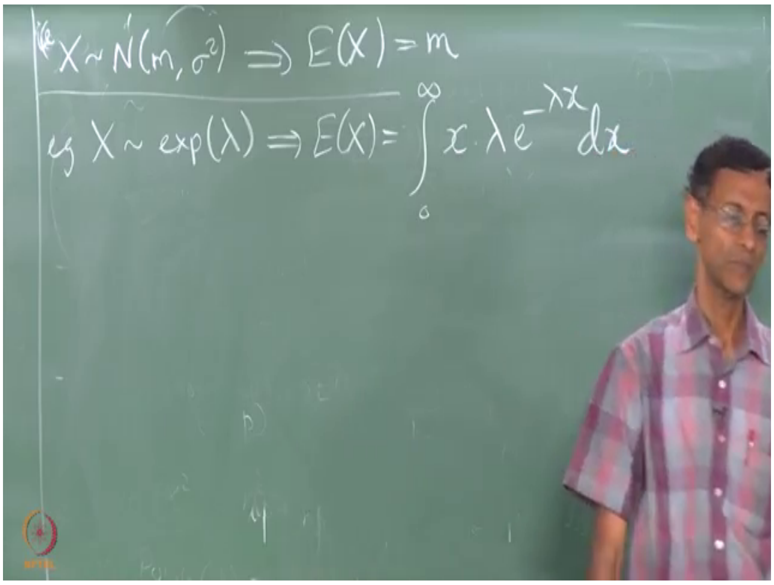
Student: (Refer Time: 17:17) point of (Refer Time: 17:19).

A plus b by 2 exactly why is there so much hesitation. So, implies that if X is u a b will mean m X is a plus b by 2, again to prove this what I meant I guess I am getting ahead of myself in some sense that, if this is true if there is a point of symmetry then I have written it here, the beam must be that point of point of symmetry.

So, for the Gaussian case there are two examples I have written one is a Gaussian the other is uniform. So, that actually right belongs to this. So, for the Gaussian pdf the mean is and how do I know the mean exists for the Gaussian pdf, it is it exists clearly because e power minus X square is such a strongly decaying function that it turns out you can do X power any integer, and the expected value of that quantity will still be finite, e power a E of X power k will exist for any k for the Gaussian case, unlike other pdfs which are dependent which shall fall off as some polynomial function like 1 by x squared.

For example, even mean does not exist, if you take 1 by X cubed, then mean will exist, but the square E of X squared will not exist, but e power minus X squared is strong enough that and e power minus X itself, even the exponential pdf itself, right is even though it does not have a point of symmetry it is still the decay is strong enough that every e power E of X power k will exist. So, it turns out that is a simple factorial function, you can we will but anyway at least or I should say the gamma function in general ok.

(Refer Slide Time: 19:30).



So, for the Gaussian case without doing any integration, I know that the answer has to be m. So, again there is a lack of time here. So, which means I have to I cannot spend more time on this, then I want to, but I think you people are mature enough now to go and think about it yourself and derive, and maybe even derive this result. All you have to do is you put you split that integral into 2 parts, and make this substitution in the 2 parts 2 parts am I talking about up to m and above m, and you make the substitution you can easily derive this mean has to be the point of symmetry.

Of course you are satisfy yourselves before doing any computation the mean actually exists which is most for most pdfs mean will exist right only the Cauchy is a very rare case right for example, the exponential itself. So, and I am among the discrete examples I have forgot to include the geometric. So, one thing again I take away from this whole course is that the exponential and geometric are very highly similar, right they are almost counterparts of each other right, you have the exponential in continuous space and geometric and discrete space, but otherwise a very lots of very similar properties especially the memory less property that we talked about right.

So, here what happens is you have E of X as integral 0 to infinity 1 example we will do just to show how this integral works out. So, I hope no right this right this is a does not need any further explanation, where is this X coming from it is because, I am calculating the E of X if I wanted E of X squared what would I put here.

Student: X square.

I would put X squared.

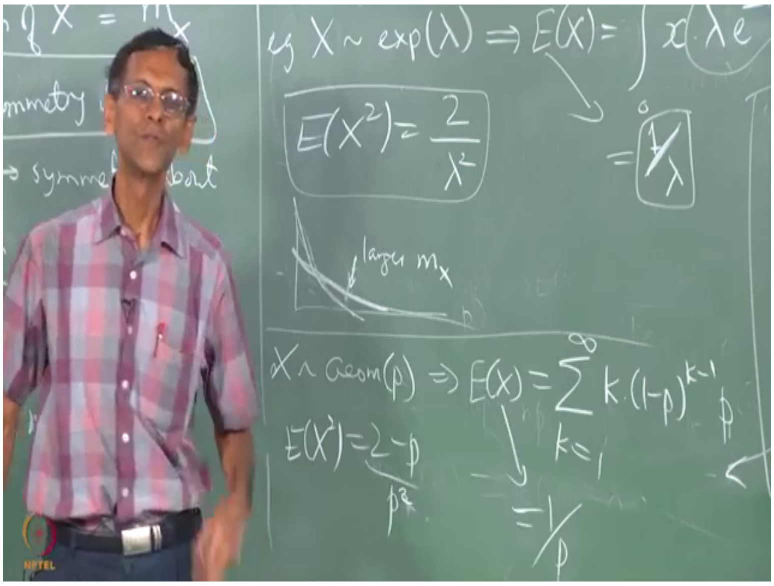
What is how do you go about if you know, if you are in twelfth standard and somebody gave this to you what would you do.

Student: Parts.

You would integrate by parts. So, the same thing you have to do here remember I said in the beginning of the course you cannot run away from integration, if you want to do this business. So, you have to brush up on integration by parts, I do not know if you have already seen it in this course, but we seeing it by the way we are seeing it here right. So, what happens to this I mean again I do not want to write out all the steps that is for the twelfth standard student to not to be done here. So, finally you get the answer what answer do you get you get.

Student: 1 by lambda.

(Refer Slide Time: 22:16)



You will get 1 by lambda 1 divided by lambda it should not be too much so, far too much of work for anybody right. So, why this reciprocal dependence on lambda, and similarly E of X squared let me see if I have written down correctly I do not want to right, I do not want to so, E of X squared is not is again two terms not one term, it is actually 1 by lambda whole squared no it is 2 by lambda square, again integrate by parts or use the gamma function definition right, what is the gamma function definition is something very similar to this right. You look up the gamma function.

So, this is basically if you take a k here, it will be k factorial divided by lambda power k or some such thing, but we are not going to go anywhere beyond too at definitely in this course right. So, these are the most important two things that we care about. So, why is this inverse relationship with lambda coming that is the intuition I want to bring out here, remember the spread of that pdf, what is how does a pdf behave for large lambda and for small lambda.

For large lambda it spreads out no sorry miss mistake for large lambda it is its tall and skinny, right probability is concentrated around 0, remember this is the pdf. So, lambda is inversely proportional to the spread of the pdf as far as the exponential is concerned unfortunately we write it as expo of only not as one by lambda right.

So, so la if you have to remember this for the exponential pdf lambda is inversely proportional to the spread, the larger the lambda the smaller the spread, and more close closer it is to 0 all the probabilities are sitting close to 0. So, the mean also will be smaller will be small if lambda is large, and similarly this E of X squared on the other hand is the smaller you make lambda the both the mean and this mean squared value will get pushed out.

So, you know. So, if you have if you have this versus this right. So this is so this will clearly have larger mean compared to this not just mean also larger mean squared value and relatedly the way and by extension the variance which we have not yet defined, but you are going to do so in a in a very short while is not it. So, this parameter lambda is very important to the expression benefit it controls everything about the exponential pdf, unlike the Gaussian which has 2 parameters this m and sigma square have two different roles to play right, m is only m only tells you how we know the location of that pdf along the X axis, sigma squared correspond right controls the spread.

But here this lambda does it everything ok. Now in I forgot to add in the discrete domain, we also have the most important geometry I mean I should say very important geometric pdf right, we just saw this counterpart of this. So, when I do so of so let us say I do X of

X is geometric P, what do you think the P are the mean value is here, how do I go about finding the mean value for this again I have to do a summation.

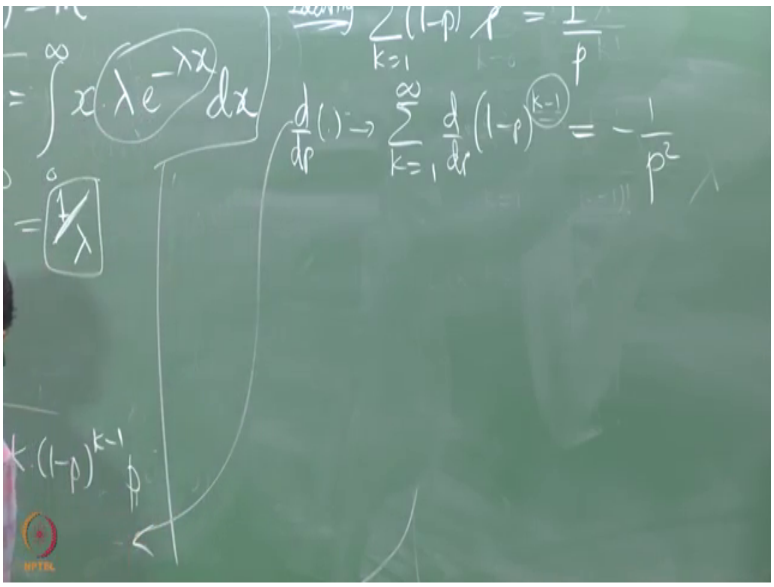
So, k times 1 minus p to the power of k minus 1 into p, what somewhat values of k k equal to 1 to infinity because I am assuming that I mean. This is my this is my definition of geometric pdf it goes from 1 to infinity and the pmf is 1 minus p whole power k minus 1 into p. So, this is the pmf and this value of X multiplied and sum how do I know it exists, because again it is an exponential fall off what kind of sum is this.

Student: AGP.

AGP and turns out that the way to sum a this it turn it turns out is to one simple a way to do it is to differentiate some identity, you do not have to remember the formula for this crazy AGP, because the AGP formula actually will work for will work for any finite sum also, but we are not doing a finite sum here we are doing infinite sum.

So, for the infinite sum you do not have to remember any formulas instead you can take the geometric identity, and differentiate with respect to p what do I mean by that, what is geometric identity here, there is some identity right. The sum identity is simply if I go and, if I look here.

(Refer Slide Time: 28:11)



Let us for the geometric I if I write The identity as I will write it as sigma I will write the full identity here, 1 minus k times p k equal to 1 to infinity is 1 is not it, this is the basic property of the geometric pmf itself. Now I will take this p and put divided by p here, just to right make it easier this is also correct.

Because p is nonzero I can divide by p. Now I want to differentiate both sides of this with respect to p, why I take the d by d p inside the summation, how am I justified in doing that I am justified in doing that because the sum is absolutely convergent that is this is actually some of positive qualities there is no plus minus terms here. So, I can take d by d p on both sides with p in and take and take the d by d p inside the summation, doing that gives me this.

What do I get by and that I may have to do some modification of this it is probably better to write this is a k prime equal to 0 to infinity and so on right so, that I do not like this k minus 1, but anyway this is minus 1 by p square. So, you can what do I get out of this manipulation, I get basically that k term coming here can you do it and see you should be able to get this the answer, what is the answer what is this what is this evaluate to hm.

Student: (Refer Time: 29:56).

It has to evaluate to one by p right, can you do complete the rest of the exercise just write this instead of one to infinity and think of it from here write rewrite it in a small minor change from 0 to infinity instead. So, which is actually just nothing, but 1 minus p power k 0 to infinity, it is still 1 by p is it not, and that is and then so, you do not have to you do not have to worry about this k minus 1 and all that right.

So, this basically equals this will equal from here this equals the important result 1 by p. So, you notice that this p and this lambda are sort of play the same roles the smaller the p the bigger the mean why is that the case, smaller the p means you have generally more number of trials to get the first success, which means the average number of successes is going to average number of trials till the first sections will increase.

So, the inverse relationship with p. Similarly this E of X squared it turns out this is E of X p of x square can also be done with k squared here, you can do one more differentiation of you write identity again differentiate once again with respect to p, and then you can get this answer, 2 minus p by p squared. So, like you have a lambda squared here, you have a p squared here. So, the smaller the p both the mean will grow as 1 by p, and the mean square grosses 1 by p squared ok. So, the exponential and geometric are highly right are you know very highly interlinked in more ways then it is obvious. So, but right but and they have a mathematically they are these kinds of things right are prove their similarity.