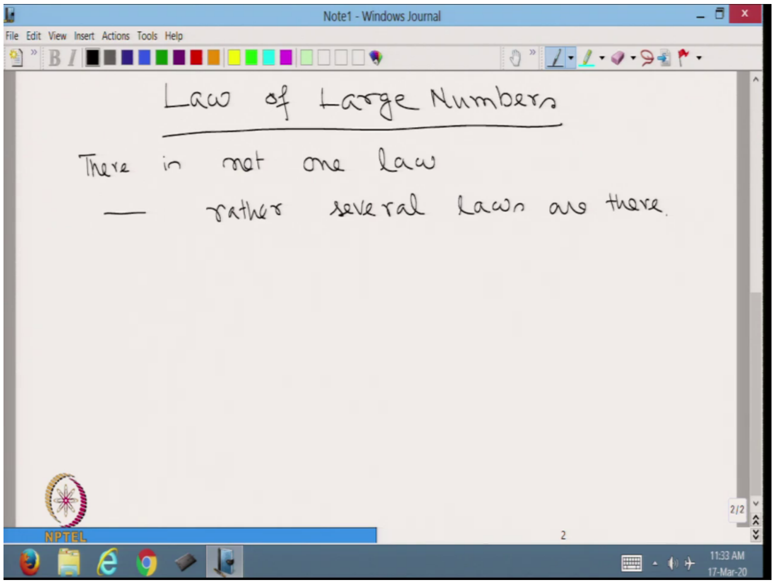
Advanced Probability Theory

Professor Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

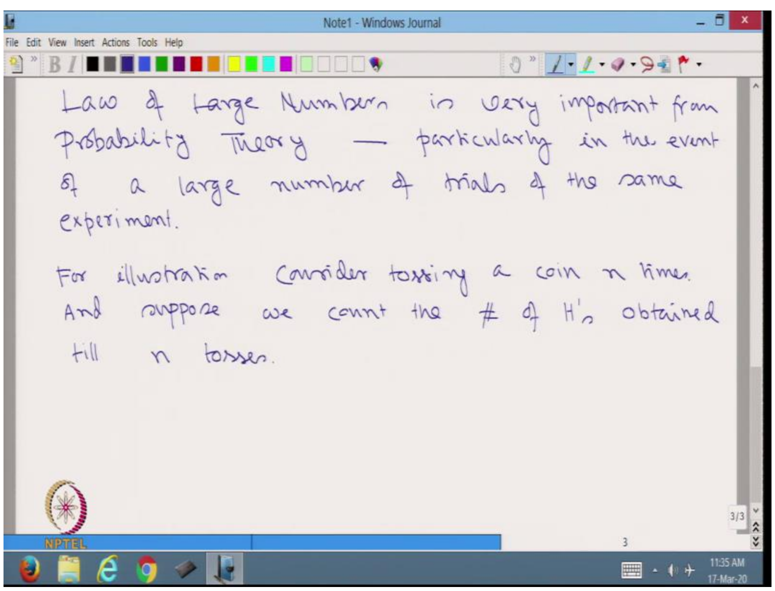
Lecture 27

(Refer Slide Time: 00:33)



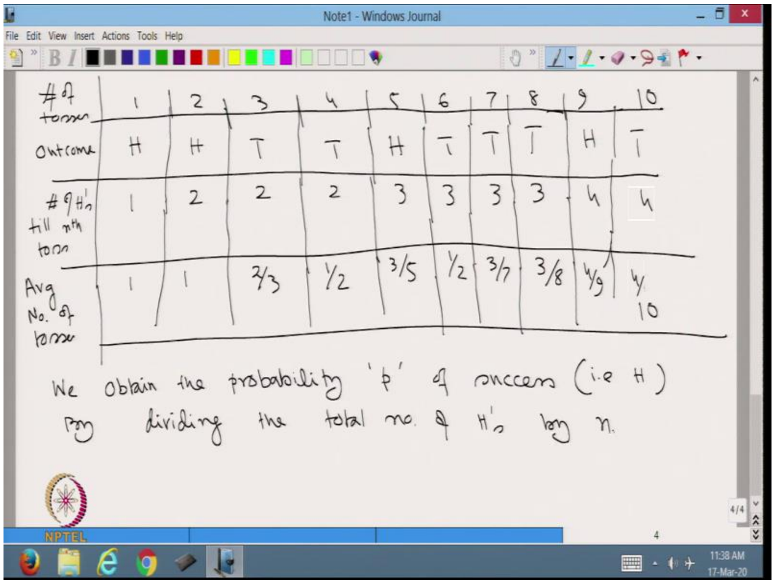
Welcome students to the mock lecture series on the Advance Probability Theory, this is lecture number 27, as I said at the end of the last class that today we will start Law of Large Numbers. In fact there is not one law rather several laws are there we shall look at some of them in detail.

(Refer Slide Time: 01:16)



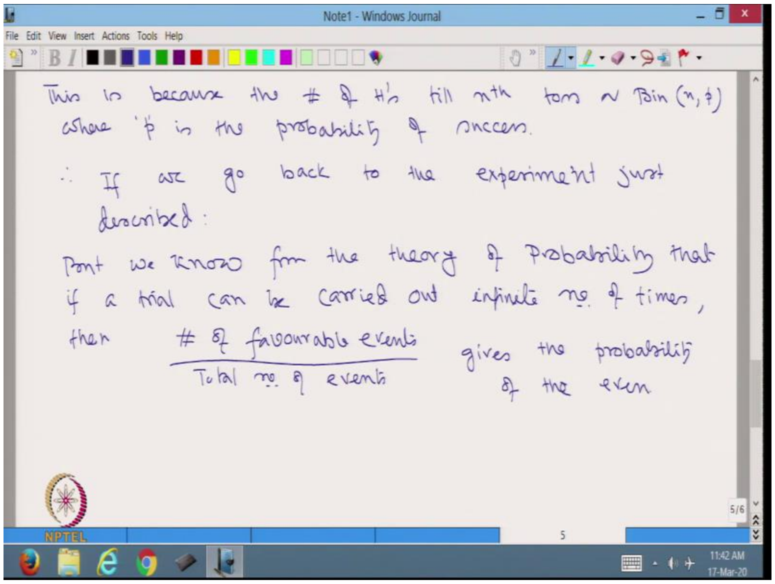
Now, Law of Large Numbers is very important from probability, probability theory particularly in the event of a large number of trials of the same experiment. For illustration consider tossing a coin n times and suppose we count the number of heads obtained till n tosses.

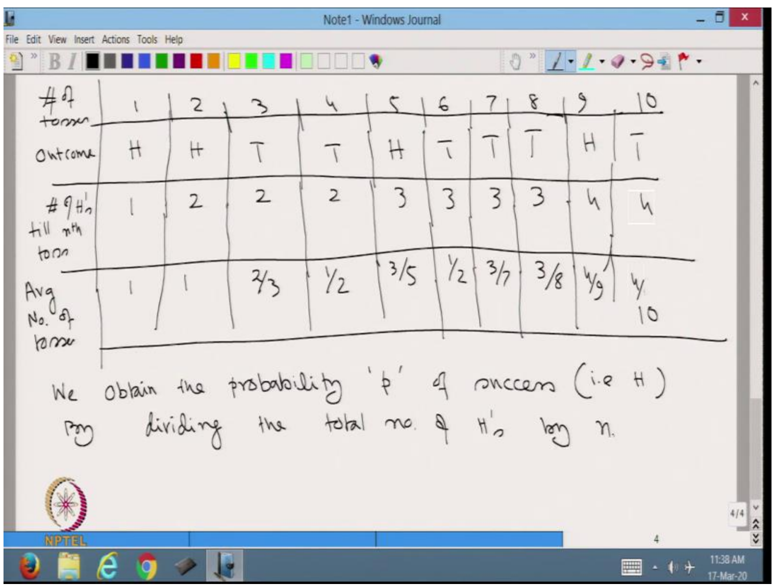
(Refer Slide Time: 03:03)



Say for example we have number of tosses say 1, 2, 3, 4, 5, 6, 7, 8 and suppose the outcomes are like this, suppose the outcomes are head, head, tail, tail, head, tail, tail, let us go 2 more steps head, tail. Then the number of heads if we count till nth toss going to be 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, so if we consider average number tosses or then this is going to be 1, 1, 2 by 3, half, 3 by 5, half, 3 by 7, 3 by 8, 4 by 9 and 4 by 10, we know that we obtain the probability p of success that is head by dividing the total number of heads by n.

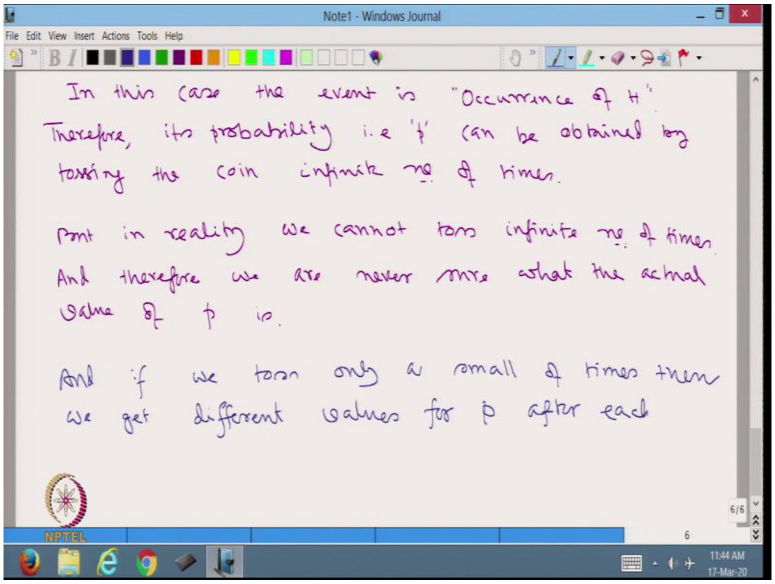
(Refer Slide Time: 05:49)





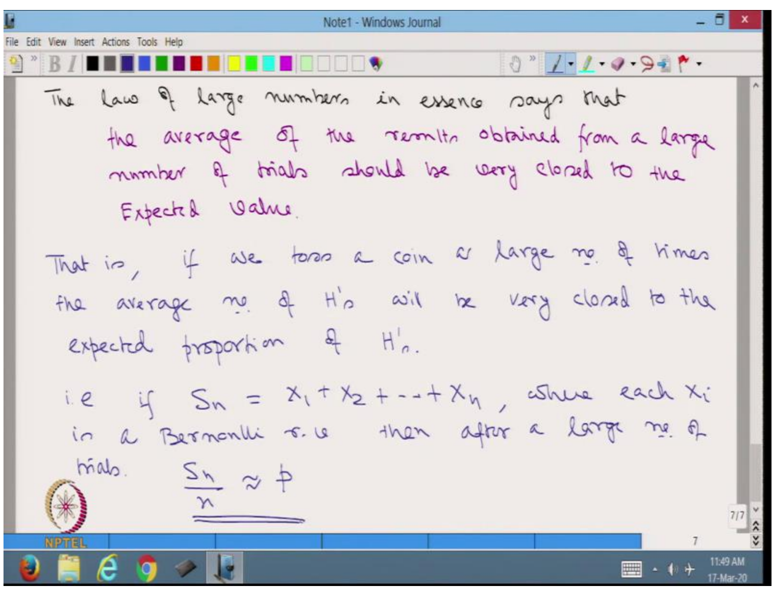
This is because the number of heads till nth toss follow binomially n, p where p is the probability of success. Therefore if we go back to the experiment just describe, therefore if we go back to the experiment just describe, we see that the number of or the proportion of heads is changing with the number of process. But we know that from the theory of probability that if a trail can be carried out infinitely number of times, then the number of favorable events divided by total number of events gives the probability of the event.

(Refer Slide Time: 08:13)



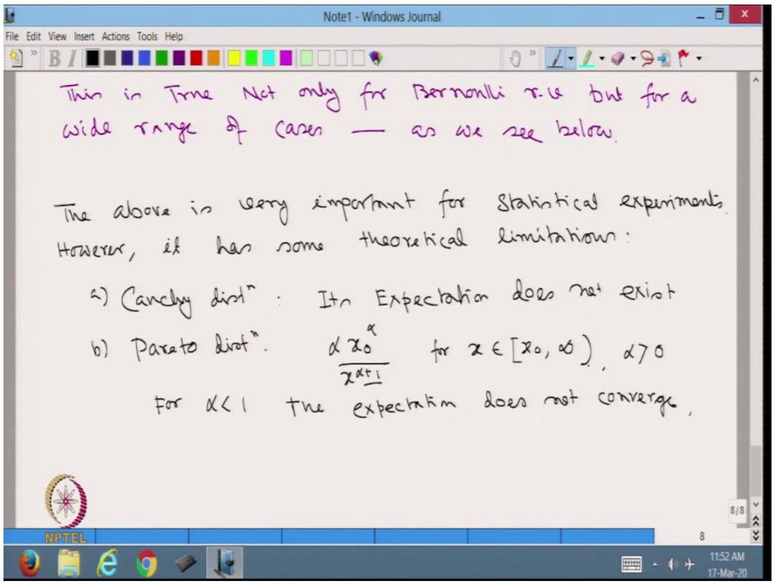
In this case the event is occurrence of head, therefore its probability that is p can be obtain by tossing the coin infinite number of times. But in reality we cannot toss infinite number of times and therefore we are never sure what the actual value of p is and if we toss only a small number of times then we get different values for p after each toss.

(Refer Slide Time: 10:24)



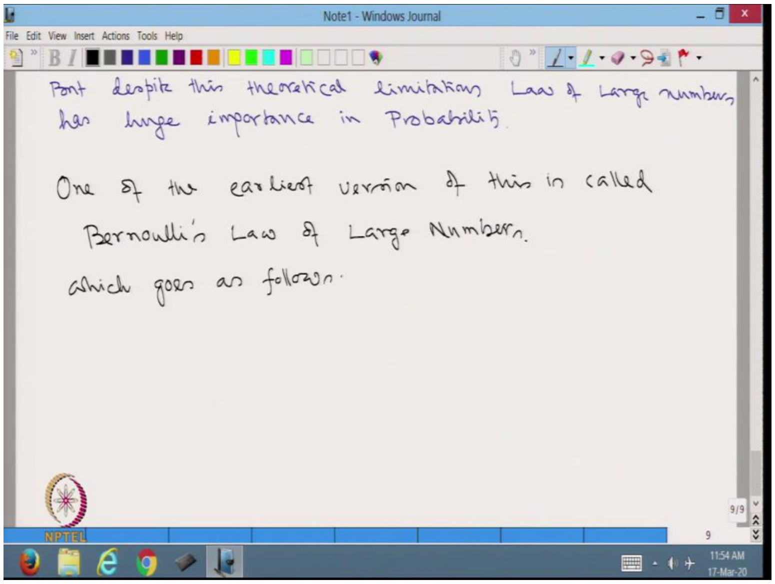
The law of large number says that in essence says that the average of the results obtained from a large number of trials should be very closed to the expected value or in essence that is, if we toss a coin in large number of times, if we toss a coin in large number of times the average number of heads will be very closed to the expected proportion of heads, that is if Sn is equal to X1 plus X2 plus Xn, where each Xi is a Bernoulli random variable than after a large number of trails Sn by n the average number of heads will be very close to the average proportion of heads will be very close to the actual proportion of heads that is p.

(Refer Slide Time: 13:33)



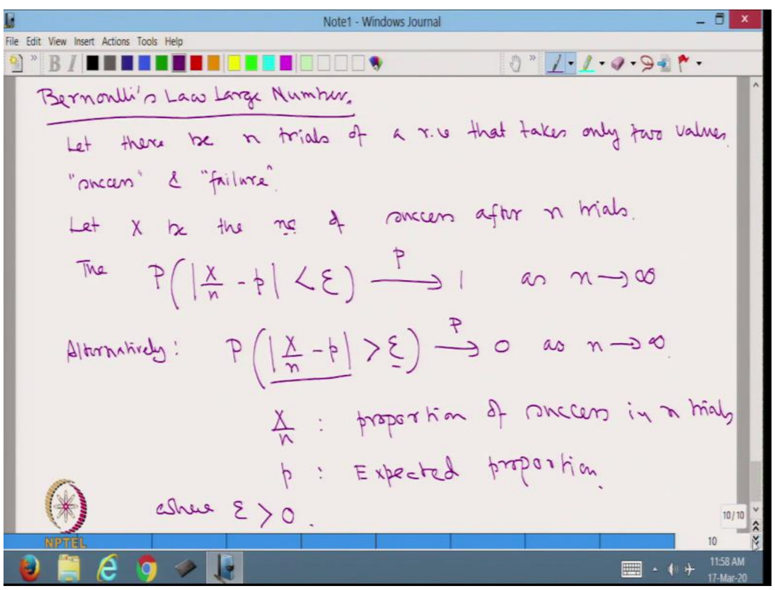
This is true not only for Bernoulli random variable but for a wide range of cases as we see below. Now, the above is very important for statistical experiments, however it has some theoretical limitations such as, Cauchy distribution, we know that its expectation does not exist, similarly Pareto distribution which is of the form alpha x naught to the alpha upon x to the power alpha plus 1 for x belonging to x naught comma infinity and alpha is greater than 0, for alpha less than 1 the expectation does not converge. Because this one will be canceled when we multiplied with x, therefore these series is not converge.

(Refer Slide Time: 16:08)



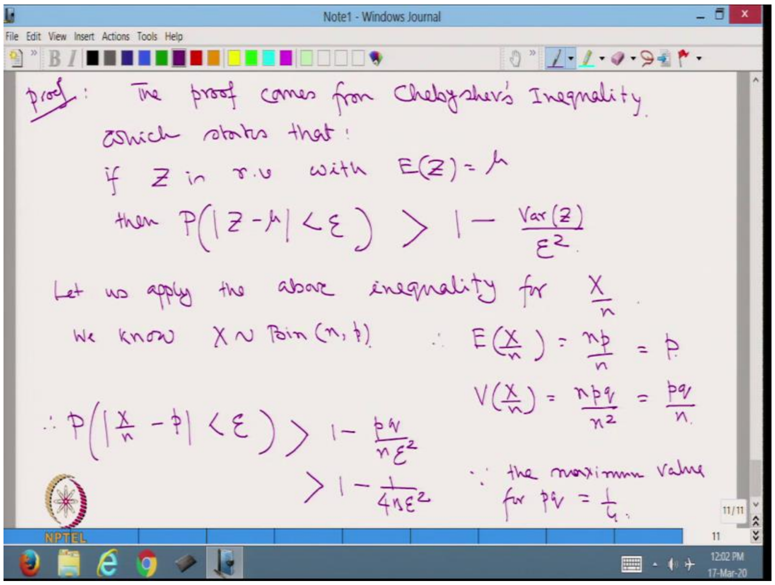
But despite this theoretical limitations law of large numbers has huge importance in probability. One of the earliest version of this is called Bernoulli's law of large number, which goes as follows.

(Refer Slide Time: 17:31)



So, Bernoulli's law of large number, which states that, let there be n trails of a random variables that takes only two values success and failure, let X be the number of success after n trials, then probability modulus of X n minus p less then epsilon goes in probability to 1 as n goes to infinity, or alternately probability modulus of X by n minus p greater then epsilon goes to 0 as n goes to infinity. So, let us analysis the statement, x by n is the proportion of success in n trials and p is the expected proportion, so Bernoulli's law suggest that the difference between them being greater than some epsilon will probabilistically will go to 0, where epsilon ay positive quantity greater than 0.

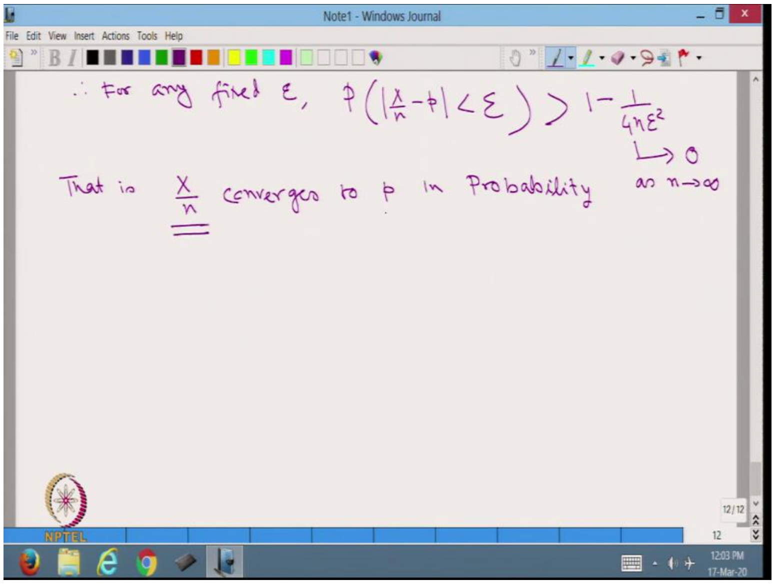
(Refer Slide Time: 20:00)



Proof, the proof comes from Chebyshev's Inequality which if you remember which I hope you remember, but let us still recollect that if which states that, if Z is a random variable with expected value of Z is equal to mu then probability modulus of Z minus mu less then epsilon is greater than 1 minus variance of Z upon epsilon square, this we have proved earlier.

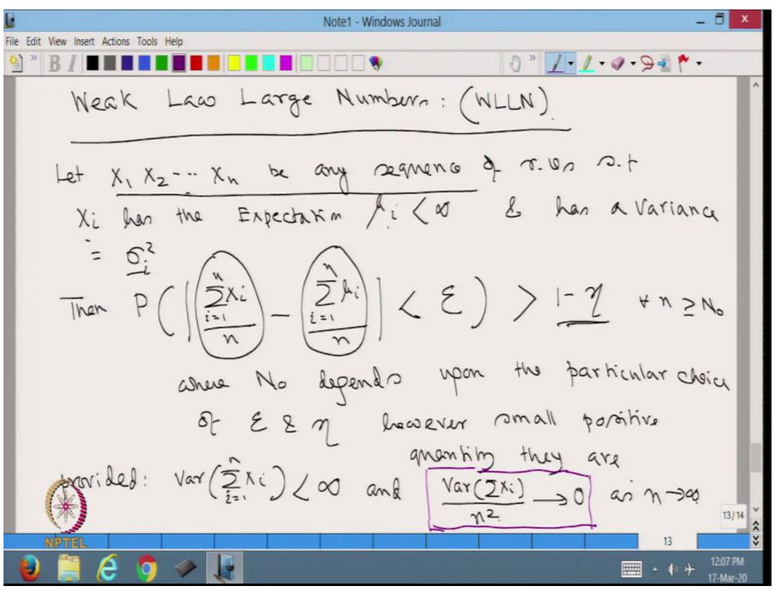
Now, let us apply the above result the above inequality for X by n, we know X is distributed as binomial n comma p, therefore expected value of X by n is equal to np by n is equal to p. And variance of X by n is equal to npq upon n square is equal to pq upon n. Therefore probability modulus of X by n minus p less then epsilon is greater than 1 minus pq by n epsilon square, which is greater than 1 minus 1 by 4 n epsilon square, since the maximum value for pq is equal to 1 by 4.

(Refer Slide Time: 22:45)



Therefore for any fixed epsilon probability modulus of X by n minus p less then epsilon is greater than 1 minus 1 upon 4n epsilon square and this part going to 0 as n goes to infinity, that is X by n converges to p in probability. So, that is the proof that, that is the proof of Bernoulli's theorem.

(Refer Slide Time: 23:44)

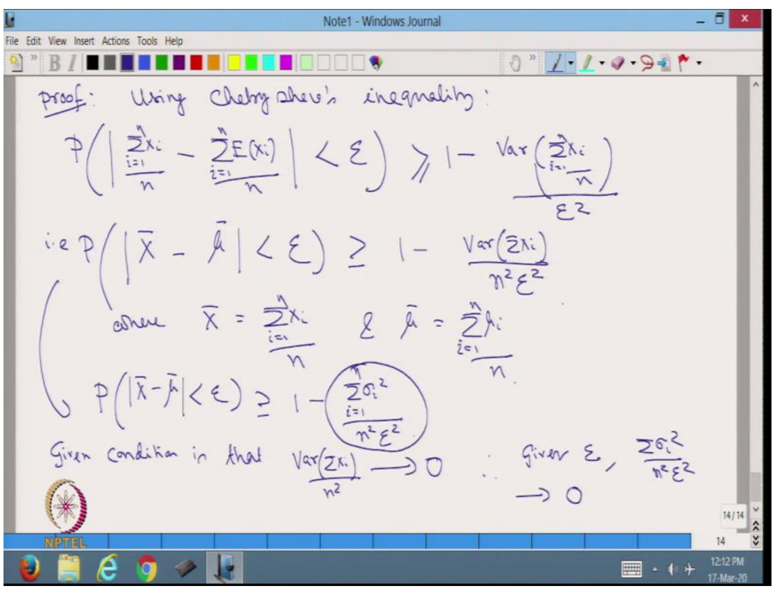


A more generalized version is weak law of large numbers which we shall write as W double L N and it is stated as follows, let X1 X2 Xn be any sequence of random variables such that Xi has the expectation mu i less than infinity and variance is equal to sigma square i, then probability that modulus of sigma Xi i is equal to 1 to n upon n minus sigma mu i, i is equal 1 to n upon n less than epsilon is greater than 1 minus eta for all n greater than equal to N naught, where N naught depends upon the particular choice of epsilon and eta however small positive quantity they are.

So, you understand that this is a much stronger statement, we are looking at a sequence of random variables we are not talking about whether they are independent, we are not talking about whether they are identically distributed only thing that if each one of them has a finite mean and a finite variance then the average of values of the random variable minus the average of the their expectations will be very close that means less then epsilon with a very high probability.

Provided variance of sigma Xi i is equal to 1 to n is finite and variance of sigma Xi upon n square goes to 0 as n goes to infinity. So, this is a very important assumption for weak law of large numbers as we will see some examples later that this is very very important.

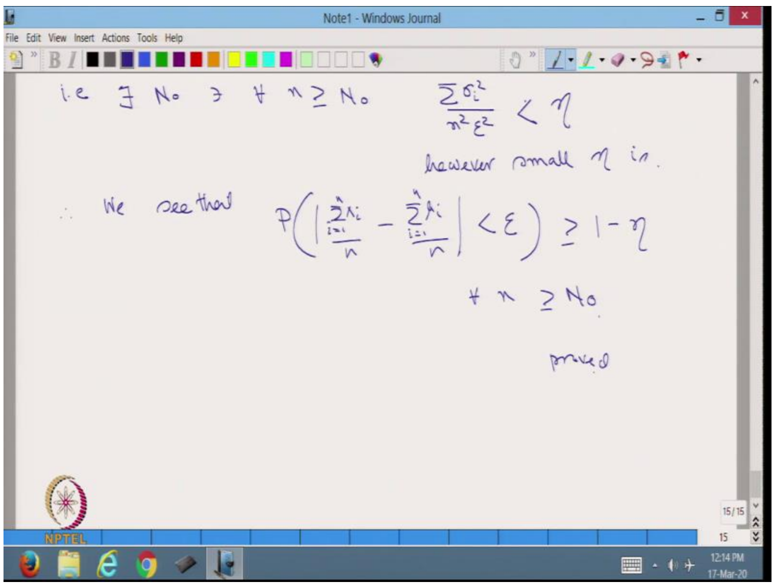
(Refer Slide Time: 27:21)



Proof again using Chebychev's inequality probability modulus of sigma Xi i is equal to 1 to n divided by n minus sigma expected value of Xi i is equal to 1 to n divided by n this less than epsilon is going to be greater than 1 minus variance of maybe greater then equal to 1 minus variance of sigma Xi i is equal to 1 to n divided by n divided by epsilon square.

That is probability modulus of X bar minus mu bar less than epsilon is greater than equal to 1 minus variance of sigma Xi upon n square epsilon square, where X bar is equal to sigma Xi i is equal to 1 to n divided by n and mu bar is equal to sigma mu i i is equal to 1 to n divided by n. Therefore we get probability modulus of X bar minus mu bar less than epsilon is greater than equal to 1 minus summation sigma i square i is equal to 1 to n upon n square epsilon square. Now, the given condition is that variance of sigma xi upon n square goes to 0, therefore for a given epsilon this quantity goes to 0.

(Refer Slide Time: 30:09)



Note that:

If X_i's are independent then

$$Var(\Sigma X_i) = \sum_{i=1}^{n} \sigma_i^2$$

If X_i's are Not independent

then $$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \sigma_i^2 + \sum_{\substack{i,j \\ i \neq j}}\sum cov(X_i, X_j)$$

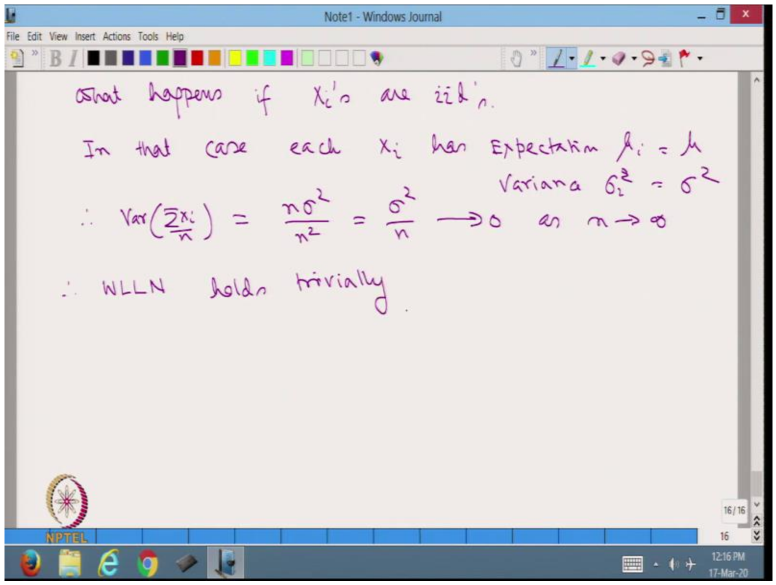
But because of the assumption

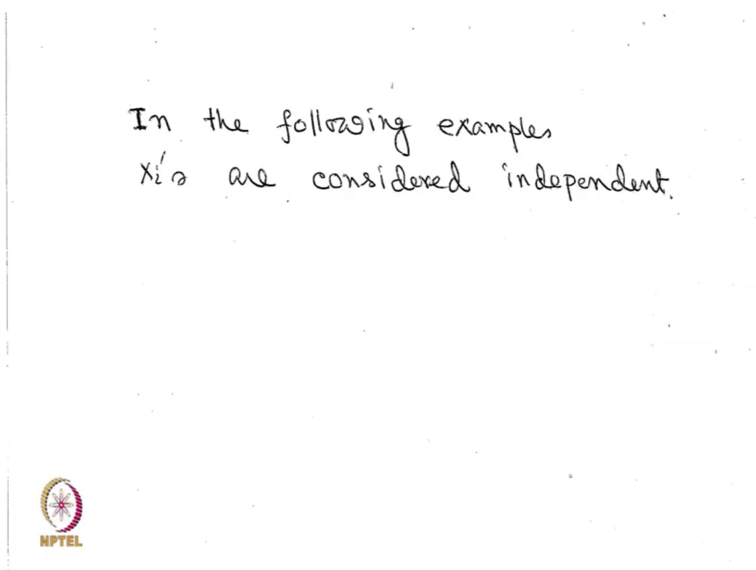
$$\frac{Var(\Sigma X_i)}{n^2} \longrightarrow 0$$

WLLN holds.

That is there exist N naught such that for all n greater then equal to N naught, summation of sigma i square upon a square epsilon square is less then eta, however small eta is therefore we see that probability modulus of sigma Xi i is equal to 1 to n upon n minus sigma mu i i is equal to 1 to n upon, that probability less then epsilon is greater than equal to 1 minus eta for all n greater then equal to N naught, hence proved.

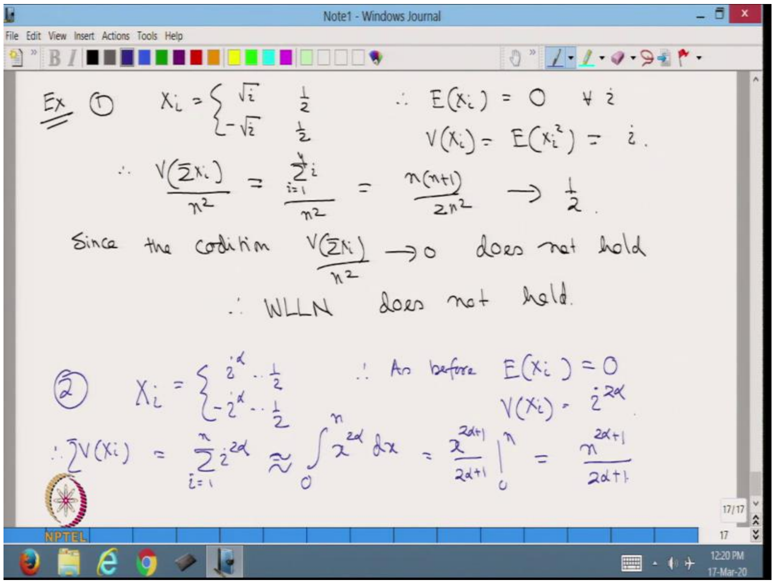
(Refer Slide Time: 31:22)





What happens if Xi’s are iid’s. In that case each Xi has expectation mu i is equal to mu and variance sigma i square is equal to sigma square. Therefore variance of sigma Xi upon n is equal to n sigma square upon n square is equal to sigma square by n goes to 0 as n goes to infinity. Therefore weak law of large number holds trivially. So, let me give you some examples.

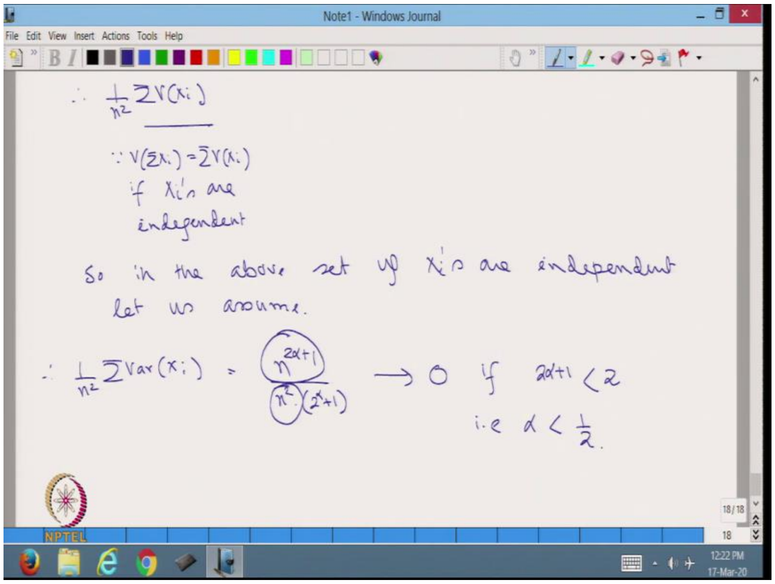
(Refer Slide Time: 32:53)



Suppose Xi is distributed as root over i with probability half and minus of root over i with probability half, therefore expected value of Xi is equal to 0 for all i and variance of Xi is equal to expectation of Xi square is equal to i, therefore variance of sigma Xi upon n square is equal to sigma over i i is equal to 1 to n upon n square is equal to n into n plus 1 by 2 n square which converges to half, since the condition variance of sigma Xi upon n square goes to 0 does not hold, therefore weak law of larger number does not hold.

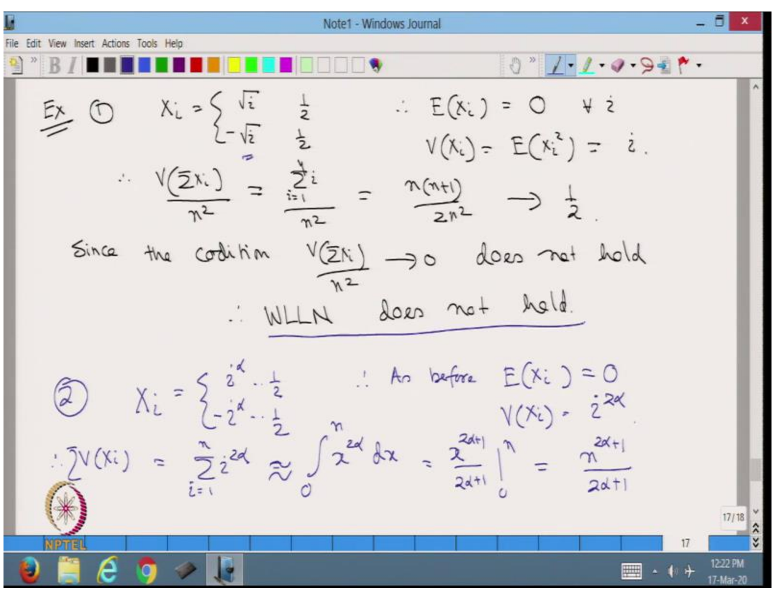
Example 2, Xi is equal to i to the power alpha with probability half and minus i to the power alpha with probability half. Therefore as before expected value of Xi is equal to 0 and variance of Xi is equal to i the power 2 alpha, therefore sigma variance of Xi is equal to sigma i to the power 2 alpha i is equal to 1 to n is equal to integration 0 to n x to the power 2 alpha dx which is is equal to x to the power 2 alpha plus 1 upon 2 alpha plus 1from 0 to n is equal to n to the power 2 alpha plus 1 upon 2 alpha plus 1.

(Refer Slide Time: 36:07)



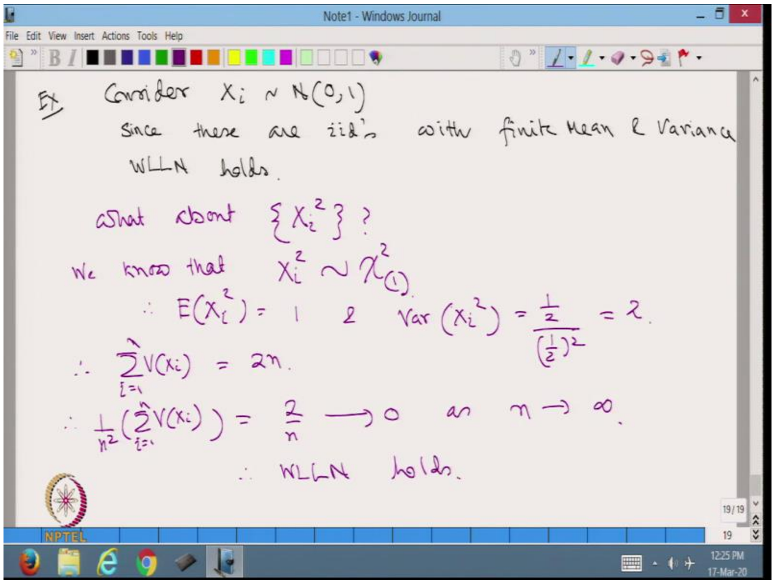
Therefore 1 by n square into sigma variance of Xi I am writing is sigma variance of Xi since variance of sigma Xi is equal to sigma variance of Xi if Xi's are independent and so in the above set up Xi's are independent let us assume. Therefore 1 by n square into sigma variance of Xi is equal to n to the power 2 alpha plus 1 upon n square into 2 to the power alpha plus 1, n to the power 2 alpha plus 1 upon n square 2 alpha plus 1 which will go to 0 if 2 alpha plus 1 is less than 2 that means this quantity has to be dominated by this quantity that is alpha is less than half.

(Refer Slide Time: 37:44)



Hence no wonder that alpha is equal to half weak law of large number does not hold.

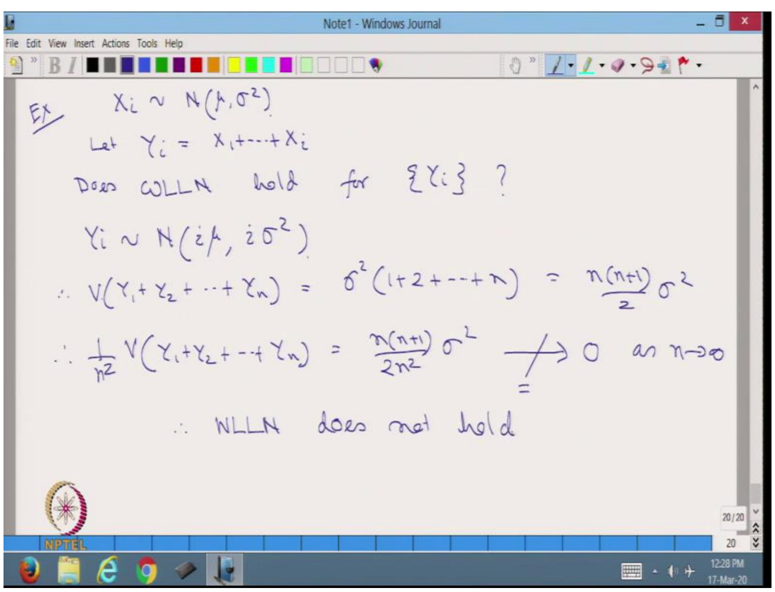
(Refer Slide Time: 37:56)



Another example, consider Xi following normal 0, 1. Since these are iid's with finite mean and variance weak law of large number holds, what about the sequence sigma Xi? We know that Xi square is distributed as Chi square with 1 degree of freedom, therefore expected value of Xi square is equal to 1 and variance of Xi square is equal to lambda upon alpha square is equal to 2.

Therefore sigma variance of Xi i is equal to 1 to n is equal to 2n.Therefore 1 by n square sigma variance of Xi i is equal to 1 to n is equal to 2 by n which goes to 0 as n geos to infinity, therefore weak law of large number holds.

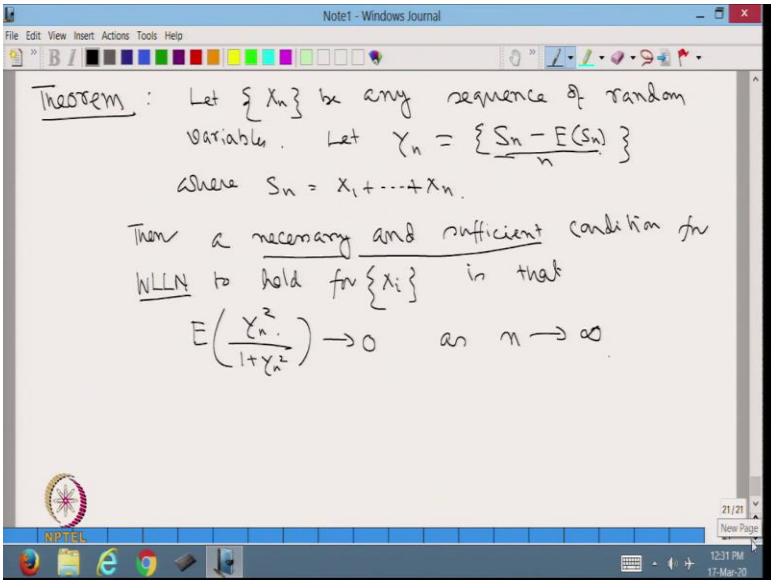
(Refer Slide Time: 39:55)



Another example, suppose Xi is equal to normal with mean mu variance sigma square let Yi is equal to X1 plus X2 up to Xi, does weak law of large number hold for the sequence Yi? That is the question. Now, each Yi, Yi is distributed as normal with i mu and variance is equal to i sigma square, therefore variance of Y1 plus Y2 plus Yn is equal to sigma square into 1 plus 2 up to n is equal to n into n plus 1 by 2 sigma square.

Therefore 1 by n square into variance of Y1 plus Y2 plus up to Yn is equal to n into n plus 1 by 2 n square into sigma square, which does not go to 0 as n goes to infinity. So, this is the notation to imply that it is not converging to 0. Therefore weak law of large number does not hold.

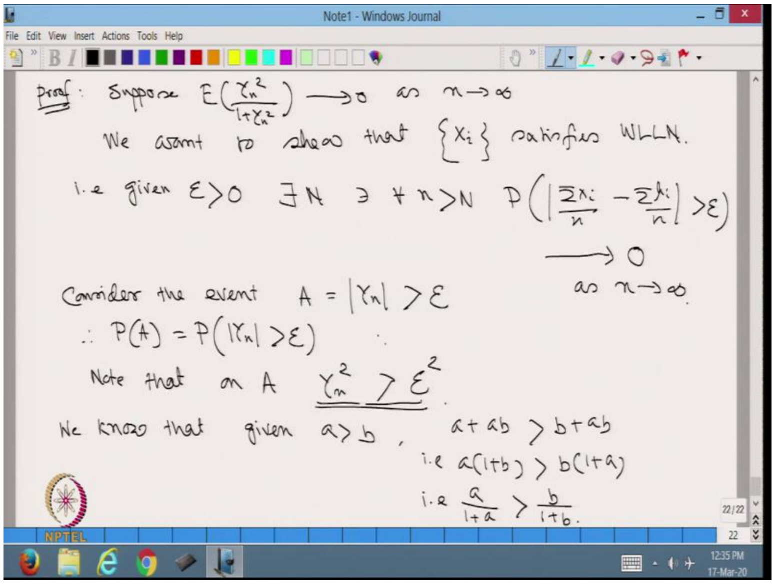
(Refer Slide Time: 41:56)



Let me now state a very important theorem, which states as follows, let Xn be a sequence, let Xn be any sequence of random variables, let Yn be Sn minus expected value of Sn divided by n, where Sn is equal to X1 plus X2 up to Xn, then a necessary and sufficient condition for weak law of large numbers to hold for Xi is that expected value of Yn square upon 1 plus Yn square goes to 0 as n goes to infinity.

So, this is a very important result and it is a necessary and sufficient condition for weak law of large numbers to hold. Now, let us analysis the result, Yn is equal to the partial sum X1 up to Xn minus its expectation divided by n, therefore yields expectation is going to 0 means expected value of Yn square is becoming very very small, so that is the implication in such a case Xn is equal to obey weak law large numbers.

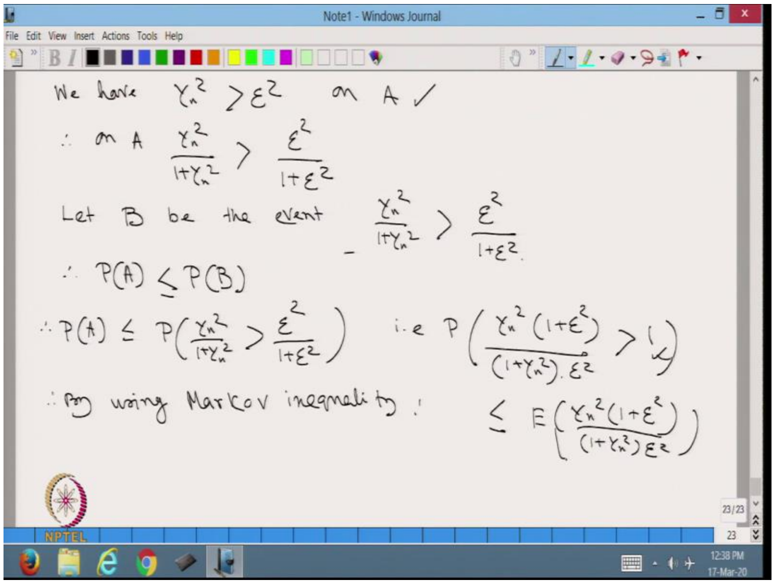
(Refer Slide Time: 44:20)



Proof, suppose expected value of Yn square upon 1 plus Yn square goes to 0 as n goes to infinity we want to show that Xi satisfies weak law of large numbers that is given epsilon greater than 0 they are exist in such that for all n greater than N probability modulus of sigma Xi upon n minus sigma mu i upon n greater than epsilon that probability will go to 0 as n goes to infinity.

Consider the event A is equal to modulus of Yn is greater than epsilon, therefore probability of A is equal to probability modulus of Yn greater than epsilon. Note that on A Yn square is greater than epsilon square. Now, we know that given a greater than b, a plus ab is greater than b plus ab, that is a into 1 plus b is greater than b into 1 plus a that is a upon 1 plus a is greater than b upon 1 plus b, we want to apply the same result here.

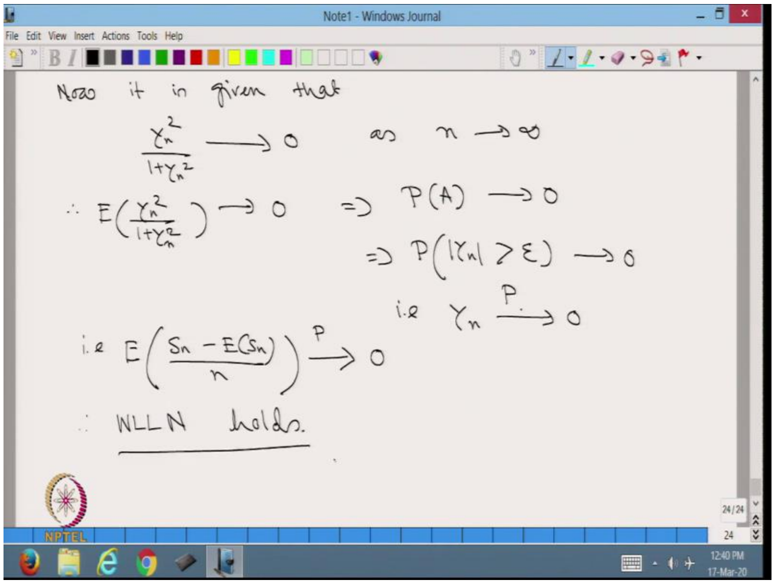
(Refer Slide Time: 46:54)



So, we have Yn square is greater than epsilon square on A therefore on A Yn square upon 1 plus Yn square is greater than epsilon square upon 1 plus epsilon square, let B be the event Yn square upon 1 plus Yn square is greater than epsilon square upon 1 plus epsilon square here probability of A is less than equal to probability of B, because this is contained in this event.

Therefore probability of A is less than or equal to probability Yn square upon 1 plus Yn square greater than epsilon square upon 1 plus epsilon square that is probability Yn square into 1 plus epsilon square upon 1 plus Yn square into epsilon square is greater than 1.Therefore by using Markov inequality which we have done several classes back we can write it as this probability is less than equal to expected value of yn square into 1 plus epsilon square upon 1 plus Yn square into epsilon square because the threshold there is 1.

(Refer Slide Time: 49:13)



Now, it is given that Yn square upon 1 plus Yn square going to 0 as n goes to infinity, therefore expected value of Yn square upon 1 plus Yn square goes to 0 imply probability of A is going to 0 implies probability modulus of Yn greater than epsilon is going to 0 that is Yn converges to 0 in probability that is expected value of Sn minus expected value of Sn divided by n converges in probability is 0.

Therefore weak law of large number holds. okay friends I stop here today we have just proved the theorem in one direction we shall look at the other side that given that Xn holds weak law of large numbers we shall show that expected value of Yn square upon 1 plus Y square goes to 0, also we shall do some more examples and some theorem in the next class. Thank you.