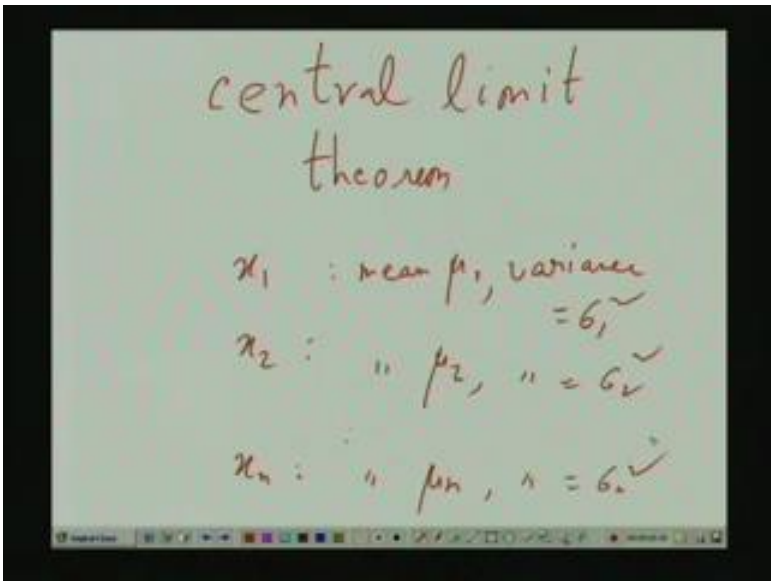
Probability and Random Variables
Prof. Dr. M. Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 28
Central Limit Theorem

So, today we discuss something which is a very important concept in the domain of probability theory. It is called central limit theorem.

(Refer Slide Time: 01:03)

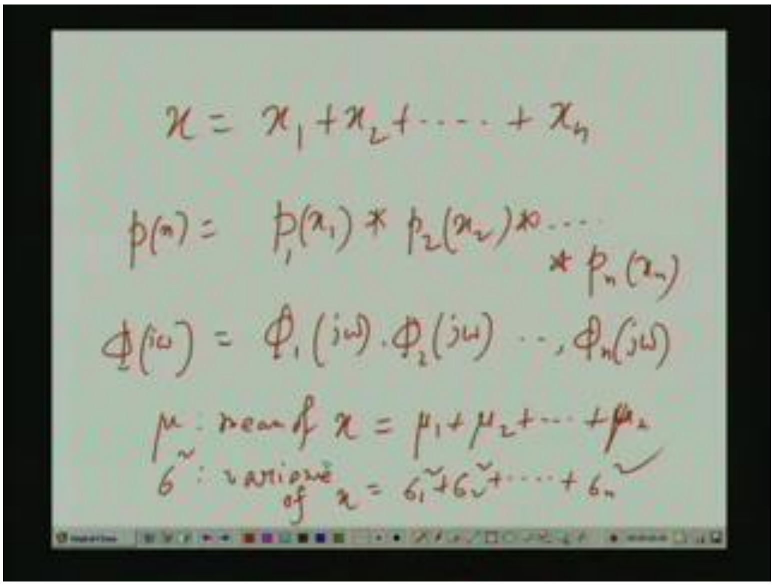


A rigorous mathematical treatment to this is fully beyond the scope of this course. So, we will not follow this, but we will take, first we will first state the fact, then explain it may be through examples and then try to construct kind of you know I mean that is not a proof, but some kind of logical basis for the central limit theorem. I repeat proper mathematically rigorous proof or treatment is beyond the scope of this course.

Now, this says that suppose we have got some random variables x 1. It has got some probability density and distribution function. It could be you know it could be binomial, it could be Poisson, it could be you know Gaussian. It has got a mean mu 1 variance sigma 1 square. Similarly, suppose we have got another x 2 mean mu 2 variance is sigma 2 square dot dot dot dot, we have up to xn, and I tell you x, they may not be same type of random variable. That is the probability density of x 1 could be different in nature from the probability density of x 2.

For instance, x 1 could be normal variable, x 2 could be uniformly distributed variable, x n could be say Poisson distributed variable likewise. So, there density and distribution functions can be widely different. We are taking n such random variables and this has got mean mu n and variance equal to say sigma n square. What is given is that this x 1 x 2 up to xn, they are statistically independent. X each random variables say xi has a mean mu y variance sigma s square.

(Refer Slide Time: 03:09)



Then, if you form a sum like this, we know that probability density of x is actually speaking is the convolution. The probability density function of x is the convolution, maybe I call it p 1. P 1 stands for the probability density of the random variable x 1. So, p 1 x 1 convolved with p 2 x 2. P 2 x 2 is the probability density function of the random variable x 2. So, p 1 x 1 convolved with star stands for convolution with p 2 x 2 star dot dot dot star pn xn. This we know all the three of characteristic functions or in terms of characteristic function phi omega is nothing but product phi 1. Phi 1 is the characteristic function of px.

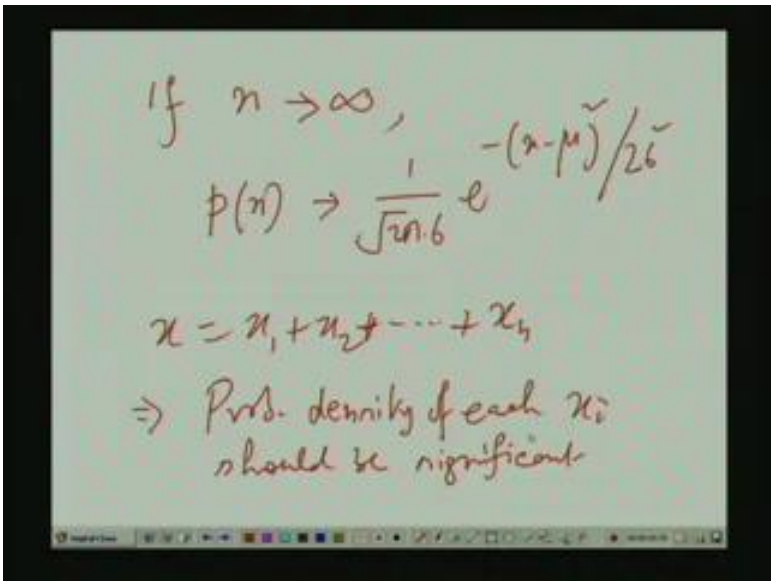
I do not know whether you wrote phi j omega or phi omega, maybe phi j omega. I am not sure. In any case, it is a function of omega only. It is not a function of j. We just write j omega because you know i mean j and omega come together. This is the product of phi 1 j omega dot dot dot phi n j omega. This is always true, but what central limit theorem says is really very interesting and it is several steps ahead of this. It says that suppose we take many such random variables, that is n that is the total number of random variables is quite large. Theoretically n approaching infinity practically n is say very large.

In that case, the summation x will become closer will tend to be very close to a Gaussian distribution. I mean the probability density of x will tend to be very close to Gaussian density function. Of course, the mean of that will be the mean of this summation and variance will be the summation. I mean if x has been mu and variance sigma square, then it will be the probability density of x will be a close approximation of Gaussian density function of mean sigma, mean mu and variance sigma square irrespective of whether x 1 is Gaussian or uniform x 2 is or Poisson or something else.

Similarly, x 2 could be I mean it does not matter whether x 2 is Gaussian or of some other. I know it has some other probability density function so on and so forth. If we add up many such random variables, then the resulting probability density function tends to be close approximation of Gaussian function Gaussian distribution. Of course, there is a condition I am coming to that, but before that let me just state this mathematically. Firstly, mu that is mean of x, obviously is equal to mean of x 1 plus mean of x 2 and then dot dot dot and then mean of xn. So, this is nothing but mu 1 plus mu 2 plus dot dot dot mu n.

So, if you know the mu, the mean of each of the random variable, we know what is the mean of x. There that is irrespective of the individual density and distribution functions. Mu is always the sum of the individual mu’s. Similarly, we have seen earlier that since x 1, x 2 up to xn, they are statistically independent and variance sigma square, that is variance of x is nothing but sigma 1 square plus sigma 2 square plus dot dot plus sigma n square. This is the summation of the variances. So, we formed a random variable x by summing n number of such random variables and I will take n to a very large theoretical tending to infinity, practically very large. In any case, the mean of the resulting sum is always sum of the individual means and variance of the resulting sum, that is x is nothing but sum of the individual variances. That is always true. So, mu and sigma square gives you the mean and variance respectively of x. That you know. Then, the central limit theorem states as we already discussed I am just putting in mathematically.

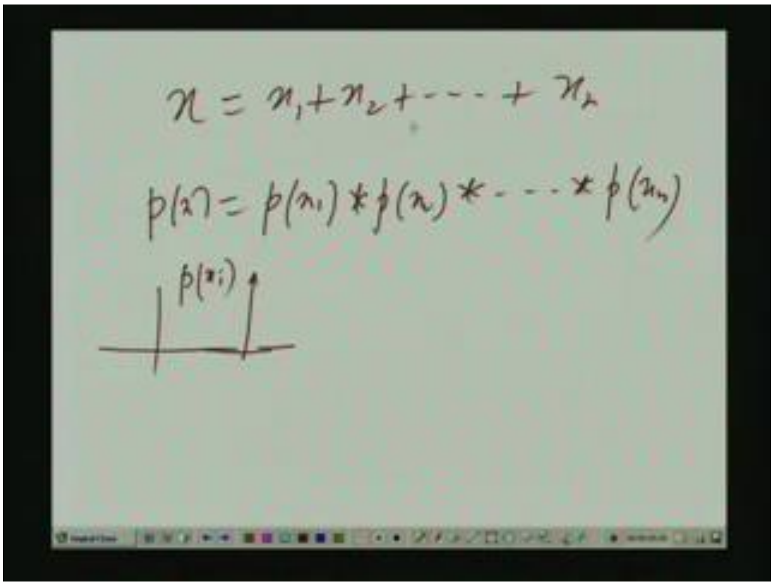
(Refer Slide Time: 07:46)



It states that p of x if n tends to infinity, that is n is very large. Px will tend to be a Gaussian curve, a Gaussian density of mean mu variance sigma square. Of course, this is true if the probability density of each random variable is insignificant. I mean everybody has a significant presence. It is not that you know I mean if you take n to be very large, it is not that just a finite subset of those random variables x 1 to xn, maybe m of them only dominates in terms of that density. The other n minus m, there are probability densities are you know their functions are very narrow and localized. So, they hardly dominate.

In that case, it will not be what I mean is that suppose you take as I said, we have x equal to x 1 plus x 2 plus dot dot dot plus xn. It means that what I am trying to say is this. The probability density of each xi should be significant by this. This is a just pure English statement. By this, what I am trying to mean is that the probability distribution should be well you know spread out function. You know it will not be very narrow and localized function and all that. That is what the axis is. If you take the x axis, it is not what the major part of axis. It has no presence and therefore, dominated by another density functions. If that be the case, then only px will tend to this.

(Refer Slide Time: 10:02)



What I mean is this. As I said x equal to x 1 plus x 2 plus dot dot dot xn which means px. As we know px 1 convolved with px 2 convolved with dot dot dot convolved with pxn. If now it so happens that out of these n random variables, several many of them may be r. Number of them are not significant in the sense that they are localized. They are like impulses like you know may be pxi is an impulse. That is interfered integrated. You will get one, but is localized here. Such functions I will call to be insignificant because over the entire x axis, they are hardly present anywhere. They are dominated by other probability density functions which are not of this type. This is only present at a particular point, maybe a b whatever.

So, this kind of functions should not be present. I mean we should not have random variables which have such very localized and narrow probability density functions which we also can view as one. So, they are weak in the sense that over the entire axis x axis in most of the places, they are dominated by other functions. This is not allowed. What happens if I have got an impulse? Obviously, if you have in the convolution if you just have an impulse, you know I mean you will get back, I mean it makes no difference. Any function convolved with an impulse gives you back the same function. So, that impulse does not make any difference in the output.

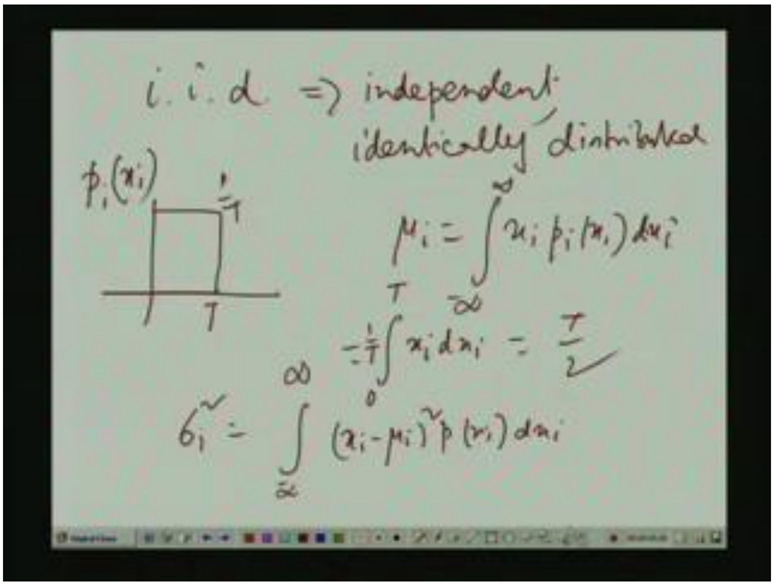
Any function convolved with by an impulse function does not produce any change. It gives you back the same original function. That means if out of this n probability density functions, suppose r of them are impulsive. That means I can just take them out because they do not make any difference. I just can take them out. They do not make any difference which means only n minus r number of probability density functions are significant, but n minus r may not be very high. In that case, I mean my statement that I take a very large number of such random variables. It is defeated because you verify, take n to be very large. Some of them which have impulsive probability density function; they have to be taken out. When you find out the resulting probability density functions, those impulsive probability densities do not make any difference.

So, it is as if only n minus r random variables are present and n minus r may not be high. So, in that case the central limit theorem may not work because for this the total number of such random variables which have significant probability density functions should be very high theoretically tending to infinity. So, that means I repeat central limit theorem will work if total number of random variables n tends to infinity or in practice, it is very large. Further each probability, each of them has got a significant probability density that is over the entire x axis. They are present in a major portion. They are dominated by other probability density functions in most of the part of the x axis. If that be the case, then the central limit theorem exits.

Now, a qualitative description, actually there is rigorous mathematical theory for it. That we will skip, but here at least through this convolution expression, we made our point clear that if some of the random variables has just impulsive probability density functions, then they do not make any contribution to p of x, whether you have the random variables or not. The probability density of x does not change because it is given by the convolution of the rest of the functions. So, such insignificant probability densities do not create any difference. So, I should not have such random variables which have such impulsive or very narrow or localized kind of probability density functions which we can say equivalently that probability density functions which are dominated in this x axis in most of the part, such random variables are not be included here. In that case, if n is large, then probability density of x p of x approaches as Gaussian function.

Now, this is a very useful theorem. This is used I mean this is used in many contexts. In fact, often it helps us in making assumptions like this that given a random variable which may be coming as a super position of various random observations, you know we can treat that random variable to be Gaussian without much problem by virtue of the central limit theorem, and once more of course x 1 x 2 xn, they have to be statistically independent. This whole convolution relationship is based on that. Let us take an example.

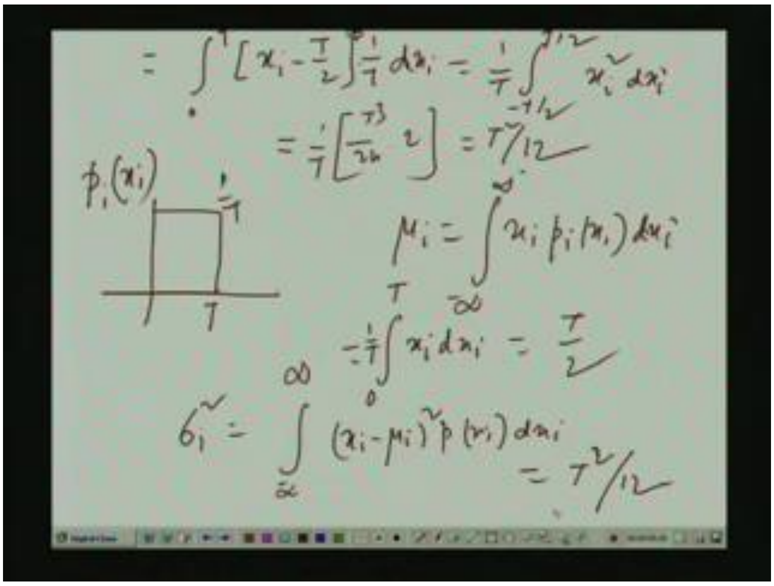
(Refer Slide Time: 15:12)



Suppose you have got random variables x 1, x 1 dot dot dot xn. Each of them, actually they are i i d, we say i i d. It means independent. So, they are mutually independent, statistically independent and identically distributed. So, they have the same density function, same probability distribution function because i i d. Suppose in this example x 1 up to x n, they are i i d random variables and their density is uniform, that is given that say pi xi is something like this. This is T, this is 1 by T. So, area is 1. So, first what is mean? What is mu i? Of course, mu i will not depend on i because this i mean is true for each i, all these i i d. So, each xi has a same distribution, identical distribution. Anyway still I write mu i.

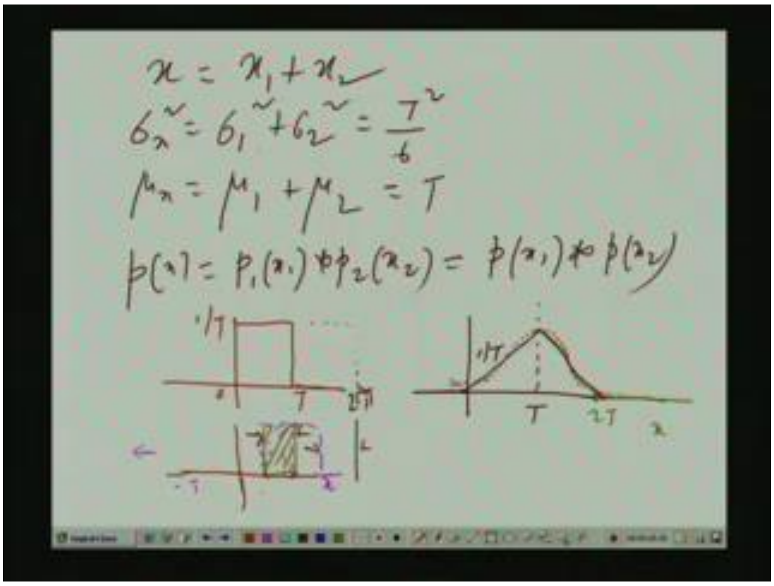
So, what is mu i? Mu i is of course as you know x into px, that is xi pi xi dxi. Xi should be from minus infinity to infinity and this is very simple. We should integrate only from 0 to T, and probability density here is 1 by T. 1 by T can go out and xi dxi. So, xi square by 2, this square by 2, then T, so basically T by 2. Obviously you know from just I mean from this figure itself you can see that since the density is constant from 0 to T, average will be T by 2. So, you will get T by 2. What is sigma i square? Sigma i square is nothing but xi minus mu i whole square pxi dxi from minus infinity to infinity and integral will be from 0 to T. Xi minus mu i mu i is T by 2, then you can say xi minus T by 2, you can call it xi prime.

(Refer Slide Time: 18:02)



That is a simple exercise, but in fact I can leave for to you to work out, but nevertheless we will do it first here. Xi minus T by 2 into 1 by T by t whole square into 1 by T whole square into 1 by T dxi, you call xi minus T by 2 as xi prime, it is just a shift of origin. So, its integral becomes this 1 by T comes out integral is from minus T by 2 to T by 2 and sorry. You call it xi prime and usually dxi and dxi prime are same. So, I again call it back as xi. So, xi square or you can it xi prime also, but no problem you can as well again write the same symbol xi. We also defined xi now dxi. So, if you integrate xi cube by 3, so T cube by 8. So, what happens is xi cube by 3. So, it is 1 by T. So, xi cube means if you put the limit T cube by 8 and 3 is 24. So, twice that T cube by 24 into 2. So, it is essentially T square by 12. So, sigma square is T square by 12. I would write this in another page.

(Refer Slide Time: 20:12)



Now, suppose I found first x is equal to just x 1 plus x 2. Obviously, sigma x square is nothing but sigma x 1 square plus sigma x 2 square, but each of them has got T square by 12. So, sigma x square is T square by 6 and mu x is nothing but mu x 1, mu x 2. Earlier I call it mu y. So, maybe I stick to the same notation and T by 2 that is T. So, x 1 x 2, each of them is uniformly distributed and they are added, they are statistically independent i i d. Therefore, it has variance T square by 6 and mean T.

What is the density of x? It is not uniform of course. So, what is probability density of x? That will be nothing but that is px is nothing but p 1 x 1 convolved with p 2 x 2. So, it happens that since these are i i d, p 1 and p 2 are the same function. They are uniform density function. So, I can as well say p x 1 convolved with p x 2. How does the convolution look like? I know px is such that the mean that we get out of it with is T and the variance that we get out is T square by 6, but what is the density function you have to convolve. So, how do we convolve? One function is like this 0 to T and then 1 by T and other function if you just reverse, you have to reverse it. So, it is like this. First you reverse it like this 0 to minus T. So, now, there is no overlap. If you now shift it to the left shift like this, the overlap between this and this is 0 and that is why any of the overlapping regions is 0.

So, you will get zero output. That means, to the left of this you will have 0 and now if I shift it to the right that is suppose this is now shifted to the right like this. So, then this is the overlapping area and actually convolution of two rectangle is triangle. We all know. So, what is the area of this? Area is if that advance by say x, if you have advances by this much x, then how much is the area x into 1 by T. So, x in to 1 by T of this as x increases and goes up to this capital T, this point area increases. So, it is a linear function with slope 1 by T. So, till x takes the value T, this goes up like this which is slope 1 by T slope, 1 by T because area is x into 1 by T. Then, suppose I am here x has come up to this place. This is x.

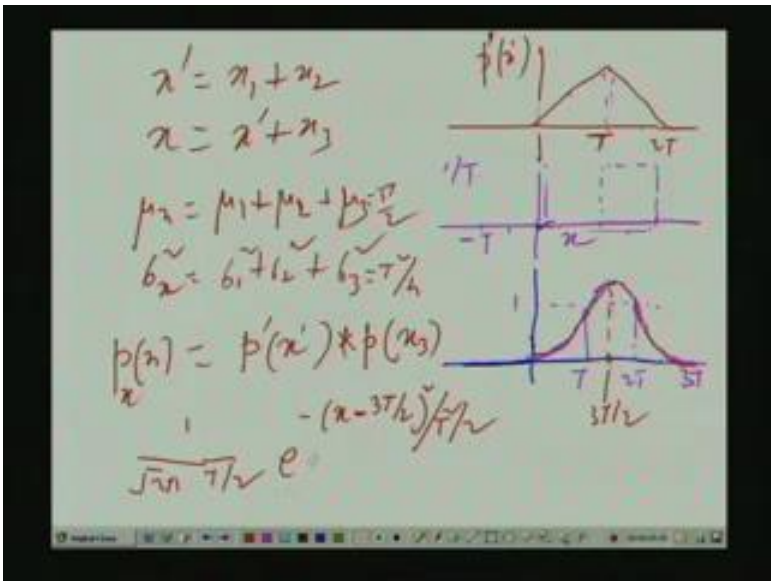
So, now, as it goes further and further to the right, the overlapping area that is this much decreases. So, finally a time will come when this rectangle will come out, that is when x is to the right of twice T, this rectangle will be going out. It is so going out to the right, so that there is no overlap between the two rectangles and area is again 0. So, that happens at x equal to 2 T and obviously, you can see that at 2 T, from 2 T onwards output will be 0 because when x is at 2 T, this rectangle is just somewhere here and the overlap is 0. As it moves further to the right, there is x. As x increases further and further, overlap still continues to remain 0. So, to the right of x equal to 2 T, i will always have the value 0 and to the left of 2 T between T to 2 T, I see that the area like here, this area actually this part will be decreasing as x becomes higher and higher because rectangle will be sliding further to the right. So, overlapping area will be less.

Obviously, you can see that area, this is a linear function actually like this. This is a linear function. It is just now very difficult to see. If this is x and total length is T, you know how much this side is. You know this is 1 by T. This is x, total length is T and this point is T. So, this is you know this is x minus T. So, how much is this? It is 2 T minus x. This length is twice T minus x and therefore, we fall down you can easily see. So, it is a triangle function.

How do you do this? 2 T minus x is very simple. Suppose I had repeated this. Suppose hypothetically I get up to 2 T and this is x. So, this gap is 2 T minus x, right and just from symmetry you see this is repetition. This portion and this portion, they are same. So, it is purely repetition. So, 2 T minus x times 1 by T. 1 by T is the height. So, there is a negative slope of minus 1 by T. So, it falls down now. So, this is the actual probability density, but if I plot now a Gaussian probability density with mean T and variance T square by sigma, how does it look like. If I plot a Gaussian distribution with mean T, then it will be like this, something like this. Mean is T and it will touch the top because Gaussian at mean is, it will be somewhere here. So, you see it is following the Gaussian curve is now some following somewhat this triangle.

Now, if instead of having just two random variables x 1 and x 2, suppose I have got another x 1, x 2 and x 3. We again do the same exercise and find out what is the resulting probability density and we will be doing it now. We will find that the resulting density is a further closer approximation or it becomes closer further to the Gaussian curve so on and so forth. How do you do that?

(Refer Slide Time: 29:29)



We have seen already that if you now call it x prime as x 1 plus x 2 and x equal to x prime plus x 3 which is nothing but x 1 plus x 2 plus x 3. Now, you have already seen px prime. If you say it is a triangle function, mean is T, centre is T and goes up to 2T. We also know what the mean of x prime is. That is mu 1 plus mu 2. What is the variance of x prime? It is sigma 1 square plus sigma 2 square. This we know. Now, obviously mu x will be mu 1, that is mu 1 plus mu 2 from x prime and mu 3 and sigma x square will be from x prime which is sigma 1 square plus sigma 2 square and then from x 3 sigma square. They are all statistically independent i i d.

So, then the summation of mean and summation of the random variances, what is the actually density? Actually density will be a convolution between px prime. May be I can call it pu p prime. Now because it is no longer uniform in appearance, so it is a different function all together. So, what is px? Px is nothing but p prime x prime. I will put that px p prime x prime, where p prime is of this form convolved with px 3, where px 3 is just uniform which we have seen earlier how does it look like. So, you have to convolve here. Px 3 is uniform. So, to convolve we have to first reverse it as before. At this moment, there is no overlap.

So, this point is x as x becomes, x goes to the left of origin, this entire rectangle moves further to the left and there is no overlap between the two figures. So, you continue to get zero value. So, as before to the left, you get 0. Actually as we know in signal processing, that convolution of any two causal functions, you see a causal function the same thing is happening here. The probability density function whether uniform or triangular, whatever you know they are all causal functions. So, naturally convolution also will give rise to causal function. So, to the left, it is 0. Now, to the right what is happening? It is bit interesting that suppose it is somewhere here now. So, this part is the overlap and you find out the area of this multiply by 1 T, I get some value, but that area if this is x, height will be 1 by T times x 1 by T is the slope.

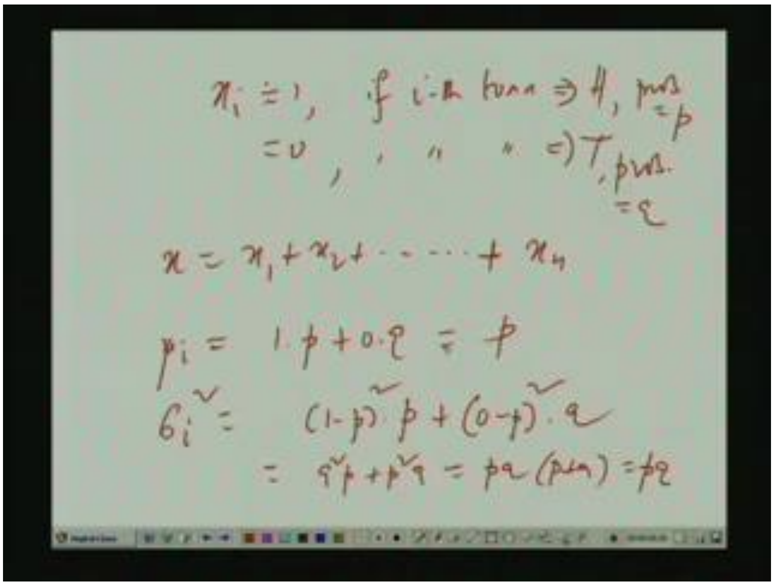
So, x into 1 by T x, that will be another square quantitative function x square by T. So, you will get a function like this, something like this. Then, as this rectangle goes further and further into it, the area increases finally when it is exactly here, that is x is T. Then, I have got the maximum overlap. I have got T into 1 by T that is 1 it goes up to 1 after this. After this what is happening? If this goes further to the right, suppose I am somewhere here. What is happening from this side? From this side I am cut off here only, but here I have been something like this. So, from this side a smaller triangle is going away, but from the right hand side, the bigger trapezium is coming in. So, area still increases. So, it will still increase.

When will it stop? When will it reach maximum? It will reach maximum somewhere here when I guess x has gone somewhere here. I mean let me draw when that rectangle has come here and that this point T is in the middle at this point, it reaches maximum. As this rectangle tries to move further to the right, then what happens is you get a trapezium from right hand side of smaller area and a bigger triangle goes out. So, the area starts falling, that is if I now goes further like this, then from this triangle a bigger trunk chunk will go away. Earlier here this much, now smaller portion will be coming. So, naturally the area will fall down and this will continue. Then, what happens is when x is here, then there is some overlap. Same thing you will get value 1.

Now, as this starts going up to the right area will fall down because you will only be losing. No new area will be added, but area will be some portion of the triangle will always goes out. It will start falling and finally, when x has gone to thrice T, that is at 3T, it will appear 0 because that time it will be just no overlap between the triangle and this T. So, this is like this, so T 2 T. Now, here mu 1 plus mu 2 plus mu 3, how much is that? Each of them has mean T by 2. So, this is 3 T by 2 and 3 T by 2 is somewhere here. This is 3 T by 2, right. This is the mean. What is the variance? Variance is T square by 12. So, how much is that? T square by 4. So, this is equal to T square by 4, this is equal to 3 T by 2. So, T square by 4.

So, if I now plot a Gaussian density function e to the power minus x minus 3 T by 2 whole square divided by twice sigma square a, sigma square is T square by 4. So, twice into T square by 4 makes it T square by 2 and 1 by root 2 pi sigma and sigma is nothing but T by 2. How will it look like? That Gaussian density will be even closer to this. It will be almost on this. Since, e as we added one extra variable x 3, we get even better even further closer approximation to a Gaussian curve. So, this is the meaning of the central limit theorem that if x 1 x 2 x 3, they have densities which are you know not dominating, one is not dominating the other, in that case if you sum them and if you sum many of them and if they are statistically independent, then the sum tends to become a Gaussian random variable. One result follows from this.

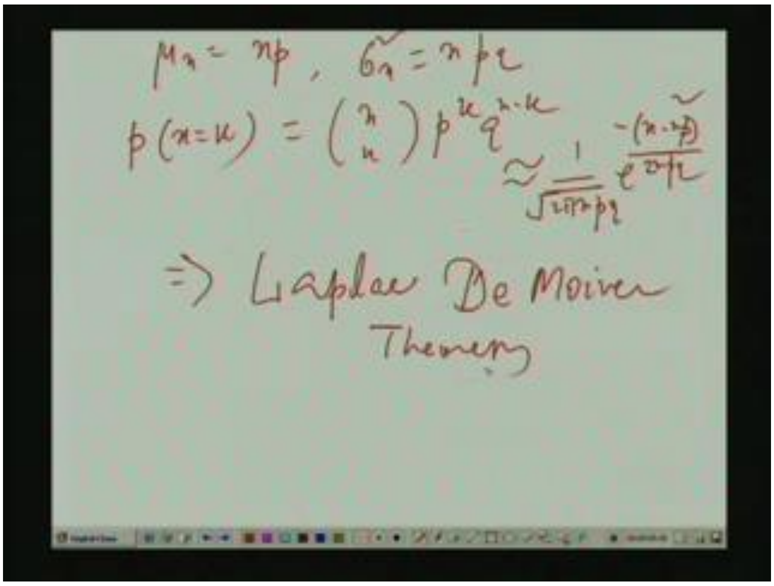
(Refer Slide Time: 39:54)



Suppose we have got this Bernoulli trial, that is we are just tossing coin n times. Xi is a random variable and it is related to the ith toss. If xi is ith toss gives rise to head, xi takes 1. If ith toss results in head equal to 0, if ith toss result is tail this has probability p, this has probability q and p plus q equal to 1. These tosses, they are mutually and statistically independent as you can see and you just form a random variable by summing them and summing many of them. Now, what is mu? It is i here. What is mu i? Mu i means xi can take value 1 with probability p. So, 1 into p and value zero with probability q. So, it is 1 to p plus 0 in to q which is nothing but p.

What is sigma i square? Sigma i square when it takes the value 1. So, its deviation from mean is 1 minus p. So, square of that times the probability of 1 is p. When it takes the value zero, then deviation from mean is 0 minus p square into probability q and 1 minus p is nothing but q. So, this is q square p plus p square q, you take pq common p plus q and p plus q is 1 which is nothing but pq.

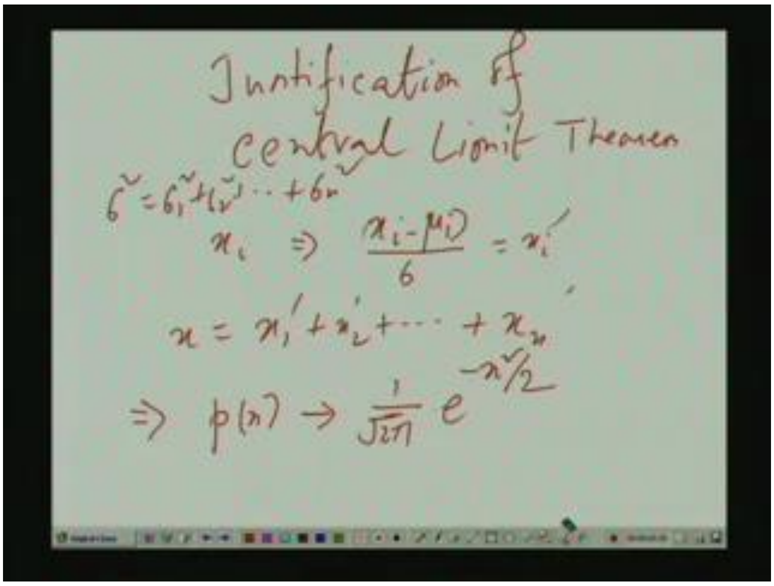
(Refer Slide Time: 41:55)



So, that means mu x is nothing but n times p and sigma x square is nothing but n times pq. We have already seen what is the exact probability density for x. That is say what are the values x can take. X can take either 0 when each of x 1 x 2 up to xn takes 0, that is always tail occurs. When a head occurs only once and tail occurs in the rest of occasions, in that case x takes the value 1. Then, 2 means head occurs twice and rest of the occasions, you get tail and likewise. So, x can take as I said earlier values either 0 or 1 or 2 up to n. So, p of x equal to k, we have seen earlier is nothing but it is that binomial distribution n ck p to the power k q to the power n minus k. That is the actual definition, but going back to central limit theorem if n is large, that is there should be approximately equal to 1 by root 2 pi square root sigma a square is pq. So, sigma x square is npq. So, square root npq e to the power minus x minus np whole square by twice npq, this result has a name. This is called Laplace De Moiver theorem.

Now, in an outline of the proof I mean I do not have that much time to carry out the proof, but let me get in to it. So, instead of calling it a proof, I will rather call it justification because it is not really a proof in the strict mathematical sense. It is just logical way of justifying this central limit theorem. It is just a logical approach.

(Refer Slide Time: 44:31)



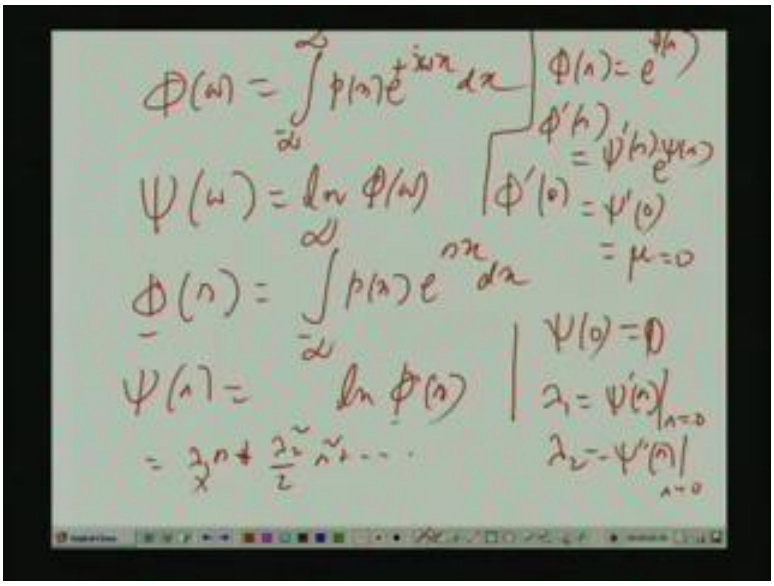
That is I will call it now before that I will just modify the things a bit. Instead of taking a variable xi, let me take it as xi minus first xi minus mu i. So, then it will have this. This variable will have mean 0, and then if I divide it by sigma and I call it xi prime, then xi prime is a random variable with mean 0 and variance 1. Such xi primes are supposed to be added. So, suppose x is nothing but x 1 prime plus x 2 prime, sorry. Then, central limit theorem wants us to prove that p of x for large n p of x will tend to be x. Then, mean of x is obviously I will change it a little bit. I will not may be it will be better if I do not divide by sigma i, but rather I divide it by sigma.

I will tell you what sigma is. Sigma square is nothing but sigma 1 square plus sigma 2 square plus dot dot dot plus sigma n square. So, if you add up all the individual variances, you will get the variance sigma square and suppose xi defined to be xi minus mu i by sigma, so obviously xi prime. Now, I mean x prime I mean defined to be like this. So, obviously xi prime has zero mean. At least that much is sure. So, x also has zero mean. What is the variance of x? These are statistically independent. X i mean x 1 prime will have a variance is equal to sigma 1 square by sigma square. X 2 prime will have a variance sigma 2 square by sigma square dot dot dot sigma n square by sigma square and if you all add of all them, then you get 1. So, x will be random variable with 0 mean and variance 1.

Now, in that case context, central limit theorem would say that for large n probability density of such x should approximate this function x square by 2. Sigma is 1 and mean is

0. So, the approximate this is what we have to prove. So, remember I am talking of x which consists of several random variables which are statistically independent having zero mean variance of this form, so that the total summation is again a random variable zero mean and you need variance. This is what we have to prove. First we consider we will be doing it in this domain, i mean using characteristic function. Since, time is short today, I will be able to just go away in to it half way and I will continue from here in the next class.

(Refer Slide Time: 48:34)



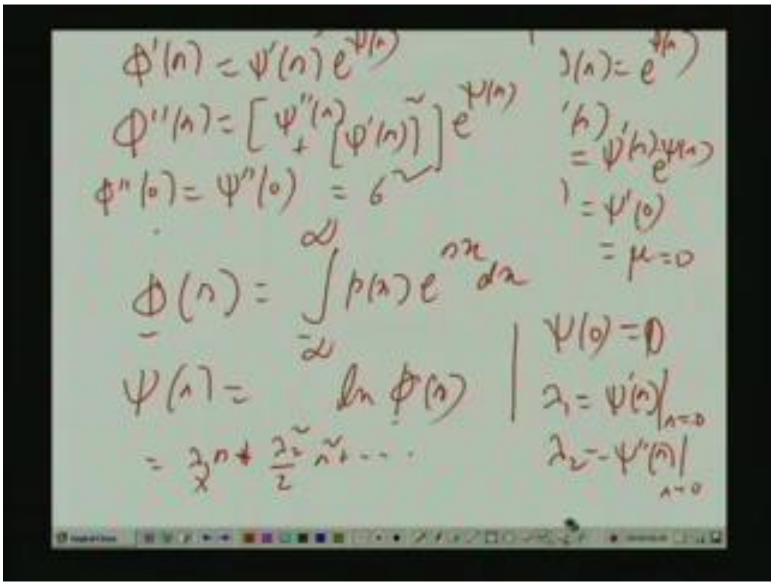
What is happening is we have to just recall certain things from characteristic function theory. That is what phi omega is. It is nothing but px e to the power e to plus j omega x dx minus infinity to infinity, and then we defined the second characteristic functions phi or may be no. We defined second characteristic function psi omega as nothing but ln phi omega. Similarly, we had phi s. It is a more general function e to the power sx dx and psi s, sorry was ln phi s, right. This we have seen. This is just repetition. Then, phi s can be expanded in McLaren series.

What is psi 0? That is ln phi 0. That is integral px dx which is 1. So, phi 0 is 1. That means, psi 0 which is ln of log of 1, that is 0 psi 0 1. So, this McLaren series will be nothing but psi 0 which is sorry, this is 0 log of 1 is 0. Phi 0 is 1 integral px dx, that is 1. So, log of that is 0 that is psi of 0 is 0. So, McLaren series of this will be what? It is psi 0 first, which is 0. So, forget it. Then, one term may be you can call it lambda 1 s. Then, another term may be lambda 2 square by 2 s square plus further term.

What is lambda 1? Lambda 1 was psi prime. This is typical McLaren series psi prime s at psi equal to 0. What is lambda 2? It is psi double prime s at s equal to 0. This is McLaren series. What is psi prime s? We have to see we have done this earlier, but still it is worth we do it again. What is psi prime s? We know that phi s, I mean psi s is ln phi s or phi s is e to the power psi s. That means, phi prime s is nothing but psi prime s times e to the power psi s. So, phi prime 0 and you know e to the power psi 0. Psi 0 is 0 e to the power 0 is 1. So, this is nothing but psi prime 0. What is psi prime zero? It is phi prime 0. What is phi prime 0? First one is phi prime s. If you differentiate this integral with respect to s, what you get is x times px e to the power sx dx. If you differentiate this with respect to s, x comes out.

So, xp x e to the power sx dx and then put s equal to zero mean e to the power 0, that is 1. So, x into px dx integral which is nothing but mean of x. So, phi prime 0 is nothing but mean of x that is say mu, but in our example we are considering a random variable x which has zero mean because all the constituents x 1 prime up to xn prime, they have zero mean. So, in our case it is 0. So, that means lambda is nothing but psi prime s with s equal to 0. So, that is 0. So, this is 0. We consider this term. What is lambda 2? Lambda 2 is this. I do not think I will be able to complete it today, but I will just do a little bit and then call off. I will start from here in the next class. Just bear with me for this.

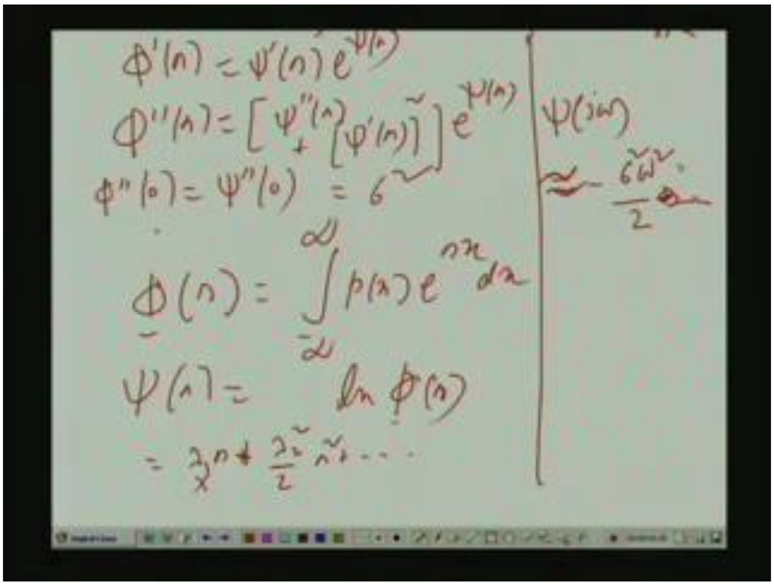
(Refer Slide Time: 53:37)



What is lambda 2? You have seen already earlier what is lambda 2. That is we know that phi prime s we have seen is nothing but psi prime s e to the power psi s. If we differentiate again phi double prime s is nothing but psi double prime s e to the power psi s. Take that outside and psi prime s whole square e to the power psi s. Put s equal to 0. If you put s equal to 0 e to the power psi s, that is e to the power psi 0. That is equal to 1 because psi 0 is 0. So, 1 psi prime 0 we have already seen is 0. So, forget that because our mean is 0. So, only psi double prime s that is psi double prime 0 that is nothing but phi double prime 0.

What is phi double prime zero? What is phi double prime s? You differentiate it. Once x into px e to the power sx differentiate, again another x comes x square px e to the power sx. Now, put s equal to 0. So, just s square px dx integral which is nothing but variance of x because x has zero mean. So, that is variance of x. So, this is nothing but sigma square. So, lambda 2 square gives us the sigma square of the random variable. So, this is for psi s. So, how is psi j omega? You just replace s by j omega. So, you get minus sigma square.

(Refer Slide Time: 55:19)



So, what is psi j omega? It is nothing but minus sigma square by 2 sigma square omega square by 2, minus sigma omega square by 2 plus other terms. Of course higher order term, but you know why I am now considering omega very close to the origin, so then higher order terms which have higher powers of omega, so omega to the power 3 omega to the power 4. Suppose we can ignore them. So, then it can be approximated to be like this. It can be approximated to be like this.

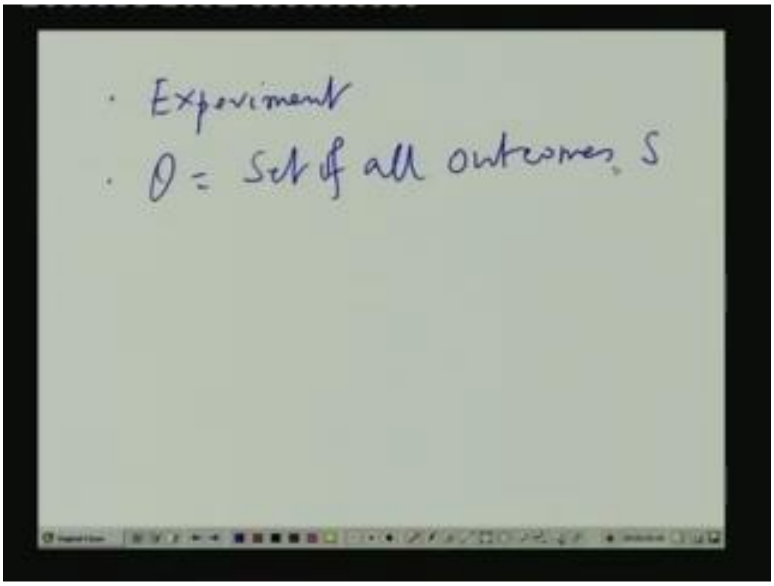
So, I will stop here today. I mean I will continue from here in the next class. Just remember this that around origin for zero mean random variable around origin, this psi j omega can be approximated as minus sigma square omega square by 2. I again repeat for any random variable which has zero mean and variance sigma square, the corresponding second characteristic function psi, the omega can be approximated around origin by a function like minus sigma square omega square by 2. This is irrespective of whatever may be the probability density of the random variable x. So, I stop here today from here.

Thank you.

So, today we begin a new topic all together. You know so far we have been discussing random variables. See one random variable, then two random variables and multiple random variables and all that. Of course, their probability density functions, probability distribution functions and various relations and properties associated with them. Now, we use those concepts to explain or to interpret various natural phenomena, and that takes us to study of random process or which is also called stochastic process.

What is a stochastic process? Essentially you remember how we defined random variables. I mean there were some experiments you know. It is not that it could be really experiment, but it is just kind of we need just assume that I mean there is an experiment and there are various outcomes that are coming out depending on the trial, and with each outcome, you assign a value to a variable. So, that variable is then called random variable. Similarly, suppose we have got some experiment that is going on and every trial there is some outcome and with each outcome S.

(Refer Slide Time: 58:10)



That is first experiment that may be say theta is a set of all outcomes S. S is a typical outcome. So, whenever a particular outcome, any experimental I mean when a particular trial takes place, you get some outcome S and depending on what you get, you assign a value to the variable x and that variable is called a random variable x, but instead of assigning a value to the random variable, suppose depending on the outcome S, we generate a function x of T. A particular function we assign instead of assigning a value to a, to this outcome S.