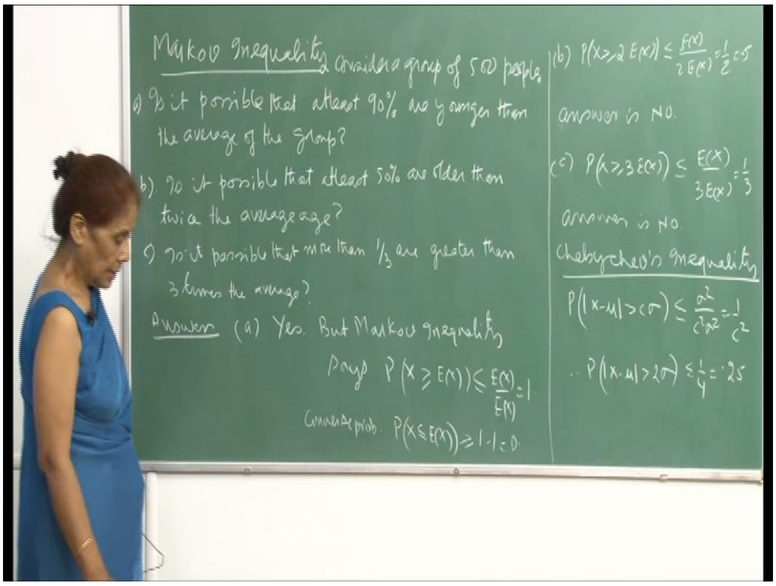
Introduction to Probability Theory and its Applications
Prof. Prabha Sharma
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 20
Convergence and Limit Theorem

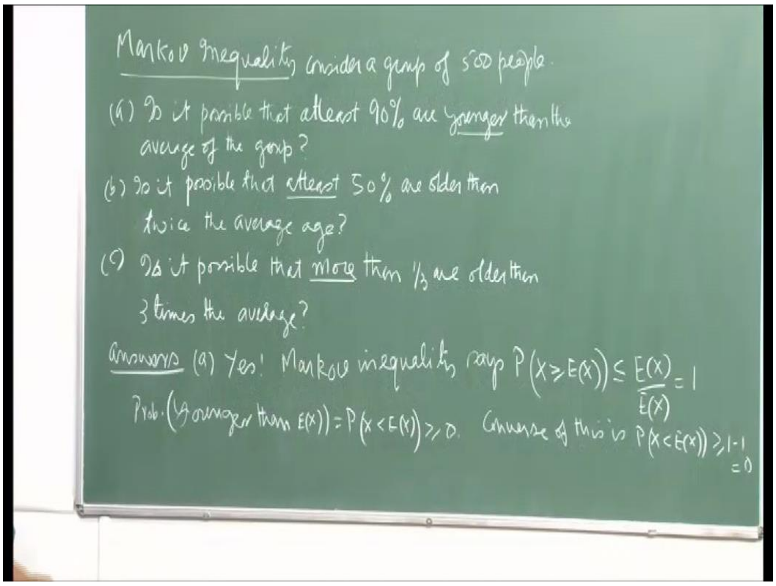
So, in the last lecture, we had introduced these inequalities – Markov inequality and Chebychev's inequality.

(Refer Slide Time: 00:15)



But I feel that revisiting them is necessary because some aspects need to be emphasized. And in fact the Markov inequality has its strength, and its simplicity and its generality, because the inequality is very simple to state, but this can be very useful and powerful at places and also the strength lies in its generality, because it just that you need to know that there is a random variable, whose expected value exists; and that is it. And then you can you know state facts about certain probabilities.

(Refer Slide Time: 01:01)



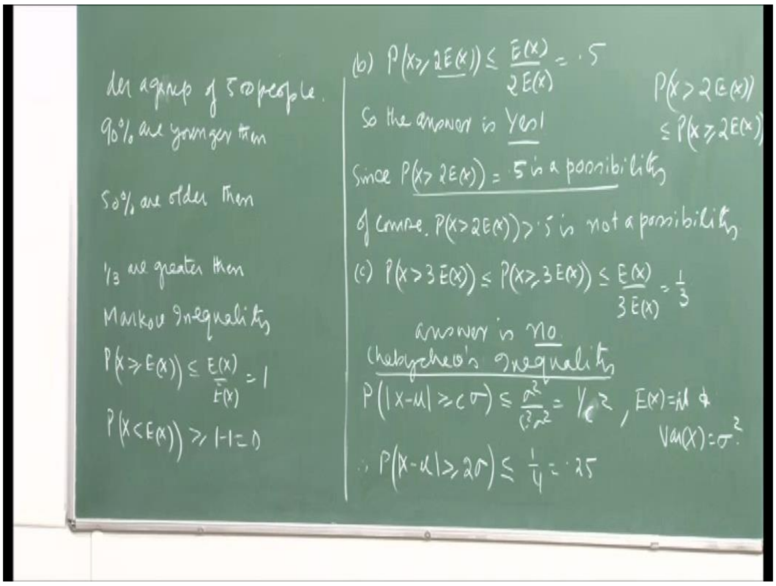
So, let us see interesting applications of the Markov inequality. Consider a group of 500 people. Now, the kind of… You are going to ask this question – is it possible that at least 90 percent are younger than the average of the group. Then, the next question is – is it possible that, at least 50 percent are older than twice the average age. And, another question could be is it possible that more than one-third are older than three times the average. So, let us try to see what kind of answers Markov inequality will give you. So, for the first part, of course the answer is yes and I will explain why. But, if you look at the…

If you try to get the bound from the Markov inequality, the inequality says that, for X greater than or equal to E X – probability that X is greater than or equal to E X will be less than or equal to E X upon E X, because you take E X of this and then divide by this, which is equal to 1. So, that is no bound, because you know that, all probabilities are less than or equal to 1 and the converse of this event would be probability X less than E X, which would then be greater than or equal to 1 minus 1 – converse of this, because this is less than or equal to 1. So, this will become 1 minus 1, which is 0.

So, again, does not give you any information. So, that is what we are trying to say. We are saying that, possible that, at least 90 percent are younger than. So, younger than means that, you want to compute the probability of the event that, X is less than E X – younger. This is what you want to compute. So, I should have said here this is comma.

Therefore, the Markov inequality just tells us that, this is greater than or equal to 0. So, that is no help. But, of course, you can rationalize the string that, the answer would be yes, because there may be some people who are very old; and therefore, they will make the average go up to… So, even if 90 percent are younger; that means what we are saying is that, the answer to this is yes, because 90 percent are younger. Even then the few people, who are very old, will lift the average. And so this inequality would be… This is the probability of 90 percent are younger would be satisfied; that means probability X less than E X is equal to 90.9 would be satisfied.

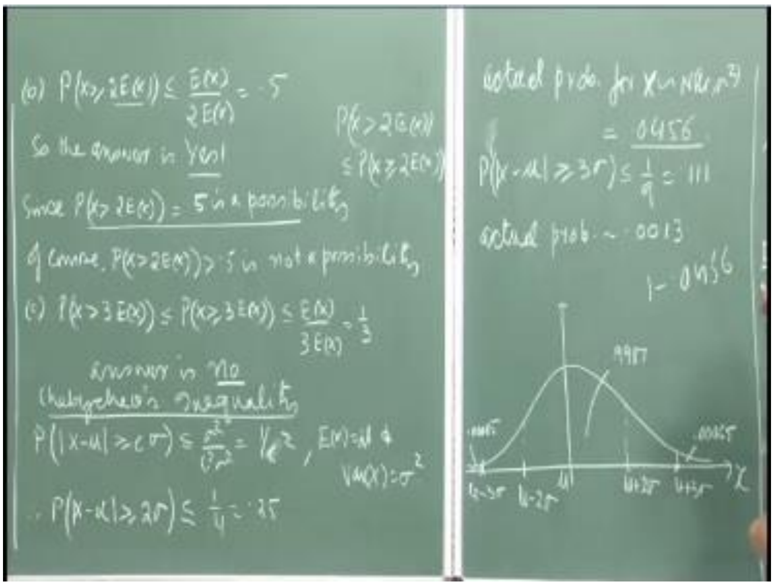
(Refer Slide Time: 03:33)



So, to answer the second question, what we want is that, older than twice the average age; that means you want the probability X greater than twice E X; and you want a bound that, this is at least 50 percent people are older than twice the average age. So, if I want this probability, then this is less than or equal to probability X greater than or equal to twice E X, because this event is bigger than this event. And, this why Markov's inequality would be less than or equal to E X upon twice E X. So, we divide by this. And so, that is equal to 0.5. Therefore, the answer would be yes, because probability that, X greater than 2 E X equal to 0.5 is a possibility. Yes, but probability X greater than 2 E X greater than 0.5 is not a possibility. But, since this is possible, we will say that, the answer is yes that, at least 50 percent will be older than twice the average age. So, interesting applications.

Then, to answer the third part, that is, probability X greater than 3 times E X; and you want a bound on this. So, this is less than or equal to probability X greater than or equal to 3 times E X; same argument is earlier. And, this by Markov's inequality is less than or equal to 1 by 3. So, here you want that, at least 1-3 are greater than the probability; that, at least one-third are greater than thrice the average age. So, the answer is no, because this is less than or equal to 1 by 3. So, this cannot be more than. So, this event – the probability of this event cannot exceed 1 by 3. So, the answer here is no. Now, similarly, let us look at the Chebychev's inequality, which says that, probability – absolute value of X minus mu greater than or equal to c times sigma is less than or equal to sigma square upon u divided by the square of this, which is c square sigma square. So, this is equal to 1 by c square – 1 by c square. And so, if you consider the event that, probability of absolute value of X minus mu greater than or equal to twice sigma, then this will be less than or equal to 1 by 4, which is 0.25.

(Refer Slide Time: 06:21)



So, now, if you look, compare this with the some of the actual probabilities, then for X being distributed as normal mu with mean mu and variance sigma square. And, you are looking at the probability that, absolute value of X minus mu is greater than or equal to twice sigma; then the actual probability is 0.456. Therefore, you can see that, this is much smaller than 0.25 and if you look at the diagram. Therefore, if this is the mean, the x-axis is this; and then this is the PDF – axis for this PDF. Then, you see here you take the area; that means what you are saying is that, this area lying between, because absolute value X minus mu greater than or equal to 2 sigma means that X lies between mu minus 2 sigma and mu plus 2 sigma. So, these are the limits.

And so, here what we are saying is that, this area would be 1 minus 0.0456. This is the area, which we are depicting here. And so, difference is quiet large. And, this becomes even more significant or more glaring the difference between the Chebychev's bound and the actual bound or the actual probability. If you take the probability of X minus mu greater than or equal to 3 sigma, then this will be less than or equal to 1 by 9, which is 0.111 by the Chebychev's inequality. But, the actual probability is actually very small; it is 0.0013, which is… See what here again… Because of the symmetries remember; so, here this would be mu minus 3 sigma and this is mu plus 3 sigma. So, you are asking for… Exactly. So, that area I am showing that means between mu minus 3 sigma and mu plus 3 sigma. So, this whole area I am saying is 0.9987. And, that is because we know that, by symmetry, this area – the tail – this part – tail part, and these two are the same. And so, we have discussed this many times before also. Therefore, that means actually, the tail… that means this tail area is half of this – 0.0065. And, here also the tail is 0.0065. And so, therefore… So, the difference becomes bigger and bigger.

One can go on and looking at these interesting parts that these inequalities. But, at times, they provide you… They are very useful tools and they… As I told you, for the Markov inequality, it can answer some very interesting questions. And, here also we will see various applications of the Chebychev's inequality. Markov inequality is not able to say much, but you can see… The thing is that, the answer would be yes, because you can always have small number of people who are very aged, whose ages are very big. And therefore, the average… Therefore, the 90 percent can still be younger than the average age, because these older people – they pull up the average. Therefore, the answer is yes.

Now, if you look at the second question, then you are asking for the probability that, X is greater than or equal to twice E X. So, twice the average age. And therefore, by Markov inequality, this would be E X upon 2 E X, which is 1 by 2, which is 0.5. So, Markov's inequality gives you the bound that, this probability cannot exceed 0.5. And so, therefore, the answer here will be no. So, the answer is no, because here they are asking is it possible that, at least 50 percent are older and twice the average age. So, no; 50 percent will not be older. So, this probability would be always less than or equal to 0.5. And similarly, for the third question, probability X greater than or equal to 3 times E X –

that will be less than or equal to E X upon 3 E X; it is 1 by 3. Therefore, again more than 1 by 3 is not possible; more than 1 by 3 are greater than 3 times, because this probability – the bound – upper bound is 1 by 3. And therefore, again the answer is no. So, I just thought that, this gives you another insight into the Markov inequality and its uses. And, one can go and discover more and more about the usage of this particular inequality.

Now similarly, for Chebychev's inequality, I wanted to just point out that, if you ask for the probability that mod of X minus mu is… Therefore, you have a random variable X, which has accepted value as mu and variance X is sigma square. So, just a random variable with mean mu and variance X sigma square; you are asking the question mod of X minus mu or absolute of X minus mu is greater than C sigma. So, Chebychev's inequality – this would be sigma square upon c square sigma square; this is 1 by c square. So, in particular, if you put c is equal to 2, then this is a probability that, mod of X minus mu is greater than 2 sigma.

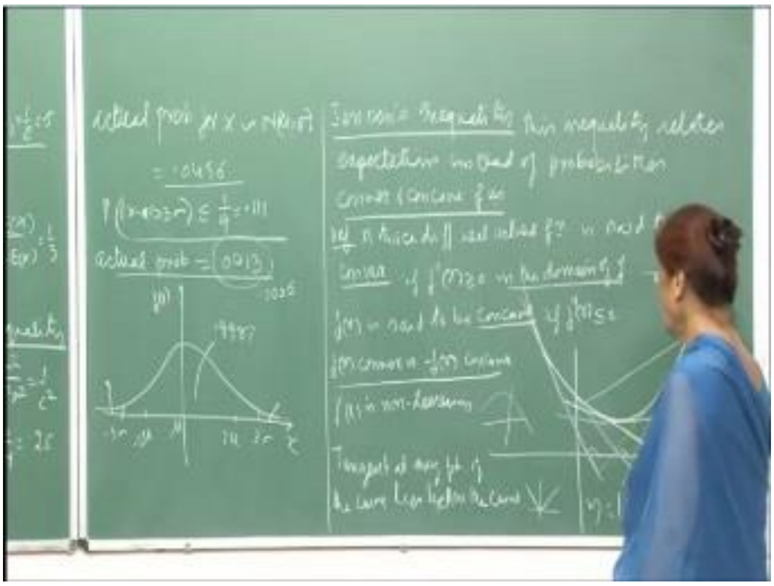
And therefore, this will be less than or equal to 1 by 4, which is 0.25. So, in other words, here if you… I have drawn the normal curve; does not matter. Therefore, this is minus 2 mu and this will be 2 mu. So, in this, we are asking for the area, that is, the probability that, this is greater than 2 sigma; that means, the area on to the left of minus 2 mu and the area to the right of 2 mu. So, that will give you the probability that, mod of X minus mu is greater than 2 sigma. And, this is less than 1 by 4 in general; universally true. This is universally true, which is 0.25.

Now, if you compare this for normal n mu sigma; that means, if a random variable X is mu sigma, then this probability is 0.0456. Therefore, compared to this, this is rarely loose bound – loose upper bound. But, later on we will see how… No matter… Because of its universality – Chebychev's inequality, this is very useful improving many other results in probability theory. So, anyway I just thought I will give you an estimate, because the normal curve is symmetric about mu and then it is bell shaped. So, the mass is concentrated around mu for normal. And therefore, this probability would be small, because the area lying on the left of minus 2 mu and to the right of 2 mu will be much smaller than compared to the area, which is around mu. Therefore, this…

And similarly, if you take c to be 3, then the difference is more marked, because probability mod X minus mu greater than 3 sigma is less than or equal to 1 by 9, which is

0.11. Anyway for… So, that means, it says is that, for most of the distributions, the area – the mass of under the curve lies the probability mass – lies within minus 3 sigma. This is minus 3 sigma and 3 sigma; then the area inside here is 0.9987. So, only this much area lies outside; which means half of this. I will have to be very sure that, this is this; then the half of this half; that means, further do it 0.0006. So, this is the area, which lies here and the both. This area is 0.006 and that is 0.006. So, this is the idea. Therefore, Chebychev's inequality is an upper bound; but, because it is applicable to all the distributions, therefore, it has its own uses and applications.

(Refer Slide Time: 14:46)



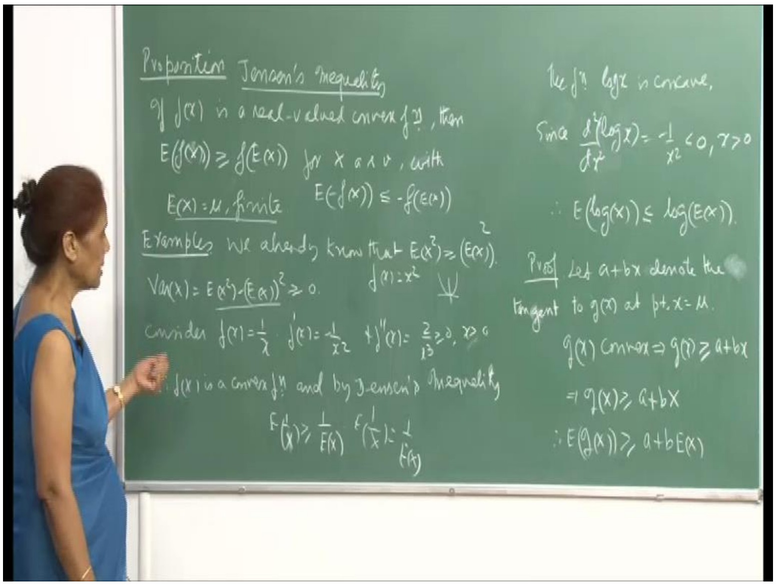
Now, the third inequality that we want to talk about is Jensen's inequality. And, this inequality relates expectations instead of probabilities. So, like for example, both these inequalities were giving you upper bounds for the probabilities of certain events. But, Jensen's inequality relates the expectations. But, before that, before I give you the Jensen's inequality, I need to define convex and concave functions. And, some you may have already come across, for example, convex lenses, concave lenses – you may have heard of. So, here the function is said to be convex or if it is twice differentiable… If a function is twice differentiable, real valued function. And, it is said to be convex, if its second derivative is non-negative in the domain of f. So, wherever f is defined, then at all those points if you are f double prime x is non-negative, then the function is said to be convex. And, if the double derivative is less than or equal to 0, then the function is said to be concave. So, therefore, the relationship between convex and concave is that, if f is convex, then it will imply that, minus of f x is concave.

So, now, here for example, I have drawn for you convex function twice differentiable. And, what we are saying is that, if f double prime x is greater than or equal to 0, then f prime that… This implies that, f prime x is not decreasing if the… Wherever you take a function f and if its derivative is non-negative, then we say the function is non-decreasing. Here f double prime x is non-decreasing. So, this implies that, f double prime x is greater than or equal to 0; that implies that, f prime x is non-decreasing. So, you see here for example, these are the tangents to the curve; and see these angles – they are negative; they are obtuse. And, if all of you remember the graph of tan x, because slope is given by… f prime x is a slope; tan of the angle – tan of this angle; tangent of the angle that, the tangent at the curve makes. So, you are…

For example, this if you take this is 0; this is pi by 2; then this is pi. And therefore, on this side of this, it is like this. So, the function is obtuse angle and the curve is increasing. So, as the angle becomes… And, then of course, this becomes… The angle becomes up to pi; and so, tan of pi is 0. So, you are derivatives – the tan of these angles are increasing. And then finally, at this point, it becomes 0. And then when you take this, then you can see that, the angles are increasing. Therefore, for obtuse angles, again, tan is increasing. So, this is the idea. Therefore, the first derivative is non-decreasing. Also, that the tangent at any point of the curve lies below the curve, because you have seen. See the function is like this. So, the tangent is this. So, tangent is always below the curve. And so, here when you say that, minus f x; minus f x means you will turn it upside down; you overt. Therefore, a convex function you can say holds water; a concave function will not hold water, because it will be upside down. So, this thing will be up and a function will be like this. So, this will be a concave function.

Now, of course, here I have given you the definition of a twice differentiable. But, for example, if you take y is equal to mod x, this is also convex. But, of course, this is not differentiable. So, none of these things… It is differentiable at these points, but not at the origin. So, this holds, because it is constant. See here the slope is minus 1; here the slope is 1. So, in any case, the slope is increasing, because this is this; here it is not defined, but the… So, this is also a convex function. And of course, there are many ways of characterizing a convex function. So, now, I will state the Jensen's inequality for convex and concave functions.

(Refer Slide Time: 19:17)



So, the Jensen's inequality says that, if f x is a real-valued convex function, then expectation of f x – this should be capital X, because function f is a function of the random variable X. Then, E – expectation of f of X is greater than or equal to f of E of X; that means you exchange f and E; then the inequality is this kind. So, for X random variable with E X equal to mu finite. So, the requirement is that, the mean – the expected value must exist for a random variable; and if a function f is convex, then this would be E f X is greater than or equal to f of E X.

Now, you can see that, if you replace this by… If you multiply the inequality by minus sign, then the minus sign will go inside and it will say that, expectation of minus f of X is less than or equal to minus f of E X. And, since… So, minus as we said earlier when we were defining a convex function that, minus f will be concave if f is convex. Therefore, for the concave function, the inequality reverses. So, this is the Jensen's inequality. So, it is just relating the expected values. And, you can… If the function is convex, then the inequality would be greater kind; and for concave, it will be less kind.

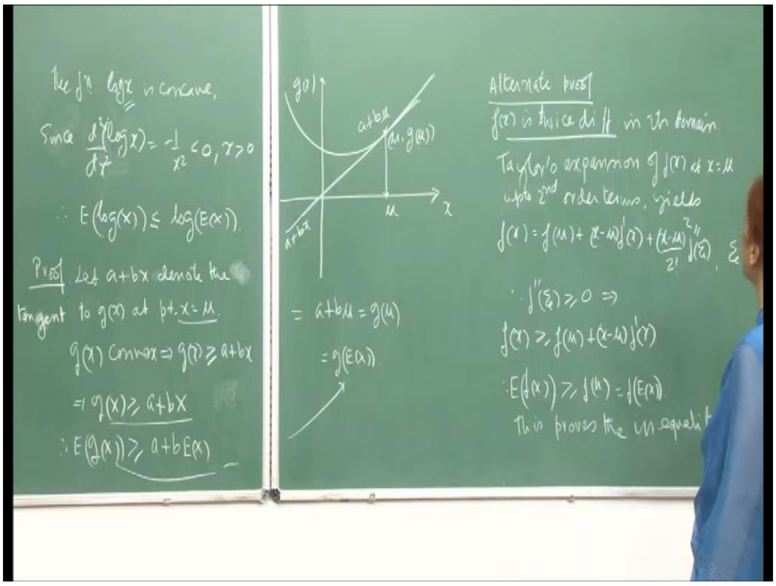
Now, you already know that, expectation of for example, X square – if the second movement exists, expectation X square is greater than or equal to expectation of X whole square; that means the function f X here is X square and this we know is convex;

everybody knows it is a parabola or the second derivative is 2 – a constant, which is non-negative. So, this is a convex function, but we already know that, variance X can be written as expectation X square minus expectation X whole square, and this is always non-negative. So, from here also, it follows that, expectation X square will be greater than or equal to square of expected X.

Consider the function f X equal to 1 by X. Then, if you just find out the first derivative, this is minus 1 by X square; and second derivative would be see X raise to minus 2. So, minus 2 and minus sign – plus 2 upon X cube. And, this is always non-negative for X positive. And therefore, this is a convex function. And so, by Jensen's inequality, expected value of 1 by X is greater than or equal to 1 upon expected X. And, quite a few people often mistake this and they say that, expectation of this will be… So, now, you know better, because the Jensen's inequality says this will be greater than or equal to; they are not the same thing; expectation of 1 by X and 1 by expectation X are not equal. So, this is also you can now assert by using Jensen's inequality.

You can consider the function log X. log X – the second derivative is minus 1 by X square; first derivative would be 1 by X. So, when you take the second derivative, it will be minus 1 by X square. And, this is less than 0 for X greater than 0. Anyway the function – this is defined for X positive. And so, by Jensen's inequality, expectation of log of X is less than or equal to log of expectation of X, because for concave function, the inequality reverses. Proof is simple.

(Refer Slide Time: 23:09)



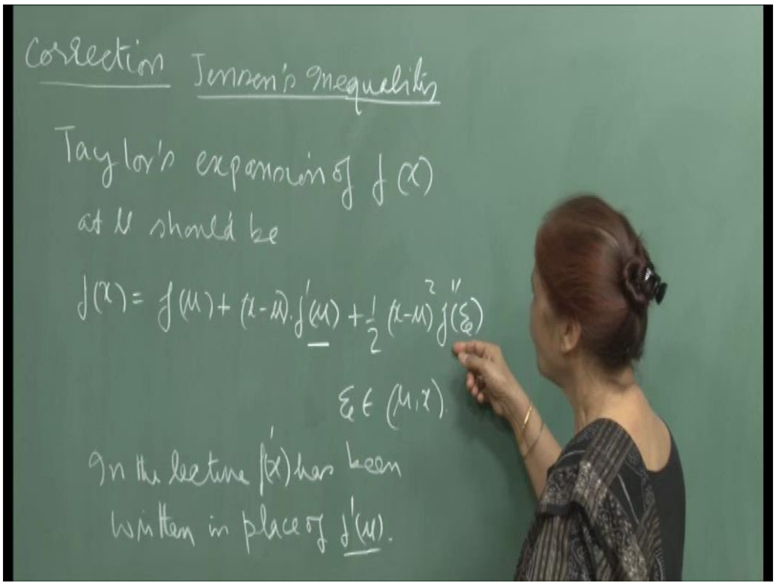
So, I will use the first property that, the tangent at any point of a convex function lies below the curve. So, the curve always goes… – it is above the ((Refer Slide Time: 23:18)) And, of course, they meet at this point. So, the tangent is at the point mu; then the value here – the coordinates are mu, g, mu. And so, if I take a plus b x as the tangent to g x at the point x is equal to mu; then g x convex implies that, g x is always is greater than or equal to a plus b x and g mu will be equal to a plus b mu, because the curve and the tangent line – they meet at this point. And therefore, since these holes… Therefore, when I replace x by a random variable, the inequality remains intact. So, g of random variable x is greater than or equal to a plus b of X. And therefore, the expectation will also… They will not change the inequality.

So, when I apply expectation on either side, it will be E of g of x is greater than or equal to a plus b E of X; a and b are constants. So, this is what the proof. And so, a plus b of E X is a plus b mu, which is g of mu; and mu is your expected value. Therefore, this is g of E of X. Therefore, from here you have shown this inequality; the simple proof using the convexity of the function; and then the fact that, when you have inequality. So, this a bigger function than this. So, I hope you all agree that, because even if you are taking X to be a continuous random variable, then if the density function of course, is non-negative. So, here you are taking the difference. So, if you take the difference of g x minus a minus b X, which is a non negative function; then integral – whatever the limits would be also non-negative. And so, this will be satisfied. Therefore, from here to here is no problem. Therefore, you can prove the Jensen's inequality. Therefore, the figure is also quite explanatory.

Now, an alternate proof, because since we have the definition of convexity, I will use the twice differentiability of the function now. So, since f is convex; so, it is twice differentiable instrument. And, Taylor's expansion of f x at x is equal to mu up to second order terms yields. So, now, those of you who feel comfortable with calculus, then you know about the Taylor's expansion that, every function can be expanded in the neighborhood of a point; where, in the neighborhood, it has all these derivatives. And so, here since I have assumed that, it is second order derivative exists. Therefore, I can write f x as f mu plus x minus mu into f prime x plus x minus mu whole square by 2 factorial into f double prime psi; where, psi belongs to mu comma x. So, such a psi exists in the interval. So, whether it is mu comma x or x comma mu does not matter, because you are taking the square here. So, there is a psi in this interval. And therefore, this would be then exact expansion; that is what Taylor says. So, Taylor's theorem says that, such a psi always exists.

Now, since f double prime psi is non-negative, because f double prime is non-negative in the whole domain. So, this is non-negative and this is a square – square of a real number. So, this quantity is non-negative. Therefore, I can say that, f x is greater than or equal to f mu plus x minus mu into f prime x. So, which… If you write this in terms of… So, f of x is greater than or equal to f mu; I should have written the step x minus mu f prime x – f prime x; just that as we did here in this first proof. And now, you can take the expectation. So, expectation f x – again this is same reasoning; the inequality will not get reversed. So, this will be f mu plus. Now, expectation of x minus mu is 0. So, you are left with only f mu here. And, f mu is f of E X. Therefore, again the Jensen's inequality has been proved.

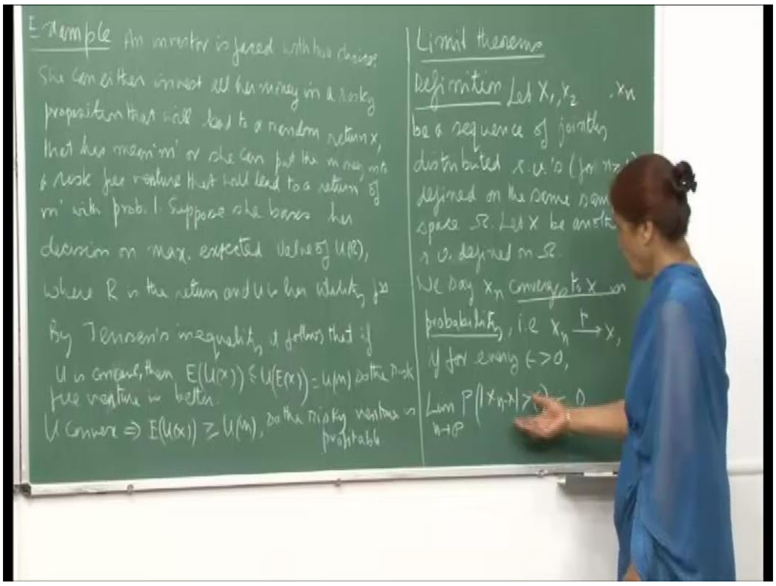
(Refer Slide Time: 28:00)



So, I just wanted to point out this correction in the Jensen's inequality proof. See I was giving you an alternate proof; and there I had to expand the function f x by Taylor's expansion at the point mu. And, the correct expansion is that, f x is equal to f mu plus x minus mu f prime mu plus half x minus mu whole square f double prime psi. Now, instead of mu, it got written as x.

Therefore, you have to read f prime mu instead of f prime x. And then of course, we know that, psi is a number, which is some number between mu and x. And, by Taylor's theorem, such as psi always exists. So, we are taking a second order expansion of the function f x at mu. And so, this should read as f prime mu instead of f prime x. And, as we go along, we might also see some more occasions to use this inequality. But, I think this gives you a good feeling about the Jensen's inequality.

(Refer Slide Time: 29:08)



So, an instructing example of the Jensen's inequality is that, investor is faced with two choices. She can either invest all her money in a risky proposition that will lead to a random written X that has mean m or she can put the money into risk-free venture that will lead to a written of m with probability 1. So, these are the two choices she has. And, suppose she bases her decision on maximizing an expected value of u R, where R is her return and u is her utility function. So, by somebody's advice or something, she has now decided that, she will base her decision to invest whether in the risk-free venture or the risky venture by maximizing the excepted value of u R; where, R is the return function and u is the utility function. So, u of R.

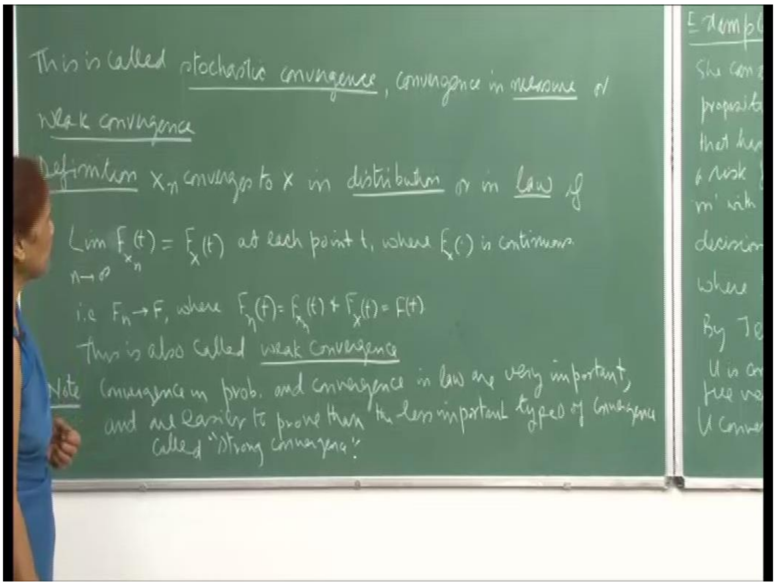
Now, by Jensen's inequality, it follows that, if u is concave, then expected u X will be less than or equal to u of E X, which will be u of m. So, the risk-free venture is better. So, here the expected return of u X will always be less than or equal to u of E X, which is u of m. Therefore, it is better to invest in the risk-free venture. Now, if u is concave, then this implies that, E of u X will be greater than or equal to u of m. So, the risky venture is profitable, because the expected return here would be greater than or equal to u of m. This is u, is her utility function; and in the risk-free venture, she gets exactly m returned. Therefore, this will be the total utility to her of the return that she gets from the risk-free venture. And, this is because X is a random returned. So, E of expected value of u X. So, that will always be greater than or equal to u m in case the utility function is convex.

Therefore, the risky venture is profitable. And, there can be many more interesting examples of these inequalities that we have just studied.

So, the next thing that we want to talk about, which again has a very important role to play; and these are the limit theorems. And so, let us just first try to understand the concept of what we mean by these limit theorems. So, the first definition that I want to make is the definition of sequence of random variables converging in probability to another random variable. So, here this is at X 1, X 2, X n, is a sequence of jointly distributed random variables for n greater than or equal to 1; that means you must have at least more than one defined on the same samples space omega. And, let X be another random variable defined on omega. Then, we say that, X n converges to X in probability, that is… So, the notation is that, X n goes to X in probability if for every epsilon greater than 0, limit of this absolute value X n minus X is greater than epsilon. So, this limit converges to 0.

So, in other words, in probability, the random variable X n is converging to X. And, please understand. So, here this is different from the concept of usual limit, where the P is missing. So, in that case, when you say that, in value X n, the sequence is converging to X; that means when n becomes larger and larger, the distance between X n and X will be very small, because epsilon is an arbitrary number greater than 0. So, I can go on making epsilon small and small. But, here the limit is in terms of probability – probability of this event; that means of this difference – X n minus X greater than epsilon becomes an impossible event, because the probability is 0. So, this is the idea of convergence in probability.

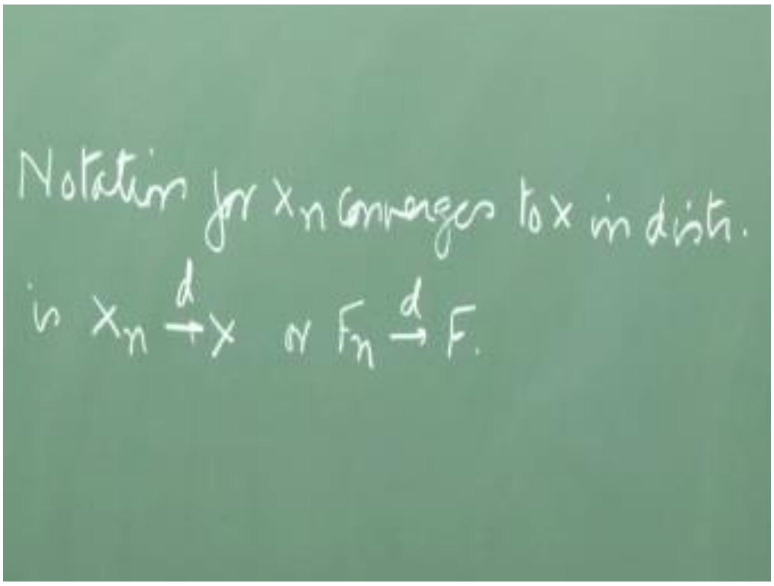
(Refer Slide Time: 33:16)



Then, the other definition that I want to make is that of… And, this is called… This convergence in probability; I have already given one name; it is also called stochastic convergence – convergence in measure – measure is the probability here or weak convergence. So, this is one definition. And, the other is the convergence in distribution. So, we will say that, X n converges to X in distribution or in law if the limit of F X n t; that means the cumulative distribution function of X n.

So, at the point t, converges to the distribution – cumulative distribution function of X at t as n goes to infinity. And, this must happen at each point t, where F X is continuous; so, that means… And, in fact, obviously, this is also continuous at that point. So, limit F X n t – the cumulative distribution function of the random variable X n – this converges to the cumulative distribution function of F X t of X as n goes to infinity. So, now, abbreviating the notation. So, this says that, F n goes to F; where, F n t is the cumulative distribution function of X n, and F we denote by the cumulative distribution function of X at t.

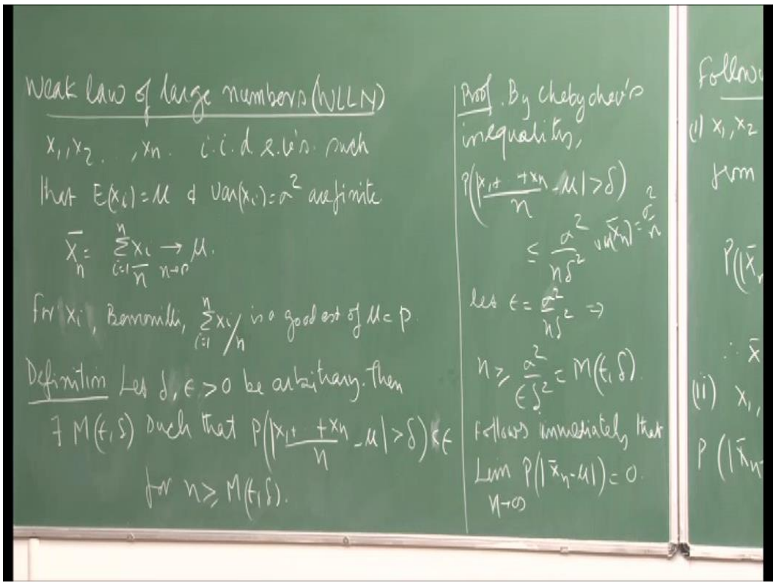
(Refer Slide Time: 34:40)



So, notation for X n converges to X in distribution. We also say that, X n going to X in distribution. So, the notation that I have written down or the cumulative distribution function F n of X n, which is F n going to F – the cumulative distribution function of X in distribution. And, d can also be replaced by l. So, both these notations are valid. So, this is also called weak convergence – weak convergence in law or weak convergence in distribution. So, you can see the difference, because here it is only we are saying that, probability of this event is becoming 0. As n goes to infinity, just the… whereas, here the whole distribution – the cumulative distribution function – the whole of the function is converging to the cumulative distribution function of X at every point t, where it is defined, where it is continuous.

Now, convergence in probability and convergence in law are very important. And, we will see as we go long that, the numerous applications of these convergences; and are easier to prove. Then, the less important types of convergence called strong convergence. So, maybe in this course, I have a chance to look at one or two strong type of convergences also. But, the more widely used are the weak convergences; and these are law and probability.

(Refer Slide Time: 36:19)



So, we will now define weak law of large numbers. Law of large numbers states that, if you have a sequence of these random variables – identically independently distributed random variables, I have said that, the expected value of each of them is mu and variance is sigma square and these are finite quantities; that means, the variance ((Refer Slide Time: 36:37)) Then, you define X n bar. So, X n bar would be the average of the values up to n. So, sigma X i; i varying from 1 to n divided by n. And then in simple terms, the weak law of large numbers says that, this sequence of averages X n bar as n goes to infinity; that means when you take n plus 1, it will be average of X 1, X 2 of X n plus x n plus 1. So, this is a sequence that you are generating by taking averages of n, n plus 1, n plus 2 and so, on. And then… So, this sequence converges to the mean of the… or the expected value of the random variables.

Idea here is that… So, actually this will happen in probability. So, the whole idea, because we say that weak law of large numbers. So, the whole convergence – the concept is in terms of probability. And so, what we are saying is that, since its converging in probability, the probability is high that… That means I can take… For large enough n, I can take X n bar as a good estimate of mu; otherwise, how do we have, because we just have these sample values, which we have taken randomly and then we are wanting to estimate the mean of the distribution. So, this would provide a good estimate for mean – for the value mu. For example, if all X i's are Bernoulli, then we know that, mu is of course, is a good estimate of mu; in the sense, this is also the probability P. If the probability of success is P, then for the expected value of each Bernoulli random variable, is also equal to P – the probability of success. And so, what it is saying is that, when you take n large enough, then this would give you good estimate of the probability of success. So, this law of large number provides way of estimating the mean of the distribution. This is the whole idea.

So, formally, if you want to define this concept that… then we will say that, given delta and epsilon greater than 0 – some arbitrary numbers, then there exists a number M, which is a function of epsilon and delta such that when you write this probability X 1 plus X 2 plus X n upon n, which is X n bar; X n bar minus mu in absolute value greater than delta. This probability will be less than epsilon for all n greater than or equal to the number dependent on epsilon and delta. So, this is simply just extending the notion of… Or, just the same notion that you have about continuity when you talk of continuous functions when you want to say that, the function values – this and this for example, can be brought as close as you wish.

So, this greater than delta will be less than epsilon provided for n begin up; that means n must be greater than or equal to some function, which is a function of number, which is dependent on which is a function of epsilon and delta. So, the whole idea is that, as long as… And, is large enough given the delta and epsilon, you will be able to say that, this probability greater than delta is less than epsilon. So, that means when I choose delta and epsilon small, then this is essentially saying that, the number X n bar comes close and close to mu. So, this is greater than delta whatever I mean… So, the event will become impossible, because if I choose epsilon very small, then this probability is very small; so, of this difference being greater than delta; so, in probability. So, the whole thing is being talked about in terms of probability. So, the proof is simple.

And, here I will use Chebychev's inequality. So, by Chebychev's inequality, this says that… Here as we have seen already that, for X n bar, the variance… because they are identically independently distributed, will be sigma square by n. And, the variance and the expected value of X n bar is mu. Therefore, this is X n bar minus is expected value. So, this difference in absolute value greater than delta would be less than or equal to sigma square upon n delta square. So, now, here I did say that, epsilon and delta are arbitrary, but see I can choose the epsilon to be sigma square upon n delta square. So, in a way, epsilon is a function of delta; that is ok. So, then this is… I will choose the epsilon to be sigma square upon and delta square. And then that will give me that, n must be…; that means this number if I denote by epsilon, then this probability is less than or equal to epsilon for n. So, from here n – the smallest value of n would be sigma square of epsilon delta square. But, for all n greater than this number, this inequality will be satisfied. And so, the number capital M epsilon delta can be chosen like this.

So, once we get that n is greater than or equal to sigma square upon epsilon delta square, this inequality is valid. So, what we have shown is that, given epsilon and delta greater than 0, we can find an n such that this inequality is satisfied for all values of n greater than or equal to sigma square by epsilon delta square. So, this is the M of epsilon delta in the definition for limit of the probability when we defined what we mean by limit in probability sense. So, then this is the M of epsilon delta. So, for all n greater than or equal to this given in epsilon and delta, then for all n greater than or equal to this number, this inequality will be satisfied. And therefore, it follows immediately that, this limit of probability of X n bar minus mu in an absolute value goes to 0 as n goes to infinity, because as n becomes larger and larger, I can choose epsilon smaller and smaller here. This was my… This is greater than or equal to delta here I have chosen; yes.

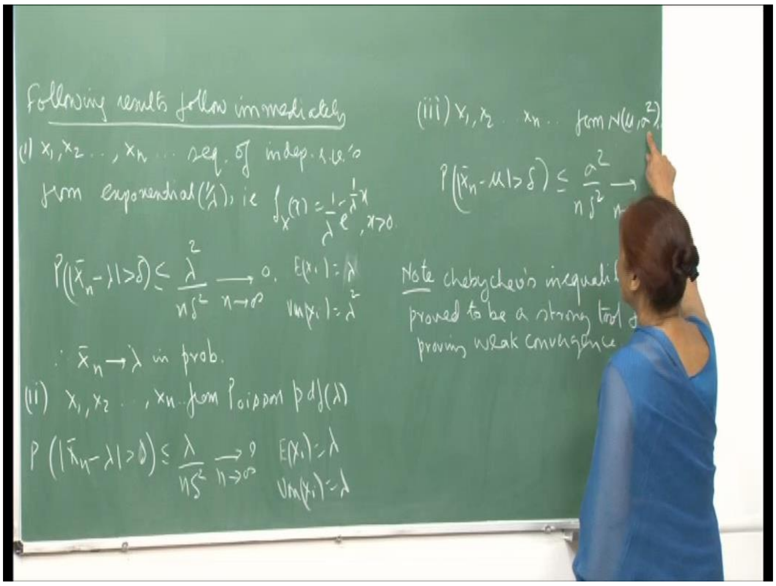
And so, in my definition for when I defined the limit of a probability, then we chose… This is the epsilon we chose – sigma square upon n delta square. So, what we are saying is that, this probability, that is, X n bar minus mu an absolute value greater than or equal to delta is less than or equal to epsilon. So, when I want… So, if I choose this equal to epsilon, then I am saying… And therefore, as epsilon becomes smaller and smaller, n will become larger and larger. And so, from my definition of limit in terms of probability, it follows that, this probability will tend to 0 as n goes to infinity. So, this is what we…

Therefore, you see again here that, I have made a very good use of Chebychev’s inequality to show you that, this probability – the limiting value of this probability of absolute value of X n bar minus mu will tend to 0 as n goes to infinity. Then, this satisfies the… So, by Chebychev’s inequality, this will be satisfied. And so, we have shown that, x n bar will converge to mu in probability. So, essentially, this is what… So, when you take the limit as n goes to infinity, then this number goes to 0, because as n goes to infinity, epsilon tends to 0. And therefore, this limit of the probability X n bar minus mu will go to 0 as n goes to infinity. So, essentially…

Now, of course, there can be different interpretations; and one of these students interpreted this as like if somebody who is practicing to be let us say a swimmer; so, what he will say is that, that means, no matter how hard I practice, my average performance will remain the same, because in probability, X n bar is converging to mu. So, that means he says that, there is no scope for improvement. But, again the fallacy in his argument is that, see here this result we are proving under the assumption that, X 1, X 2, X n – this sequence is independently identically distributed.

So, the identity part is not valid when you are practicing; obviously, these things are improving. So, your performance is improving every day. And therefore, to say that, you will never rise above the… that means, your average performance will remain the same no matter how hard you work, is not correct, because your ((Refer Slide Time: 46:31)) themselves are changing; they are no longer identically distributed. Therefore, this is not a good way to interpret the weak law of large numbers, but it certainly gives you a tool for estimating the value of the mean of the distribution from which the random variables are coming.

(Refer Slide Time: 46:55)



So, we can now look at these examples to see the application of the weak law of large numbers. So, for example, if the sequence is from exponential 1 by lambda; that means, they are all identically independently distributed random… These samples you are taking from an exponential distribution with parameter 1 by lambda, that is, the PDF is 1 by lambda e raise to minus 1 by lambda x for all x positive. Then, this probability – if you take it X n bar here; X n bar minus lambda in absolute value greater than delta would be less than or equal to again by Chebychev's inequality, because the… So, here expected X i is lambda – inverse of the parameter here, and variance X i is lambda square for the exponential distribution. Therefore, this would be less than or equal to lambda square upon… So, for the variance of X n bar would be therefore, lambda square by n. So, lambda square by n 1 upon delta square; and this goes to 0 as n goes to infinity. So, we can interact… We can choose… For any delta, we can choose epsilon as I showed you here; and it will satisfy the definition anyway. Therefore, what we are saying is that, X n bar would be a good estimate for large enough n, would be a good estimate for lambda for the mean of the distribution.

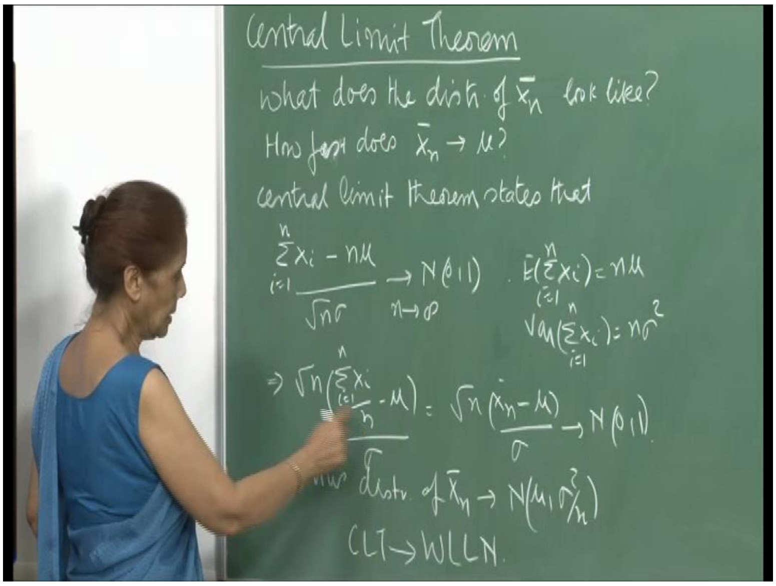
Similarly, if you have a Poisson… If you have this family; if the sequence is coming from a Poisson distribution with weight as lambda, then again this will be… So, here you have E X i is lambda. And, variance also is the same for a Poisson. So, this is also lambda. And so, for variance of X n bar would be lambda by n. And so, this probability greater than delta would be less than or equal to lambda upon n delta square. And, this will again go to 0, because lambda and delta are finite as we said that, we are talking about the situation, where the mean and the variance are finite. So, this will again go to 0 as n goes to infinity. And similarly, if you take this sample from… So, I am just giving you a few examples, but you will see that, this is universally true, because there we did not specify; we simply said they should be dependent identically distributed random variables. So, give 3 examples here.

And, if this sequence is from a normal mu sigma square, these are the sample values; then again this will be less than or equal to… So, now, here again E X i is mu and variance X i of course is given to be sigma square. So, variance X n bar would be sigma square by n. This will also go to 0 as n goes to infinity. So, Chebychev's inequality has proved to be a strong tool for proving weak convergence. And, we will see that, the other… I showed you application of Jensen's inequality also. And, we will also again look at some more limit theorems, where also we will make use of these inequalities. Therefore, the whole idea is that… Again one needs to emphasize the fact that, we are not saying that, the value that, the X n bar will… In value tend to mu, what we are saying in probably – it will tend to… Therefore, when we say it is a good estimate, this is in terms of probability; if the probability is very high – of this number becoming closer and closer to mu…

So, again, as I said, matter of interpretation, you might say that you go to a casino and you go on putting money in the machines – slot machine; and say for a number of times, you are not successful; so, you will say that, no, it will soon happen. But, that is not true, because again it is the matter of probability. Yes, the probability is high, because the event is getting impossible; I mean this probability is getting to 0; that is fine. But, it may happen that you may have to go on playing at the slot machine for a long time before your luck turns; that means the things change.

Therefore, one should not say that, yes, surely, what we are saying here is that, it will happen; that means if you flip a coin and you keep getting tails; then surely after sometime you will get heads also. But, it does not say when. And, this is a matter of… So, the important thing to understand is that, we are talking in terms of convergence in probability. And so, this gives you a good way of estimating the mean of the distribution; that means you go on taking large enough samples and then you take the average, and that will give you an idea of what the mean of the distribution is.

(Refer Slide Time: 52:43)



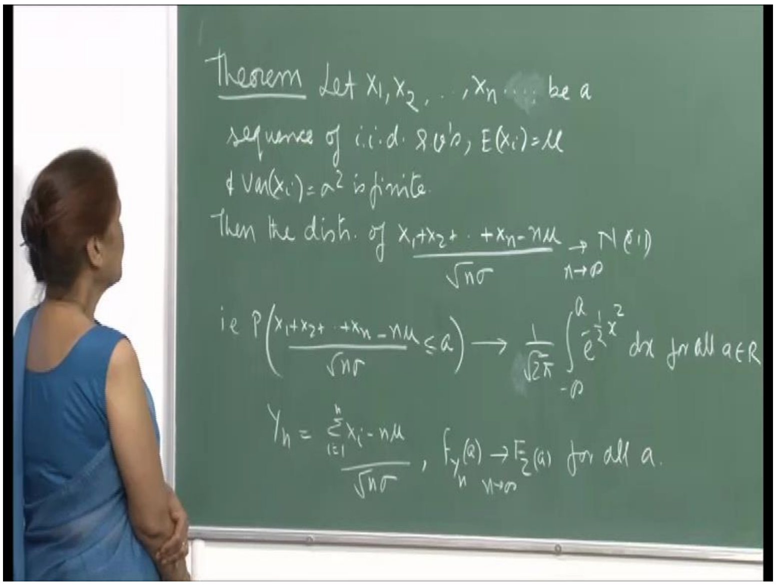
So, we will continue the discussion with the central limit theorem and what we are saying is so… Here I want to address the questions for example, what does the distribution of X n bar look like? This is one question we want to answer; and we will use the central limit theorem to do that. And then the second question would be how fast does X n bar converge to mu? So, now, let us look at the… The central limit theorem states that, sigma X i minus n mu upon under root n sigma will converge to n 0… that means, normal – standard normal distribution as n goes to infinity; that means, this variate will… because this is a random variable for all n. So, this will converge to the standard normal variate as n goes to infinity. Now, here because expected value of sigma X i – i varying from 1 to n will be n mu; and variance of sigma X i; i varying from 1 to n will be n sigma square; the X i's are sequence of independently identically distributed random variables. So, this is… And therefore, you are standardizing by subtracting the mean of this variate. So, minus n mu divided by the standard deviation, which is root n sigma. Therefore, this we are saying that, after standardizing the variate sigma X i, i varying from 1 to n, central limit theorem says that, this will go to n 0 1. So, in this distribution.

And, the weak law of large numbers said that, in probability, sigma X i, that is, sigma X i by n will converge to mu in probability. But, what we are going to say here show… This is to answer the first question, that is, if you now divide by n, then this becomes sigma X i; i varying from 1 to n divided by n. And, there will be an n here and there is a root n. So, that becomes root n times divided by sigma. So, this whole thing. And, we are saying that, this was… Therefore, now, this is… And therefore, the central limit theorem says that, this converges to this variate, will converge to the normal 0 1. So, I can write down sigma upon root n here. And so, essentially, what we are saying is that, X n will converge; that means the distribution of X n bar as limiting distribution of X n bar will be…

So, right now, the distribution of X n bar for large n we are saying will be close to mu normal – mean mu and sigma and variance sigma square by n. And then of course, as n goes to infinity, we are saying that… So, in other words that, the central limit theorem says that, if you take any distribution, the X 1, X 2, X n were coming from any distribution; but, then when you talk of X n bar and for large enough n, then the curve will become bell-shaped; it will get closer and closer to the normal curve for large n. And, the limiting value – this will converge to variate, which has the normal – standard normal distribution. And so, CLT – the central limit theorem implies the weak law of large numbers, because weak law of large numbers only said in probability X n bar will converge to mu. The probability of mod X n bar minus mu will converge to 0. And so…

But, here it is saying that, in distribution. So, X n bar in distribution will converge to standard normal… I should not say, because if I am taking X n bar; if I am simply taking X n bar, then this will converge to n mu of… So, I have simply said it here for X n bar; I have not talked of the limiting value. What we are saying is that, this will be approximated by normal mu comma sigma square by n. So, the proper statement is that, X n bar – the distribution of X n bar for large enough n will look like a normal mu comma sigma square by n. But, you can see that, as n goes to infinity, this thing will become… So, the whole mass will get concentrated on mu only for X n bar. But, then if you look at X n bar minus mu, this absolute value. Then, we are saying that, the… Or, if you are looking at X n bar minus mu upon sigma by root n; then this will converge to… so that this can be approximated by standard normal. But, when you look at X n bar, then this will be approximately normal mu comma sigma square by n.

(Refer Slide Time: 57:37)



So, the final theorem we can now state as… So, if you have X 1, X 2, X n and so on – sequence of identically independently distributed random variables; each X i having mean mu and variance sigma square, and this variance is finite. So, if the variance is finite; that means the variance exists; then the means will exist. So, we do not have to separately say that, mu is also finite and variance is also finite. It is enough if you say that, the variance is finite. Then, it implies that, the mean also exists. Then, the distribution of – see this is important – of X 1 plus X 2 plus X n minus n mu upon root n sigma. This converges to the standard normal distribution – 0 1 as n goes to infinity. This is what… that is, in other words, we want to say the same thing is that, the probability that X 1 plus X 2 plus X n minus n mu upon root n sigma is less than or equal to a.

This will converge to form – there 1 upon root 2 pi integral minus infinity to a e raise to minus 1 by 2 x square dx for all a belonging to R, because this is the cumulative distribution function for… So, this is what you are saying is this is probability Z less than or equal to a; which I have written down here; that is, if you define the random variable Y n as sigma i varying from 1 to n of X i minus n mu upon root n sigma; then the cumulative distribution function of Y n as n goes to infinity will converge to the cumulative distribution function of the standard normal variate Z, and this is for all a.

And, this is what remember; earlier I had defined convergence in distribution or in law, which said that, the cumulative distribution function of sequence of random variables converges to a particular cumulative distribution function; then we say that, this sequence of random variables converges to that particular random variable in law or in distribution. And so, here this is what we are saying that, the sequence of random variables Y n as n goes to 1, 2, 3 up to infinity; then this sequence of random variables converges to standard normal variate in law. So, now, we had looked at the central limit theorem in various forms; its implications. And of course, we will continue looking at its applications more and more.