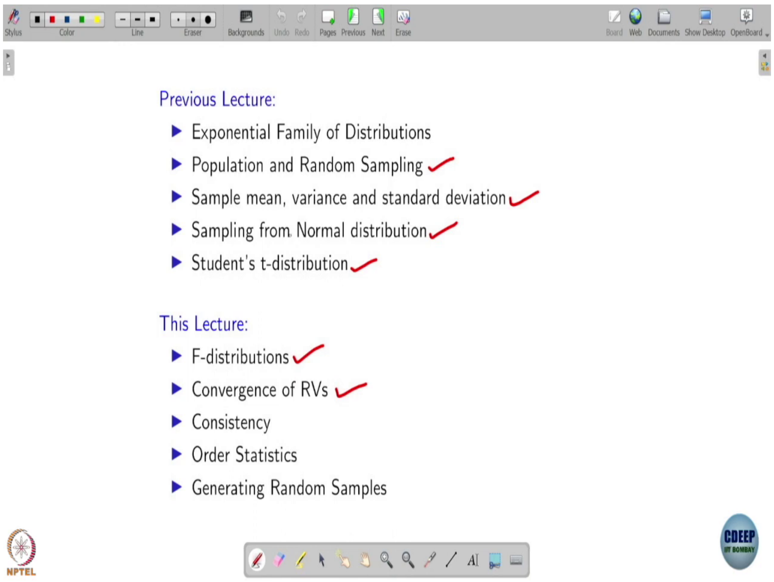
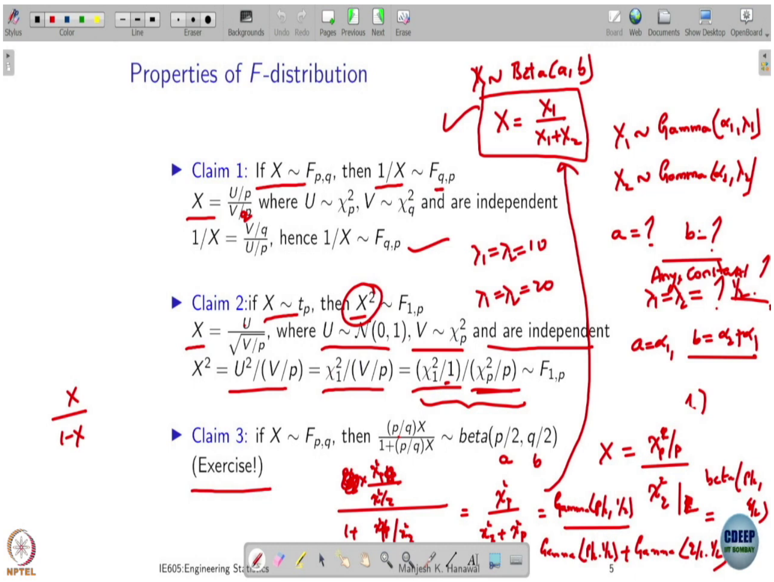
Engineering Statistics
Professor Manjesh Hanawal
Department of Industrial Engineering and Operations Research
Indian Institute of Technology, Bombay
Lecture 25
Convergence of Random variables and Consistency

(Refer Slide Time: 00:14)



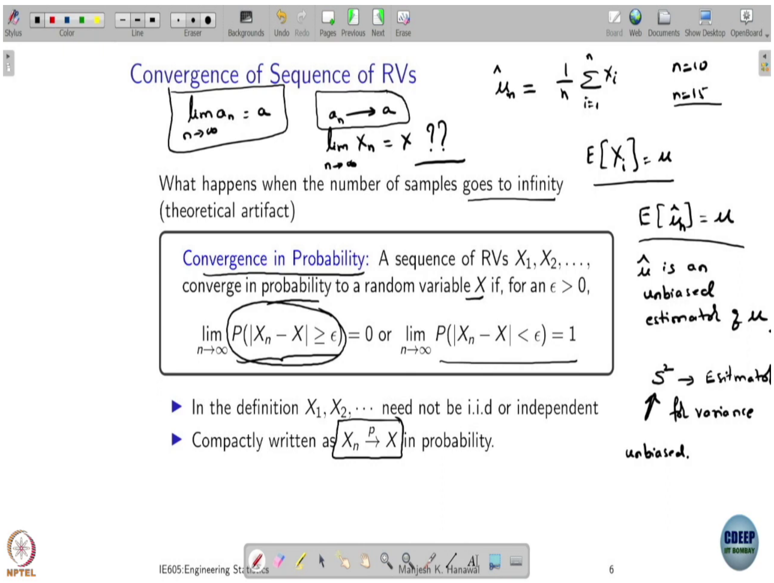
So, let us get started. So, in the previous lecture and previous to that lecture, we started talking about exponential family of distributions then talked about random sampling, we talked about a sample mean, sample variance, sample deviations, we talked about the properties and then started looking into sampling from normal distributions. So, that led us to two special distributions called as student t-distribution and student F-distributions.

(Refer Slide Time: 00:57)



And we also discussed about some relation between various distributions, like how I can obtain like F-distributions, reciprocal, what is the relation between T-distribution F-distribution and this mapping of F-distribution to beta distribution. So, you also saw the last one in quiz.

(Refer Slide Time: 01:22)



We have been talking about basically statistics which is about we said simply is some function of the bunch of random variables we have observed and as a special case we observed sample mean, sample variance and sample standard deviation. That time we discussed about one notion called unbiased like sample mean we called it as a estimator for the mean value of the distribution, sample variance is an estimator for the variance of the distribution.

So, at that time we talked about something called unbiased. So, if you have this, like we had this, if you have i samples we had by x and if all of these xi's are such that their mean value is mu then we said if you average them what we said is we are going to get an estimator for this mean value which we called as mu hat. And this mu hat was a random quantity because it depends on the random variables.

Then we say that if expectation of mu hat is equals to mu, this is, this estimator is mu hat is an unbiased estimator of mu. And similarly, we also saw S square, this was an estimator for variance and we said that this is also unbiased. Now, we will look into another property of this estimator called consistency. But before we define that notion of consistency, we need some more definition. Because that consistency notion is an asymptotic notion.

You notice that here, this unbiasedness you can define for any n, you just take it for n equals to 10 or n equals to 15 whatever. If you take n equals to 10 you are taking average of those 10 sample and its mean should be close to mu and it happens to be mu if this xi samples are i.i.d. And if you take n equals to 15 and you take, maybe I should make it n, subscript n, just to indicate that it depends on n samples and here I should take mu n.

So, here irrespective of what is the n samples you are going to use expectation of mu n hat is going to be mu. And this is, if it is an unbiased estimator, this is true for every n. But now I am going to introduce another notion called consistency which holds for n tending to infinity. So, for that we are going to look into convergence of these random variables as we have infinitely many of them. Our first notion of convergence is called convergence in probability.

Did you hear about convergence in probability in IE621? No, okay. We are going to say that a sequence of random variable X1, X2 convergence in probability to a random variable. If you look into the difference of, absolute difference of random variable Xn with X that being larger than epsilon that goes to 0 as n goes to infinity. So, just a contrast this what we are trying to say with a standard limit.

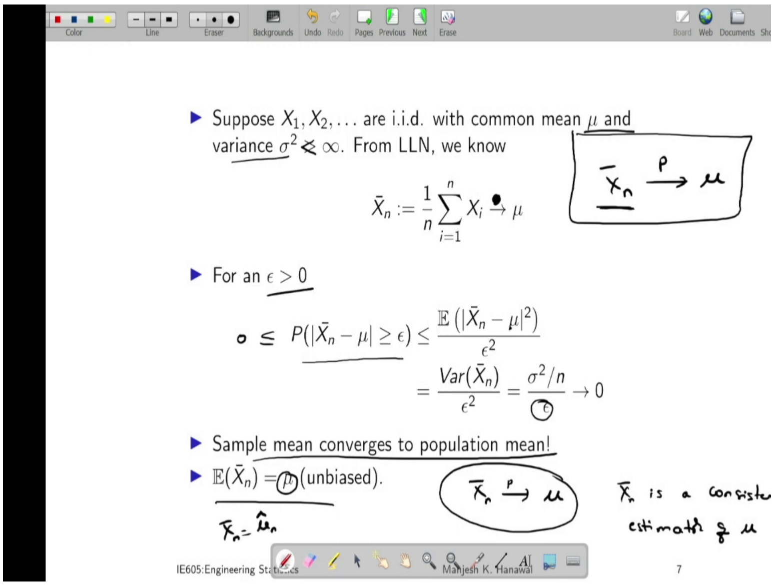
So, what does in the standard limits what we talk about? Sequence of random variables, sorry, sequence of numbers tending to a. What does this mean? Or we simply write a goes to a, an goes to a. And you know the definition of this limit, it is a standard definition. But now, what I am facing with is sequence of random numbers. What does this mean?

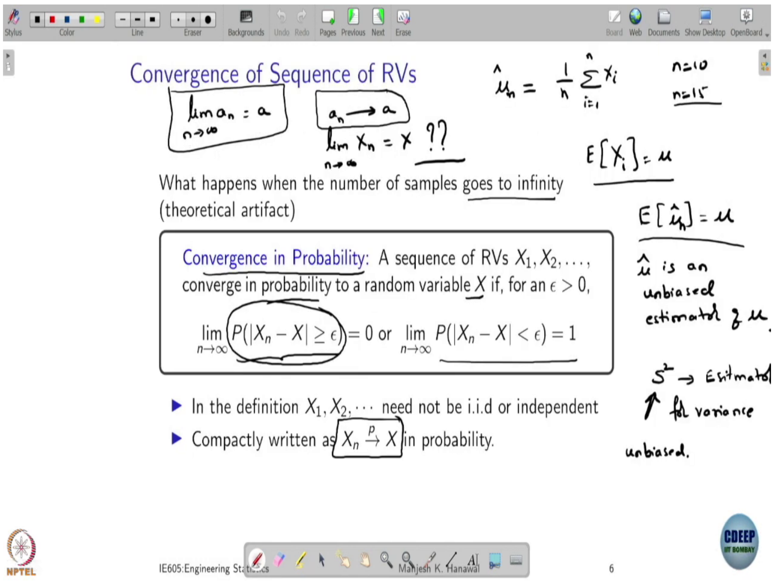
So, the sequence of random numbers converging to something we need to appropriately define some notions. And the first notion is what we call it as convergence in probability. And I just defined what is that definition is. If you look into… Now, once you take this probability this is some number and now that sequence converges into 0, we are talking about that sequence converging to 0.

And the complement of definition is if I am going to look into less than or equal to epsilon, this limit goes to 1. This is just a complement of this. And at this point I did not say that this X1, X2 there has to be i.i.d. Now, neither they need to be in identically distributed or independent, this is just a definition. If the sequence of random variables such that this holds, then we are going to call the convergent probability.

And the shorthand notation like the way we use it for a deterministic case like this, in the stochastic case, use this notion that xn goes to x with a superscript p written on the arrow mark. So, I will not just go into that I am just introducing this definition just to tell you, what is consistency.

(Refer Slide Time: 08:34)





Now, let us say you people have already encountered sequence of random numbers earlier. Where did you encounter a sequence of random variables?

Student: CLT.

Professor: When you dealt with law of large numbers and central limit theorems. So, now let us see. Now suppose, let us assume that my sequence are i.i.d. with a common mean mu and variance sigma, and there is a typo here this should have been less than infinity, it cannot be greater than infinity. And now, law of large numbers said that if you take the average of these samples this value went to mu, maybe I will not write p here at this point. This is what law of large number told me.

If you take the average of this number, and if you continue to take average of these numbers as n tends to infinity, that value is going to mu. Now, let us see, is there any connection between this statement and the statement and the definition of convergence in probability. So, the claim is that this simply says that xn converges to mu in probability. Now, let us see how is that. So, I am now going to go and apply this definition here.

Choose any epsilon positive. Now, I am interested in knowing this chi. Now, these are my sequence of random variables xn bar minus mu I will be interested in. The difference being greater than epsilon. I know this has to be greater than or equal to epsilon 0 that is by definition. Now, can somebody tell me how did I get this upper bound?

Students: Chebyshev inequality.

Students: Markov inequality.

Professor: Markov inequality. So, I obtained a Markov inequality.

Students: Chebyshev inequality.

Professor: Markov or Chebyshev?

Students: Chebyshev inequality.

Students: Markov inequality.

Professor: This is already in a way that because of this difference we have written, it is fine you can say any of this. But actually, what I have done is basically I have take the square on both sides. If I do like this, what you are saying? If I take a square on both sides?

Student: Markov.

Professor: Now, this will Markov. Now, expectation of square of this absolute value is nothing but variance n by epsilon. Now, when I actually got this, this is actually I could have directly written this and called it a Chebyshev inequality. But I first… We also did our Chebyshev inequality using Markov inequality only. And this is exactly the way we derived Chebyshev inequality from but I am just repeating those steps.

Now, what is this quantity? Variance of Xn bar. We computed this, we computed when we calculated the sample means variance when you can computed, we calculated we computed a sigma square by n.

Student: (())(12:25).

Professor: Now, I am interested in n going to infinity, when n goes to infinity this quantity vanishes n goes to 0 irrespective what is the epsilon you choose. So, now notice that this is lower bounded by 0 and now also upper bounded by 0 as n goes to infinity. So, this value is converging to what, this probability is converging to?

Student: 0.

Professor: 0. By definition that means your Xn bar is converging to?

Student: Mu.

Professor: Mu? So, that is why this statement. Law of large numbers if you interpret in other way, it is just or the formal way of interpreting law of large numbers says the sample mean converges to that population mean in?

Student: Probability.

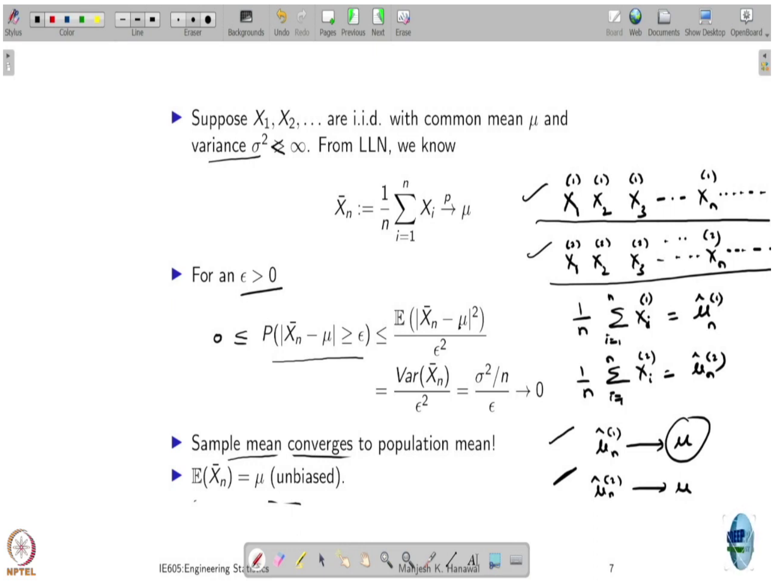
Professor: Probability. So, when we say it law of large number, we did not mention the notion of convergence, we just said if you take the average it will go to some number, but that time we did not formally define what is the limiting convergence mean? But now, we are just formalizing that saying that convergence was convergence in probability. So, sample mean converges to the population mean. And we already know this. Sample mean is unbiased for any n.

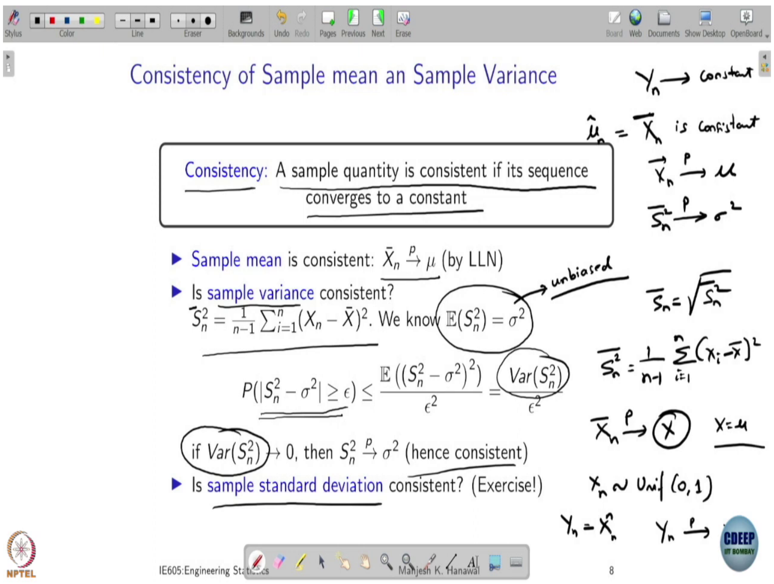
Now, I think I missed to write it here. Now, whenever it so happens that if expectation of Xn equals to mu and Xn bar also converges to the same parameter that I am interested in here. That expectation of sample mean equals to this mu, and it converges to the same value mu, then I am going to call Xn bar is consistent, is a consistent estimator of, of what?

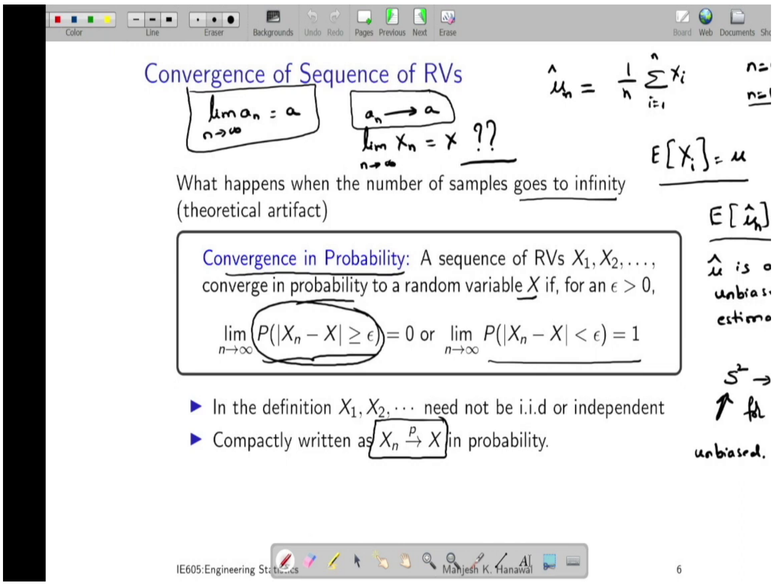
Student: Mean.

Professor: Mean. What is the difference between this and this? Agreed it is a limiting property. But is there anything more than that? Yes, this guy is that is asymptotic as n goes to infinity, but is there anything more, you see the expectation here and no expectation here. So, what is this quantity, this is Xn bar, we denoted it as a mu, mu n bar, mu n hat, this is basically average of n samples, and we say this is still a random quantity. So, if you give me one sample, one bunch of n samples, I will get some value. If you give me another bunch of n samples, I will get this. So, let us say I will do this.

(Refer Slide Time: 15:59)







Let us say one of you give me one sample one set of samples, x1, x2, x3 like this xn. And I am going to call the test set 1 and to denote that I am superscript in them with 1. Another one of you give me another set of samples. And I am going to superscript with 2, and each one of you can give me n samples but let us stick to. If I have to go and take the samples from x first guy this is like i equals to 1 to n 1 by n.

And he is whatever the value he gets me let us call this 1. And similarly, whatever the second guy gives me, the average of the second guy's give me let us call that. And all these, this is and these all samples are coming from the same underlying population. Now is mu 1 hat 1 and mu 1 hat 2 are going to be the same?

Student: Not necessary.

Professor: Not necessary, but are the same in expectation.

Student: Yes.

Professor: Yes, they are same in expectation because we know that sample means is an unbiased estimator. But suppose you give me lot of samples like you keep on give me many samples. And both of you give me a lot of samples. And now, this mu 1 hat you are giving me more and more samples I am going to average. And this is where it is going to converge?

Student: Mu.

Professor: Mu. And where is this mu and hat is going to converge?

Student: Mu.

Professor: Even though they need not be equal, but when you let n goes to infinity both the samples they are going to converge to the same mu. So, that is why like if one of you just give me n tending to infinity sample, I will just average and I know this is already mu I do not care about another guy now, as I already have enough samples. So, this is what, this is why it is called consistent whether I use this this as long as n tending to infinity, it is consistently giving me the same estimated value, same value of mu.

So, this is what we are going to say, consistently. As sample quantity is consistent with sequence converges to a constant. And in case of sample mean, we know that that constant is simply the population mean. And that has come from law of large numbers. Law of large numbers say to what constant it converges to. Now, let us say we now just to talk about sample mean, as estimator for your population mean.

Now, let us look into sample variance and see whether this is a consistent estimator for my population variance. How you are going to do that? This is my definition of sample variance when you have n samples. And we know that this, we actually discuss this, we know that sample variance is unbiased. Now again, I want to check whether if I take Sn bar and see that how much it is away from the sigma square, I want to compute this probability.

So, I know how to compute this probability. Or at least get an upper bound of that. This one is nothing but by again applying your Markov inequality you will get this value. Now, what is this quantity? This is nothing but the variance of your sample variance. So, Sn square your sample variance, its variance. So, whether did we compute variance of sample variance? If not, you can verify that even that value as n goes to infinity goes to 0.

So, variance of sample variance also goes to 0 as n tends to infinity. Now, because of that, by definition Sn square converges to sigma square in probability. And hence, it is consistent. Let us revisit this definition of consistency. So, a sample quantity is consistent if its sequence converges to a constant, whatever it is, let us say I am just going to say Yn, if it going to converge to constant, then you are just saying definition of consistencies or sample quantities consistent with sequence converges to a constant.

Let us say Yn is a sequence of random variable, which goes to some constant value, we are calling it as consistency. Now, what I am now trying to ask is whether Xn bar is consistent. Does Xn bar converges to a constant?

Student: Yes.

Professor: Yes. How?

Student: Law of Large Numbers.

Professor: Law of Large Numbers. Law of large numbers already told us that Xn bar converges to the mu value. That is why it is consistent and that convergence is actually in convergence in probability. And similarly, we just said that if you are interested in Sn bar we just computed this, that it also goes to a constant sigma square in probability. You just computed this, Sn squared and that is why the sample deviation is also consistent.

And similarly, you can compute that sample standard deviation is also going to be consistent. And what is that that sample standard deviation is simply square root of your Sn bar. Any question about this convergence in probability and consistency? You people are in sync with me or lost or anything is not consistent here. That is our notation because we said these are like estimators, sample mean is an estimator.

And if we obtained it by average of n samples, so we said that was our kind of informal we are going to whenever mu is the parameter basically we are trying to estimate so we will write it as mu n bar, mu n hat. And whenever it is an average of certain number of samples, we will put a bar on it, Xn bar is nothing but summation of xi 1 by n. So, that is Y bar.

Student: (())(24:29).

Professor: Sn square? This is same reason. This is also we obtain as an average of n samples. What was Sn square? Given it is n-1.

Student: (()))(24:50).

Professor: Even though there is some kind of centralization here, but it is still average of certain number of. I think when… Maybe this is why you are getting confused. Like sometimes I have put it bar and sometimes I have not put it bar.

Student: Yes, sir, we can use simply Sn square.

Professor: Sn square also you can use, no issues. I just like to be consistent with this Xn bar. I put bar because just to indicate it is an average of certain number. Sn bar is also average a certain number quantities, that is why I put bar on this. So, I think we first be clear with this definition of convergence in probability. Now, one more exercise. I say Xn bar goes to X in probability. What is X has to be?

Student: Random variable.

Professor: Xn bar by our definition is a sample mean that mu has to be population mean it cannot be anything that is a constant. That x is not a random variable. It is a constant. That is why the definition here, that is why we are also calling it consistent because it is converging to a constant. When that constant happens to be sample mean here. Now, one more simple example, I will take.

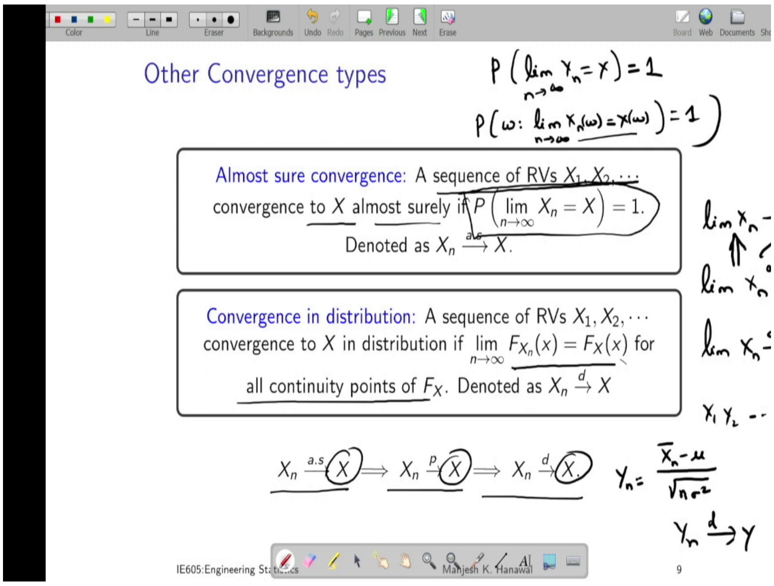
I will give you Xn is uniform in the interval 0, 1. Notice that open interval, it is an open interval and 0, 1 are not included. And now, I define another random variable Xn equals to Xn square. Does Yn converges to anything? If Yn converges to Y in probability what that y has to be?

Student: 1 by 4.

Professor: 1 by 4? Sorry, this should not Xn square this should be X and n. I am just refining it Xn n. Xn, what you understand what is Xn? Xn is a uniform random variable between 0, 1. Now, I am defining Yn to be n-th power of that Xn. See Xn's are all going to be between 0, 1. Everything is between 0, 1 and if I start experimenting with n which is tending to infinity what will happen to that value.

It is going to be 0. Does this why is 0 then, yes or no? Then, this y is 0. This Y is the limiting value is no more a random variable it is a constant. Then the definition of this convergence in probability applies. So, then we should say, yeah, Yn converges to 0 in probability.

(Refer Slide Time: 28:43)



Next. This definition we will quickly go through not spend much time these are just definitions. What are the other notions of convergence in probability? Or what are the other notions of convergence? There are two three other notions of convergence called almost sure convergence and convergence in distribution. You do not need to know much but it is just better to know these definitions. We say that a sequence of random variables converges to X almost surely, if probability that limit n tends to infinity Xn is X with probability 1. So, anyone understand what does this mean?

Student: (())(29:41).

Professor: So, let us try to understand what is this, maybe this is a compact definition, but let us see what is this means, limit n tends to infinity Xn equals to X equals to 1. What does this mean? This basically saying that this is probability of all samples on which this Xn of omega is. So, you take a omega, at that omega that sample point, you compute the Xn values and see whether that limit, this is like a Xn omega.

If you fix omega Xn omega is a deterministic sequence. Agree or no? Once you pick a sample omega, you just compute your Xn on that particular sample value, this is like a standard limit. Whatever the X you have chosen, if this holds, then that omega lies here. And if for some omega this condition does not hold then it does not belong to this set. What it is saying that if collection of all those omegas will be such that their probability is 1 then I am going to say that Xn converges to X almost surely.

So, this is different from almost, sorry, convergence in probability, you can always construct random variables which converges in probability, but not in almost sure, but it is the case that if it converges in almost sure, it always converges in probability. So, if let us say a limit of Xn converges to p and I have a limit of Xn converges to, no X here and X and this is P and almost sure is written like this.

So, this implies this, but this need not imply this. So, which one of them is a stronger notion of convergence? Almost sure is stronger notion or convergence in probability stronger notion?

Students: Almost sure.

Professor: Almost sure. Fine. The other notion is convergence in distribution. Suppose you have a sequence of distributions and it so happens that the CDF converges point wise. You take some point x and compute the CDF of all the random variables at that point x. And if that value converges to the CDF of your limiting random variable x at that point x and this should happen for all continuity points of F of x.

Notice that F of, your CDF can have jumps. Only at those points, it is, where it continuous if this happens, then we are going to call it as convergence in distribution. And we are going to denote it as goes to Xn distribution. I am now show you a little bit to think carefully. Did we come across converges in distribution before?

Student: CLT.

Professor: CLT. So, what did the CLT said?

Student: Normal Distribution.

Professor: Normal Distribution. So, what CLT said is, let us say if your sequence X1, X2 have any, these are i.i.d.'s. And if you look into that Xn bar minus mu and?

Student: n sigma square.

Professor: Let us call this Yn. Now, Yn converges to?

Student: Normal(0, 1).

Professor: In what sense?

Student: In distribution.

Professor: In distribution sense. Now, actually what I can say is instead of this I am can say Yn converges to Y where Y is normal 0, 1, so n converges to Y in distribution. I think this is good enough. So, that is what the relation between them is, if you have almost sure convergence, you know it converges in probability. And further if you know it converges in probability, you can know that it converges in distribution also.

And when all these convergence happening to the same limiting X. So, if Xn converges to X almost surely, then Xn converges to the same x in probability and that Xn converges to the same X in distribution also. We do not need to go much more details than this. I think you will study all this converges and distribution in a much more elaborate way when you take Probability 2, you are doing IE621 now.

There is a second part of that course. I do not know what is the label. So, if you take the next version of that, you will study all of these distributions.