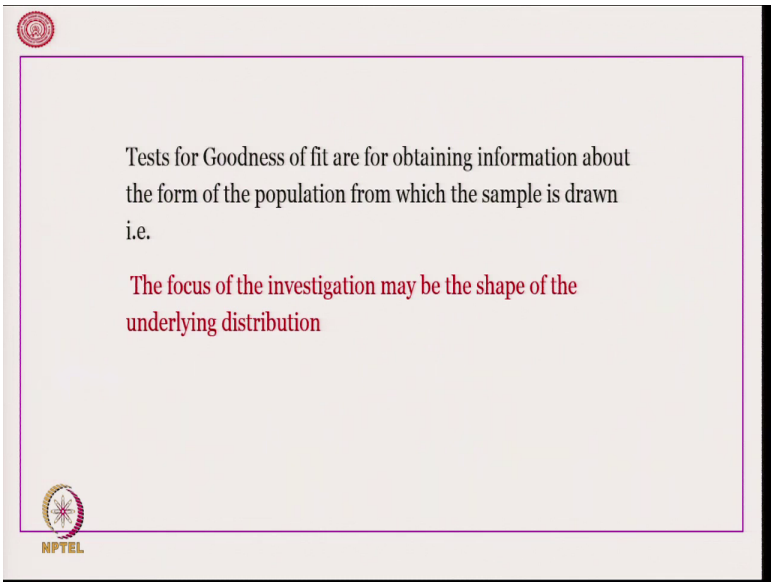
valueNonparametric Statistical Inference
Professor. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi
Lecture No. 06

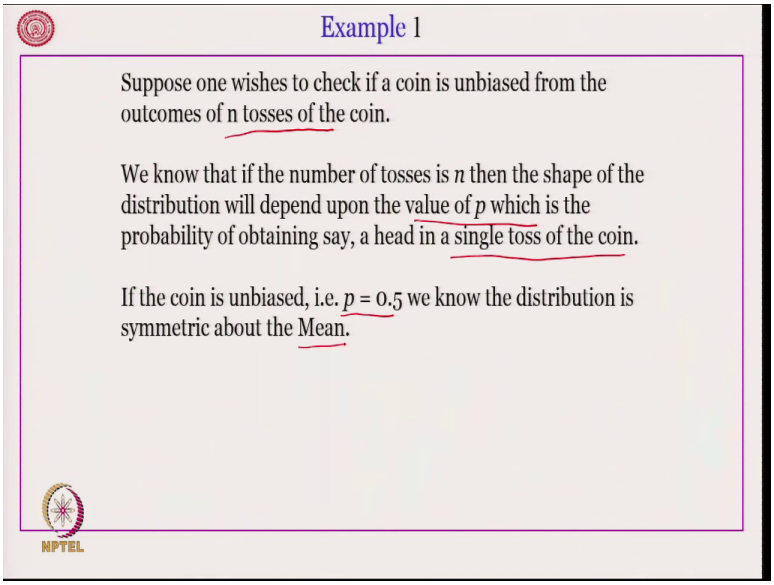
Welcome students to MOOC's series of lectures on Nonparametric Statistical Inference, this is lecture number 6. As I said at the end of last class that in this lecture, we shall study some tests for goodness of fit.

(Refer Slide Time: 0:41)



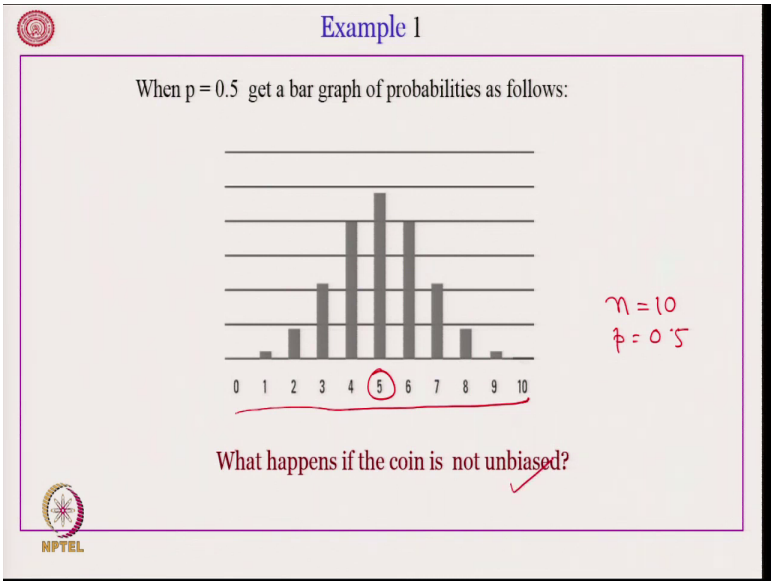
Now, the question is what are test for goodness of fit? Basically this is about obtaining information about the form of the population from which the sample is drawn. And the focus of the investigation may be the shape of the underlying distribution.

(Refer Slide Time: 1:02)



Let me now illustrate with respect to one example. Suppose one wishes to check if a coin is unbiased from the outcomes of n tosses of the coin. We know that if the number of tosses is n, then the shape of the distribution will depend upon the value p; which is the probability of obtaining a head in a single toss of the coin. Or, in other words we call it the probability of success. If the coin is unbiased that is p is equal to 0.5; we know that the distribution is symmetric about the mean.

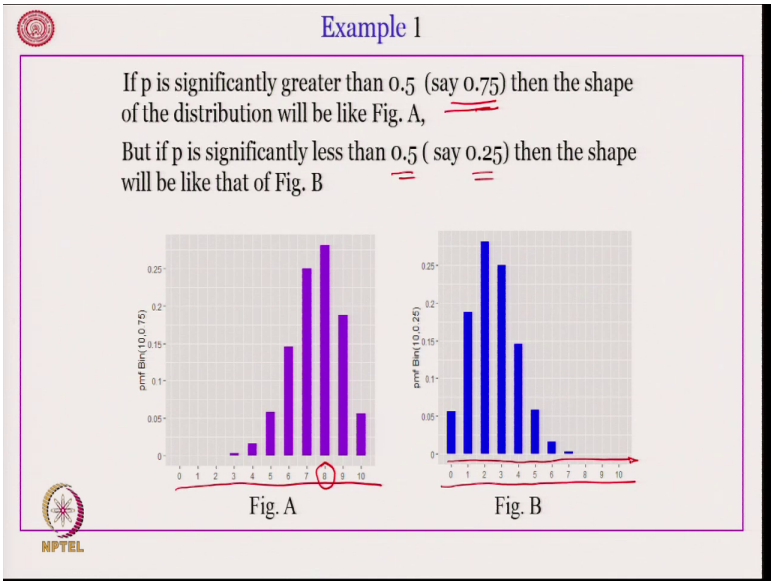
(Refer Slide Time: 1:50)



For illustration let us look at the graph that we have plotted; here n is equal to 10, p is equal to 0.5. And therefore mean

is 5 and we can see that the distribution is symmetrical in the range 0 to 10, for the binomial 10 comma 0.5 variable. The question is what happens if the coin is not unbiased?

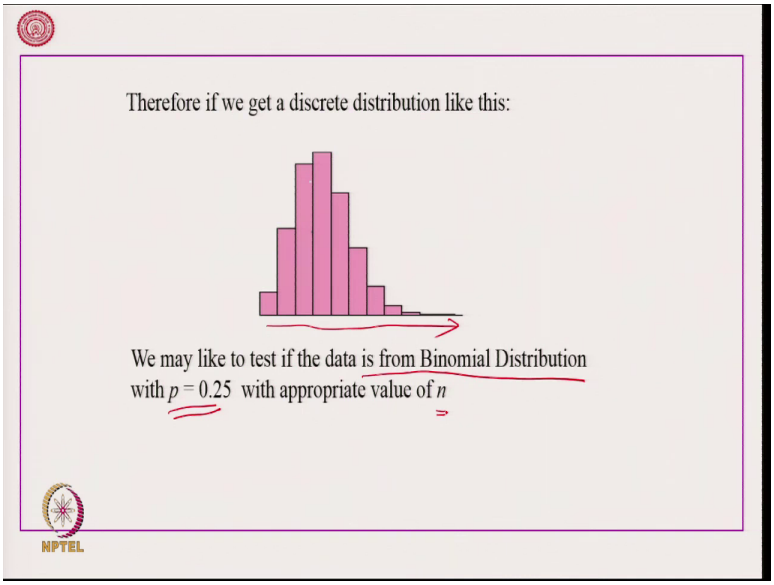
(Refer Slide Time: 2:25)



So, if the coin is not unbiased, P is different from 0.5, it maybe higher, it maybe lower than that. So, let us consider a p which is say 0.75. Then for the same n that is 10, we get a distribution which will look like this; which is given in the figure A. And as you can understand that since the value of p is equal to 0.75, the probability of getting more heads is higher than getting a lesser number of heads. And consequently this is the type of distribution that one can obtain, with the mode value at 8. On the other hand, if the value of p is significantly lower than 0.5; so for this illustration we have considered it to be 0.25.

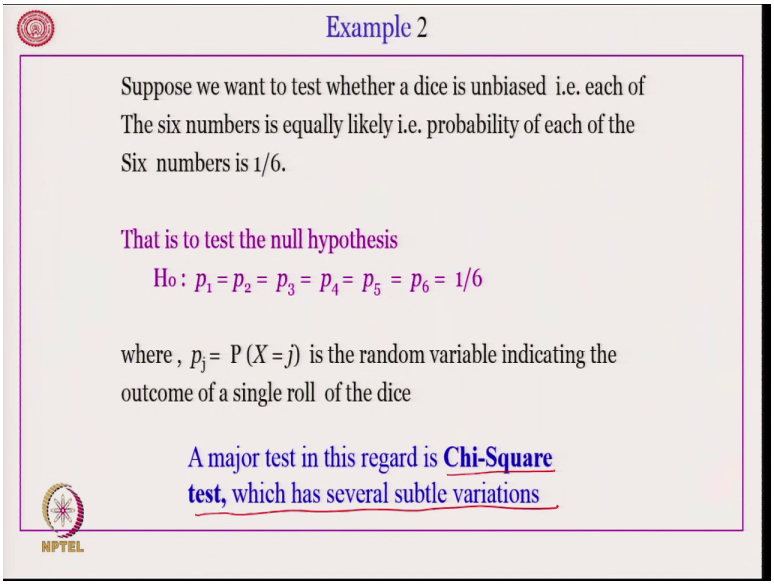
And we can see that the bar chart describing the probabilities look like this, which is not symmetric around the mean at all. And we can see that it has a skewed distribution.

(Refer Slide Time: 3:41)



Therefore, suppose we get a discrete distribution from the observed sample something similar to this; which we can see you skewed on this side. Therefore, it is quite natural that we like to test whether it is coming from a binomial distribution with p is equal to 0.25, for an appropriate value of n. And that is what testing of hypothesis is all about.

(Refer Slide Time: 4:15)

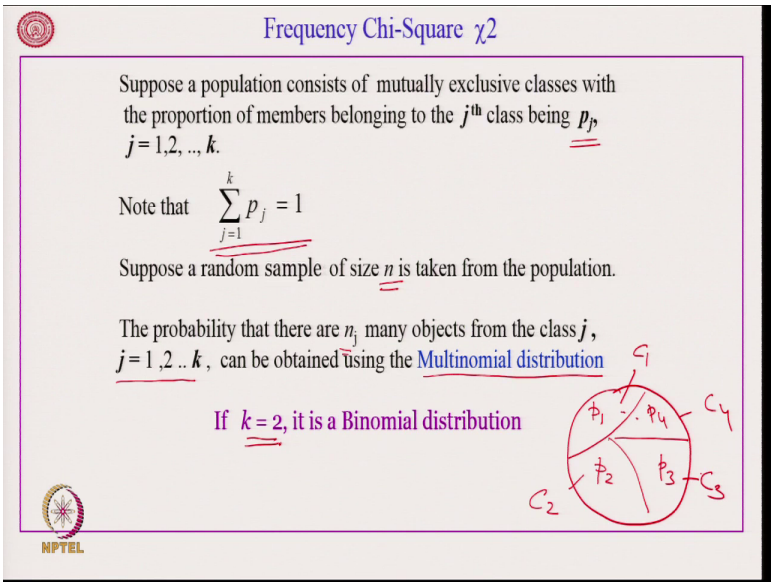


Let me give you another example. Suppose we want to test whether a dice is unbiased, that is each of the 6 faces of the die will have equal probability. Therefore, the hypothesis that we are testing is p1 is equal to p2 is equal to say up to p6; and each one of them is 1 by 6. Where, Pj is

probability that X is taking the value j; where X is the random variable indicating the outcome of a single roll of the dice. Question is how to test that? Let us understand that if the coin is unbiased; it does not mean that in 100 tosses we shall get exactly 50 heads and 50 tails. That may not happen, but there will be deviations from this obtained value from the expected value np.

However, if the deviation is too much, then we would say that it is not acceptable; and or in other words we are going to reject the null hypothesis. So, corresponding test we need to do, the major test in this regard is Chi-Square test, which has several subtle variations; but, we shall study the basic Chi-Square test today. So, the idea of frequency Chi-Square is like this.

(Refer Slide Time: 5:59)



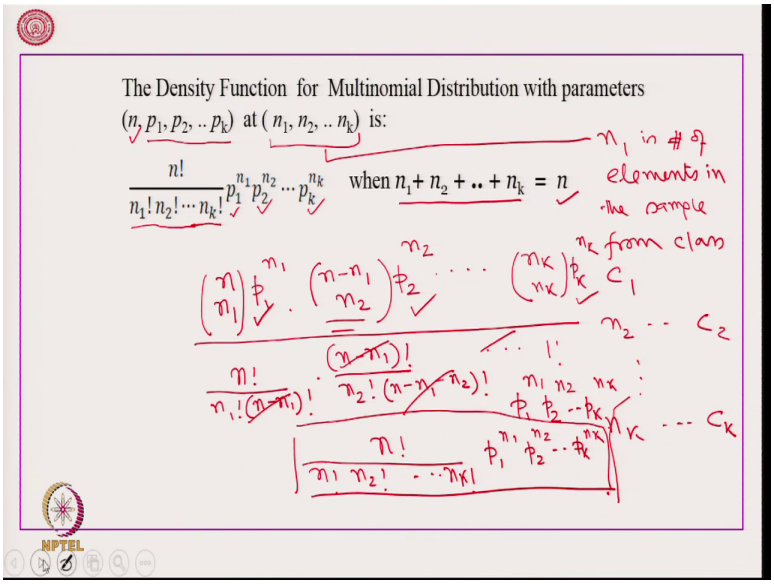
Suppose a population consists of mutually exclusive classes with the proportion of members belonging to the jth class is pj. What does it mean? It means that suppose this is the entire population; it is divided into several classes. Suppose in this case I have taken four classes; so p1 is the proportion of elements belonging to class 1. p2 is the proportion of elements belonging to class 2, like that for class 3 and class 4. And we have to understand that they are mutually exclusive; that means that no element can belong to two of these classes at the same time. Now,

$$\sum_{j=1}^{k} p_j = 1$$

where k is the number of classes.

Now, suppose we have taken n samples from the population; the probability that there are nj many objects from the class j. j is equal to 1 to k, can be obtained using the multinomial distribution. I hope you are familiar with multinomial, but I will explain it now. If k is equal to 2, then we get a binomial distribution.

(Refer Slide Time: 7:33)



So, what is multinomial distribution with parameters p1, p2, pk, and n which is the number of samples. The probability distribution will look like this, n factorial upon n1 factorial, n2 factorial into nk factorial, into p1 to the power n1, p2 to the power n2, pk to the power nk; where n1 n2 nk is that n1 is number of elements in the sample from class c1. Similarly, n2 from c2 and nk from ck; n1 plus n2 plus nk is equal to n, the total number of samples.

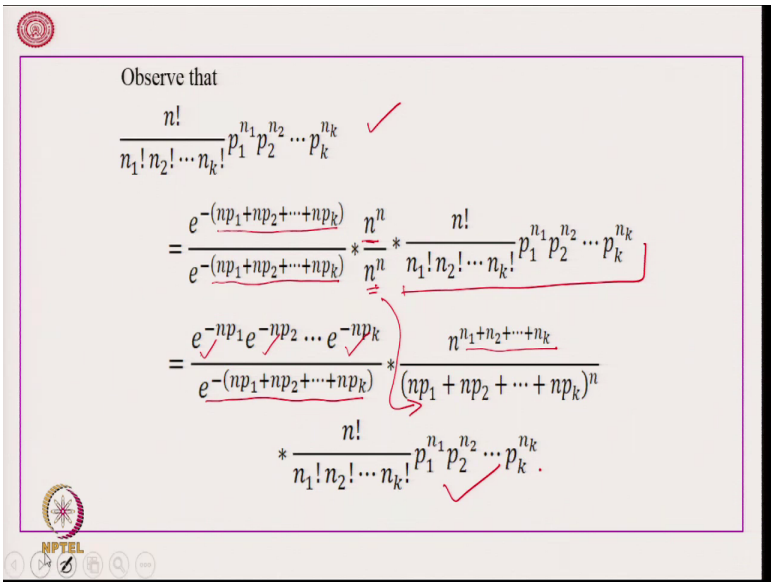
Now, how do we get that, it is very simple; so out of n we first select n1 of them that can be done in $^{n}C_{n1}$. And let us put them in class number one that is c1, so that probability is p1 to the power n1. Now, how many are remaining? n minus n1 are there; so out of that n minus n1, we choose n2 of them and put them in the class c2. That probability is p2 to the power n2; like that finally we will have there only nk many elements left.

So, nk cnk and that will give to the class K; that probability is pk to the power nk. So, this is the basic derivation of the multinomial distribution; if we simplify, it will see that n factorial upon n1 factorial into n minus n1 factorial, multiplied by let us simplify this. So, that will give you n minus n1 factorial, upon n2 factorial into n minus n1 minus n2 factorial. Like that we will go and

here we will get 1 multiplied by, as you can understand p1 to the power n1, p2 to the power n2, pk to the power nk.

And therefore if we look at these cancels with this in a similar way corresponding to class 3; this is going to cancel with the numerator. So, finally we will end up with n factorial upon n1 factorial, n2 factorial up to nk factorial, into p1 to the power n1, p2 to the power n2, pk to the power nk. So, that is how the multinomial density function is obtained. I hope most of you are familiar with this, but still I wanted to clarify the density function.

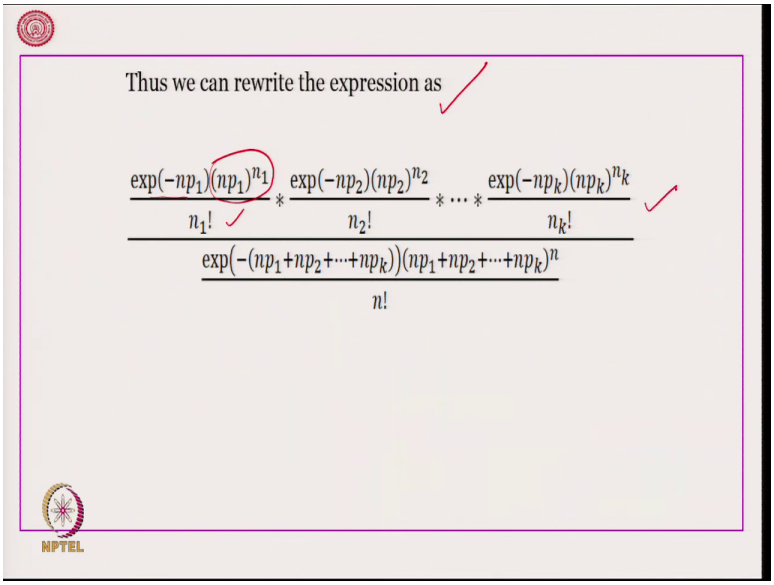
(Refer Slide Time: 11:23)



Now, we shall work with this formula, we write it slightly complicated way. So, we write it as e to power minus np1 plus np2 plus npk, and divide by the same quantity. Similarly, again we multiply it by n to the power n and divide by the same quantity; and then this is the formula that we already had. Now, let us simplify this term, so we can write this as e to the power minus np1, e to the power minus np2, up to e to the power minus npk; denominator we keep unchanged.

And numerator again this one we write, n to the power n1 plus n2 plus nk, divided by this n to the power n. We write it as np1 plus np2 plus npk whole to the power n, multiplied by the standard density thing; n factorial upon n1 factorial et-cetera p1 to the power n1, p2 to the power n2 into pk to the power nk. So, this is little bit of I gave jugglery that we have done, but for some purpose as it will be clear in the next slide.

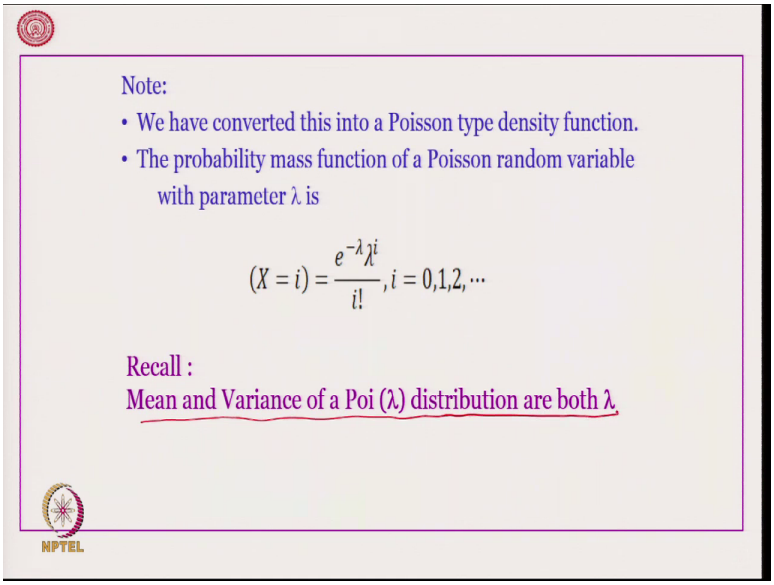
(Refer Slide Time: 12:50)



So, we write the entire expression, now as e to the power minus np1 into np1to the power n1, upon n1 factorial. So, let us go back to the slide to the power minus np1, from here then n to the power n1 multiplied by p1 to the power n1. This we write it as np1 to the power n1, then there is a n1 factorial which comes here; and therefore entire expression as, e to the power minus , into np1 to the power n1, upon n1 factorial, into like that up to e to the power minus npk, npk to the power nk upon nk factorial.

This whole thing divided by e to the power minus np1 plus np2 plus npk, multiplied by np1 plus np2 plus npk whole to the power n. This is in the denominator divided by n factorial, this comes from here.

(Refer Slide Time: 14:10)

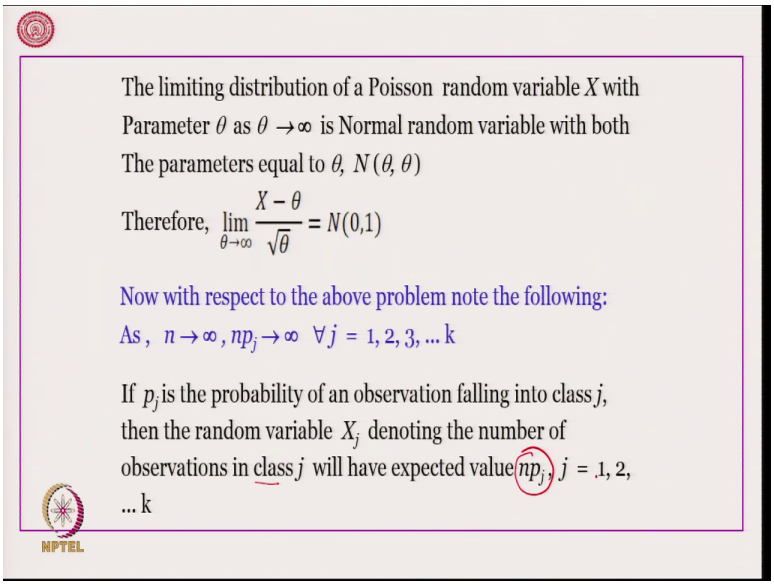


Therefore, what we have done? So, we convert the whole thing into a Poisson type density function, and I am sure all of you are familiar with Poisson. A Poisson random variable X takes the value i with the probability,

$$\frac{e^{-\lambda}\lambda^i}{i!}, i = 0,1,2,\cdots$$

where lambda is the parameter of the distribution. Now, recall that the mean and variance of a Poisson lambda distribution are both lambda.

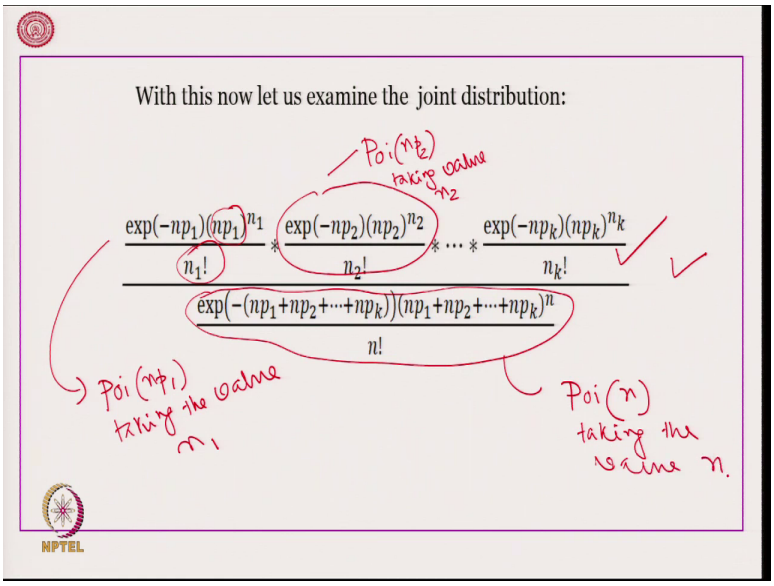
(Refer Slide Time: 14:59)



Now, the limiting distribution of a Poisson distribution or a Poisson random variable X with parameter theta. As theta goes to infinity is a normal random variable with both the parameters equal to theta, that is normal theta, theta. Therefore, limit theta going to infinity, X minus theta upon root theta; this converges to normal 0, 1; because, we have standardized it by subtracting the mean and dividing it by the standard deviation; which is root over theta.

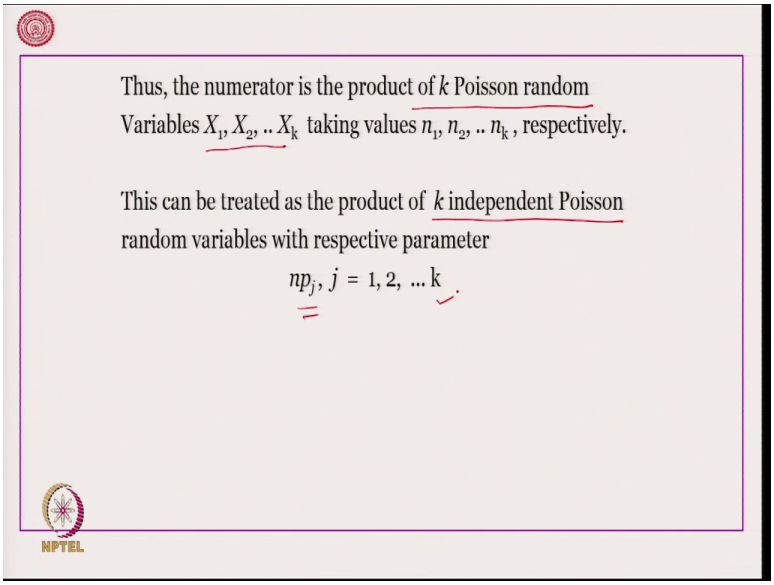
Now, with respect to the above problem that is the multinomial distribution note that, if n goes to ∞; then npj goes to ∞ for all j, j is equal to 1 to k. Hence, the pj is the probability of an observation falling into the class j, then the random variable Xj denoting the number of observations of class j; will have expected value npj. This is very clear because the proportion coming from class j is pj, for j is equal to 1 to k.

(Refer Slide Time: 16:22)



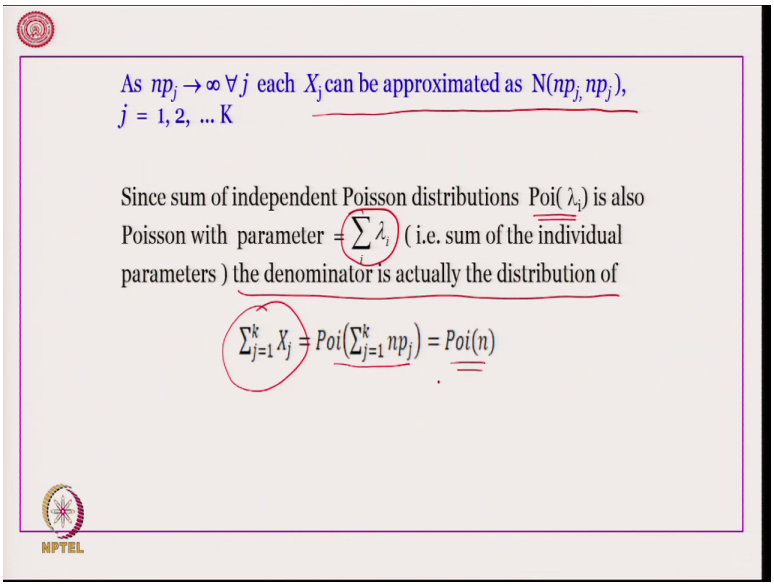
With this now let us examine the expression just we have just obtained some time back. So, it is e to the power minus lambda lambda power n1, upon factorial n1; so, this looks like a Poisson random variable with parameter np1 taking the value n1. Similarly, this is a Poisson random variable with np2 as the parameter or we can write it as Poi with np2, taking the value n2. Similarly for the kth one and in the denominator we have it is a Poisson random variable with parameter np1 plus np2 plus npk. That means, it is a Poisson random variable with parameter n taking the value n.

(Refer Slide Time: 17:38)



Therefore, the whole numerator is the product of k Poisson random variables X1, X2, Xk taking values n1, n2, nk. And this can be treated as a product of k independent Poisson random variables with respect parameters np1, np2 up to npk.

(Refer Slide Time: 18:02)



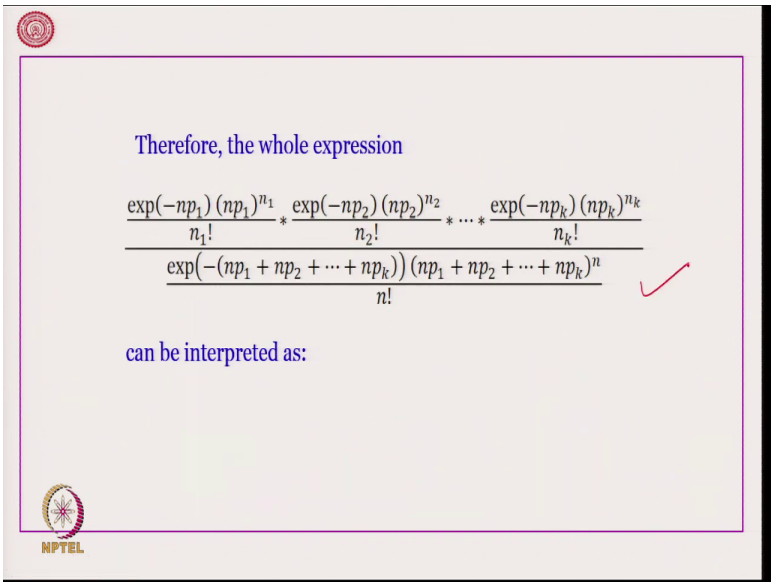
Now, as npj goes to infinity for all j; because if we take large samples each Xj can be approximated as normal with npj comma npj. This we have just shown that as theta goes to infinity, Poisson converges to normal with theta comma theta; here theta is npj for the jth class.

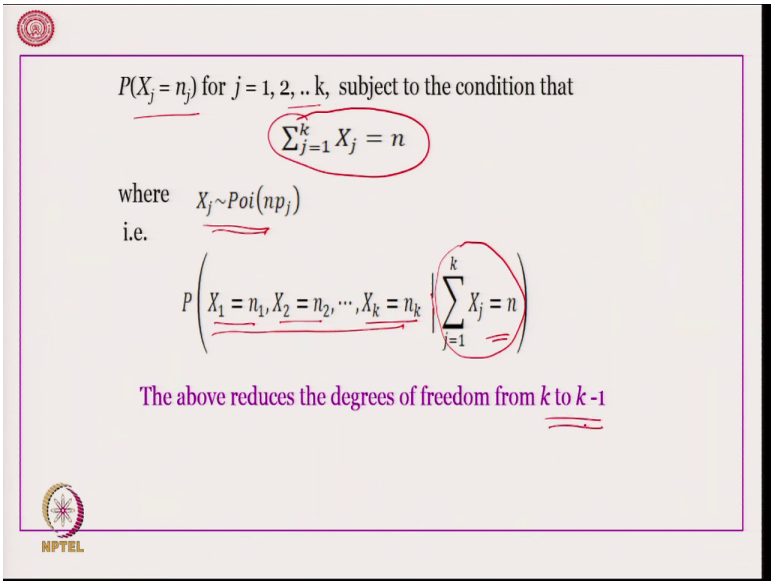
Now, since sum of independent Poisson distributions, what the ith one is Poisson with lambda i, is also a Poisson distribution with parameters

$$\sum_i \lambda_i$$

That is the sum of individual parameters, the denominator of the earlier distribution is equal to sigma Xj that is Poisson with parameter n, as we have discussed just now.

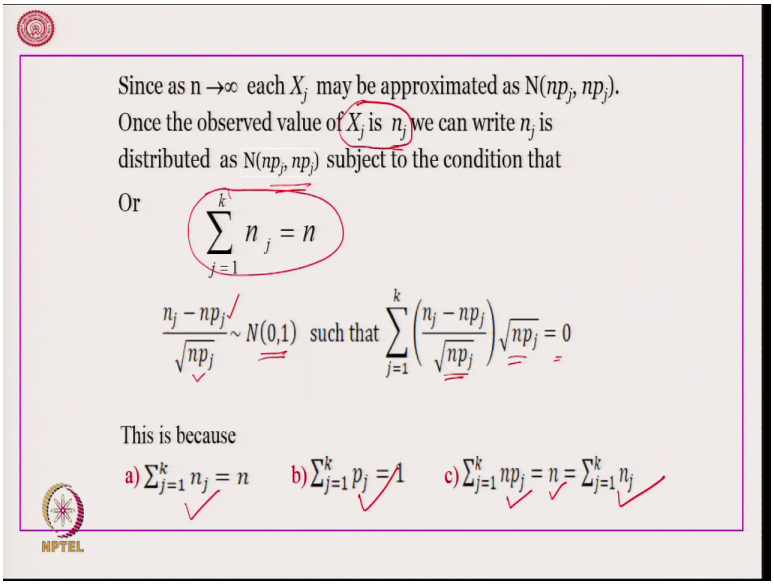
(Refer Slide Time: 18:56)





Therefore, the entire expression can be interpreted as probability Xj is equal to nj, for j is equal to 1 to k, subject to the condition that sigma Xj is equal to n; because the sum of total observations has to be n; where Xj is equal to Poisson of with npj. That is we are looking at the probability of the event X1 is equal to n1, X2 is equal to n2, Xk is equal to nk; subject to the condition that sigma Xj is equal to n. Since all the k values are not independent, there is a constraint there; therefore the degrees of freedom is reduced from k to k - 1.

(Refer Slide Time: 19:52)



Now, since n goes to ∞, each Xj may be approximated as normal npj comma npj. And therefore once the observed value of Xj is nj, we can write nj is distributed as

$$N(n_j, n_j)$$

subject to the condition that

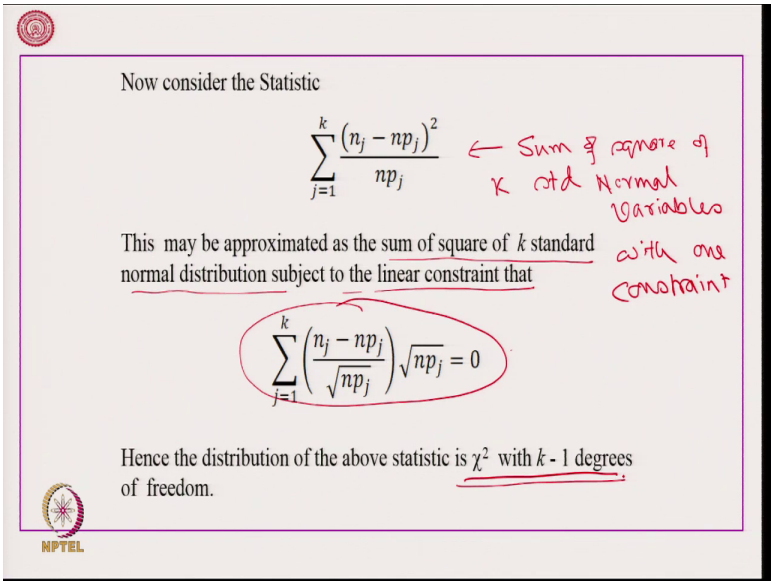
$$\sum_{j=1}^{k} n_j = n$$

Therefore, nj is the observed value, npj is the expected value divided by square root of npj. That is going to give a normal 0, 1 distribution, subject to the condition that

$$\sum_{j=1}^{k}\left(\frac{n_j - np_j}{\sqrt{np_j}}\right)\sqrt{np_j} = 0$$

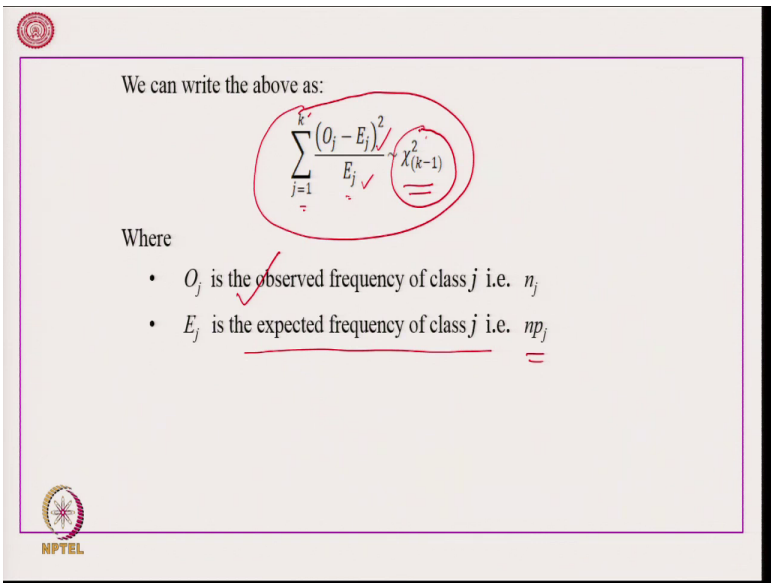
This is obvious because if we cancel it, then this numerator will give me 0. Why it is happening? Because sigma nj is equal to n, sigma pj is equal to 1, and sigma npj is equal to n is equal to nj; where nj is the observed value of the Xj.

(Refer Slide Time: 21:00)



Now, consider the statistic nj minus npj whole square divided by npj. This maybe approximated as the sum of square of k standard normal distribution subject to the linear constraint that, nj minus npj upon root over npj, multiplied by root over npj, and summation over K is equal to 0. Hence, what is this? It is the sum of square of k standard normal variables with one constraint. Therefore, this is a Chi-Square distribution with k minus 1 degrees of freedom; because we know that if there are m standard normal distribution x1, x2, xn. Then x1 square plus x2 square plus xn square is chi square with m degrees of freedom.

(Refer Slide Time: 22:20)

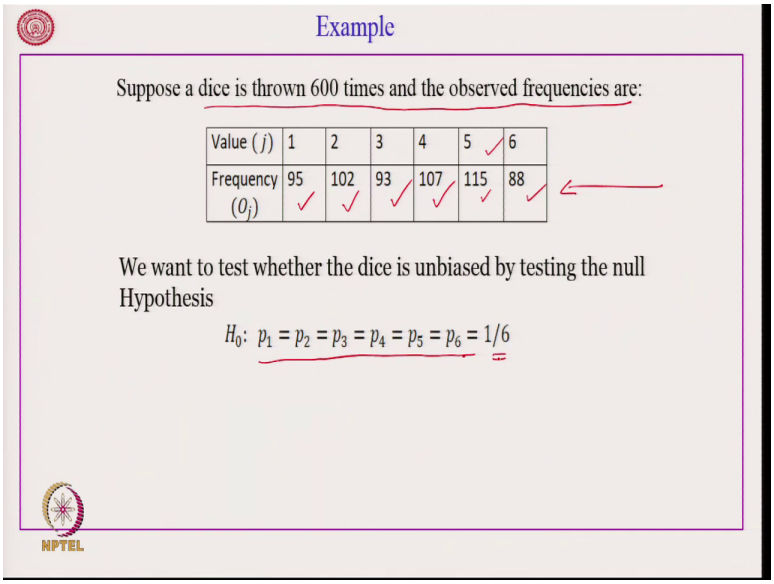


So, this same thing typically we write it as

$$\sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j} \sim \chi^2_{(k-1)}$$

where Oj is the observed frequency of the class j. That is nj, the number of observation coming from the class j, and Ej is the expected frequency of the class j; which we have just seen to be npj. Therefore, observed minus expected whole square divided by expected, this sum for j is equal to 1 to k; that is the number of classes is chi square with k minus one degrees of freedom.

(Refer Slide Time: 23:10)

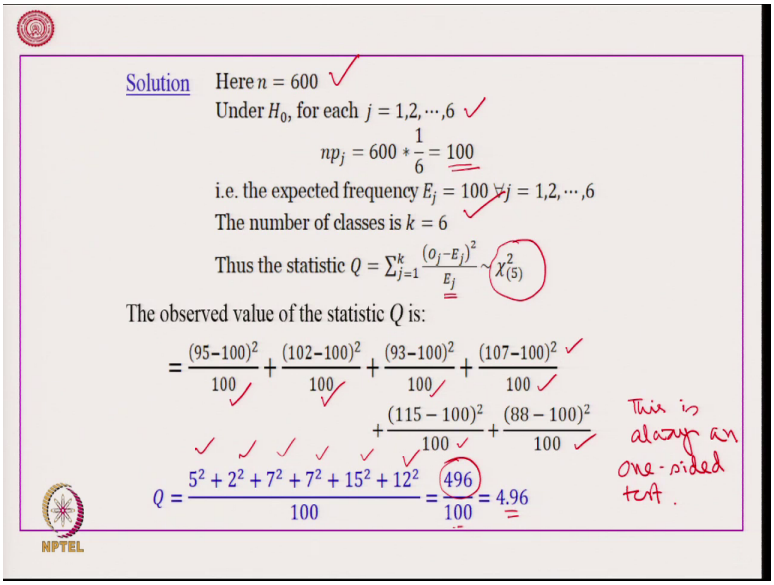


So, let us go back to an example, suppose a dice is thrown 600 times and the observed frequencies are like this: 1 appeared 95 times, 2 appeared 102 times, 3 appears 93 times, 4 appears 107 times, 5 115 times, 6 is 88 times. So, what we can see that there is some variation in the number of observations corresponding to the 6 different faces of the die. But, does it mean that the die is not unbiased, we are not sure; we need to test that. So, what we are testing is

$$H_0: p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$$

So, this is the null hypothesis that we are testing against the dice not unbiased.

(Refer Slide Time: 24:08)



So, what we do? We do the following n is 600, there are 6 classes, under the null hypothesis npj is equal to 100; that is expected number of observations coming from each class is 100 that is my Ej. Therefore, the statistic is going to be

$$\sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j},$$

and we know that this is going to be chi square with 5 degrees of freedom.

Now, let us compute the value of the statistic, therefore it is

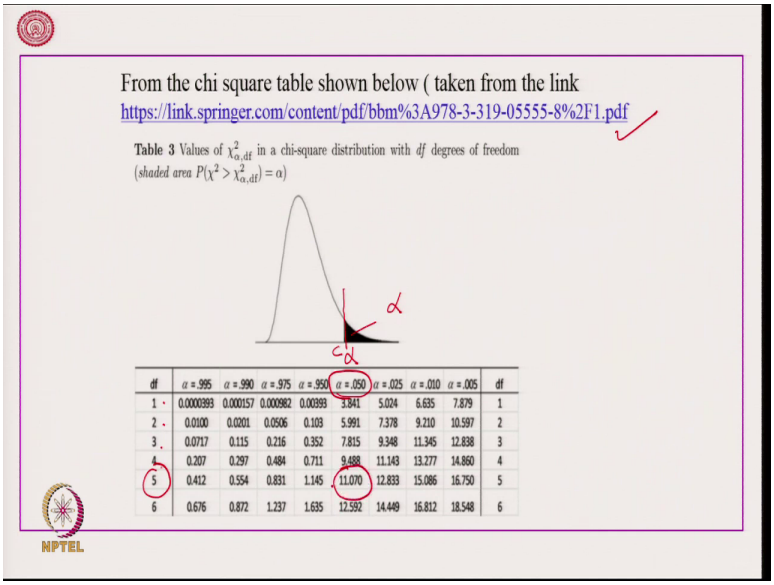
$$\frac{(95-100)^2}{100} + \frac{(102-100)^2}{100} + \frac{(93-100)^2}{100} + \frac{(107-100)^2}{100}$$

$$+ \frac{(115-100)^2}{100} + \frac{(88-100)^2}{100}$$

now, you understand how to compute the statistic. So, this value is coming out to be 5 square plus 2 square plus 7 square plus 7 square plus 15 square plus 12 square; which is coming out to be 496, divided by 100, that is 4.96.

Now, we shall reject the null hypothesis, if this value is very large; therefore this is always going to be one sided test. So, how it will work?

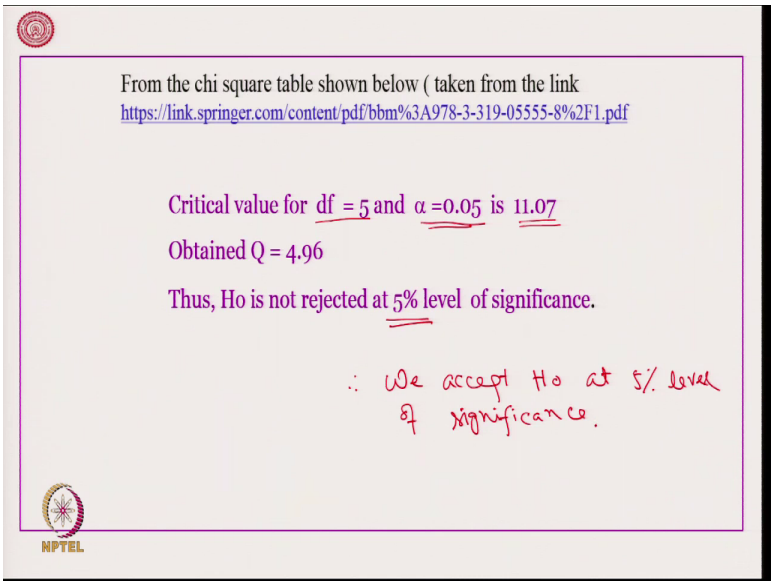
(Refer Slide Time: 25:53)



df	α = .995	α = .990	α = .975	α = .950	α = .050	α = .025	α = .010	α = .005	df
1	0.0000393	0.000157	0.000982	0.00393	3.841	5.024	6.635	7.879	1
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597	2
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	11.070	12.833	15.086	16.750	5
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548	6

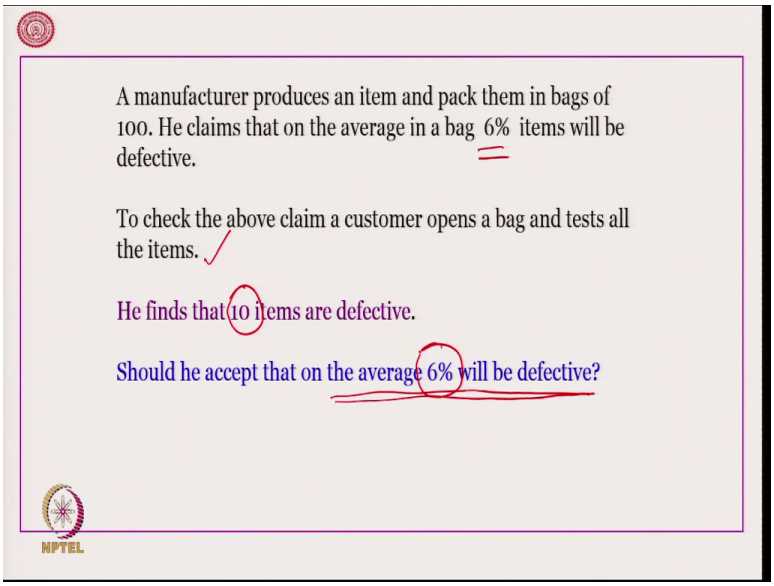
So, here is a chi square distribution, there are chi square table available; I have given you one link for different degrees of freedom. Here I have just copied for 1, 2, 3, 4, 5, 6 and the cutoffs are given, which are the critical values for different alpha. We are looking at 5 percent of level of significance that means we are looking at an α, such that the probability the chi square distribution is taking a value greater than this $C\alpha$ is α. So, these $C\alpha$ values are tabulated here, we have chi square with 5 degrees of freedom; we are looking at α is equal to 0.05, and the $C\alpha$ value is given out to be 11.07.

(Refer Slide Time: 26:54)



Therefore, what I am writing here is that the degrees of freedom is 5, alpha is equal to 0.05, the critical value is 11.07. The obtained value is 4.96, therefore we cannot reject at 5 percent level of significance; therefore, we can accept the hypothesis at 5 percent level of significance.

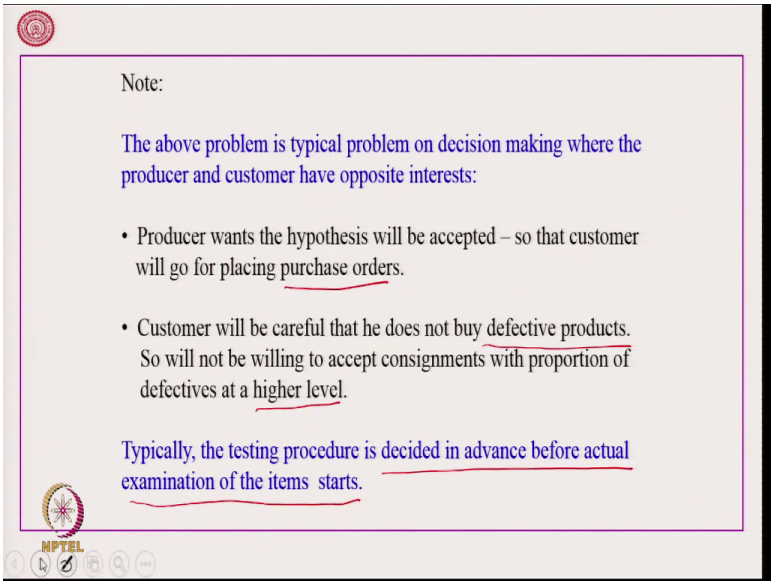
(Refer Slide Time: 27:40)



Let me now give you a practical example, suppose a manufacturer produces an item and pack them in bags of 100. He claims that on the average in a bag 6% items will be defective. Now, to check the above claim a customer opens a bag and tests all the items; he finds that 10 of the items out of 100 are defective. Question comes whether he should accept that on the average 6% of the items are going to be defective?

That means that in another bag there maybe lesser number of defectives; like that when we keep on taking different bags of samples. We may end up with at the end that on the average, the number of defectives is going to be about 6% the total number of products.

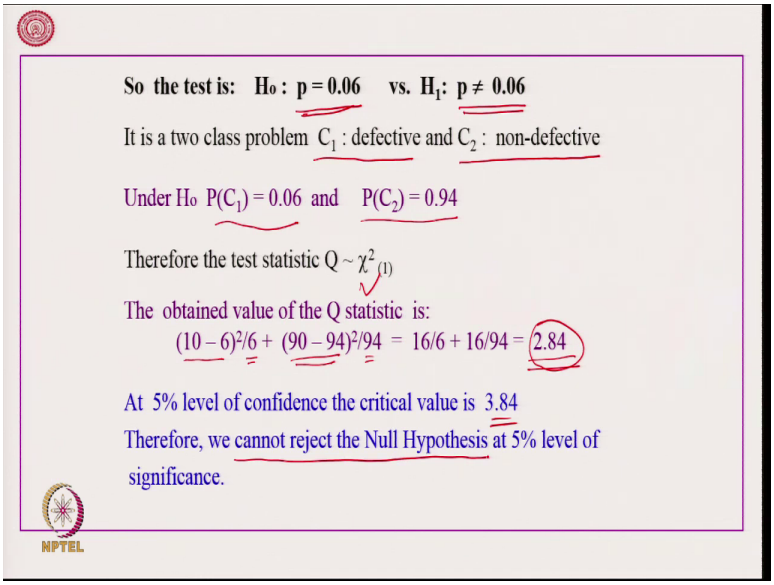
(Refer Slide Time: 28:46)



Now, this is a very very typical problem on decision making; where the producer and the customer have opposite interests. A producer wants the hypothesis will be accepted, so that the customer will go for placing purchase orders. Because, if the purchaser is convinced that the number of defectives is small; then he is likely to go for purchasing the items from the producer. However, the customer or the purchaser will be careful that he does not buy defective products; so, will not be willing to accept consignments with proportion of defectives at a higher level.

Therefore, if the test fails that the number of defectives on the average is greater than 6% then customer may not like to purchase these items from the producer. So, typically how the test will be carried out that is decided in advance before the actual examination of the item starts. So, we are not going into the detail of this statistical procedures.

(Refer Slide Time: 30:00)

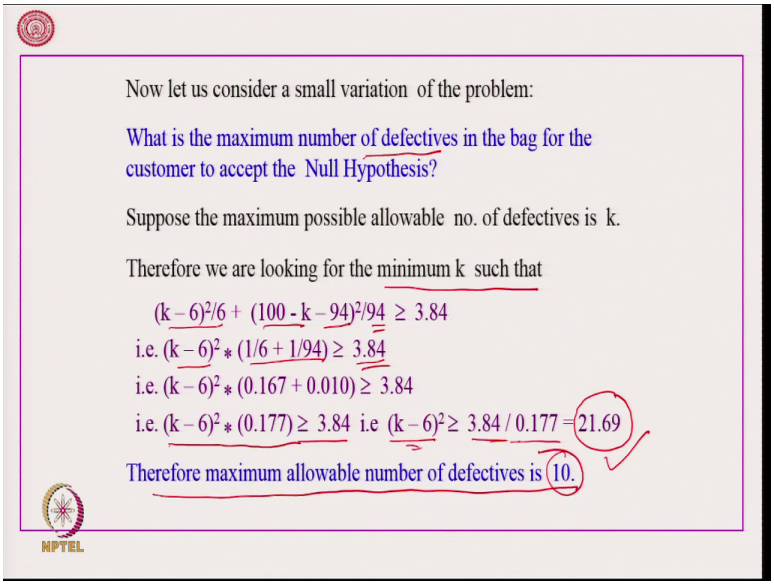


So, let us look at how we test for Ho p = 0.06 versus H1 p ≠ 0.06. So, we are looking at two sided tests. Therefore, it is a two class problems, where c1 is the class of defectives and C2 is the class of non-defectives. Therefore, under H naught probability of class 1 is equal to 0.06, and probability of class 2 is equal to 0.94. Therefore, the test statistic Q is chi square with 1 degrees of freedom. Why 1 degree? Because there are two classes. And we know that there are k classes then the chi square will have k-1degrees of freedom.

The obtained value of the Q statistic is therefore 10 minus 6 whole square divided by 6, plus 90 minus 94 whole square divided by 94; which is coming out to be 2.84. Therefore, now the problem is whether we should accept the null hypothesis or reject, for that we have to refer to the chi square table. And we see the following that for 1 degrees of freedom at 5% level of significance; the critical value is 3.84.

We have taken the table from this link, so that you can also check how to see a chi square table. But, going back to the problem, we can see that our obtained value is somewhat smaller than the tabulated value of $C\alpha$; when $\alpha = 0.05$. Therefore, we cannot reject the null hypothesis at 5% level of significance.

(Refer Slide Time: 32:10)



Now, let us consider a small variation of the problem. Suppose we want to ask, what is the maximum number of defectives in the bag, for the customer to accept the null hypothesis? That means that if the number of defective is less than that value say k, then the customer is going to accept the null hypothesis; otherwise he is going to reject the null hypothesis. So, what is going to be the maximum number of defectives that is going to be the question?

So, we are therefore looking at the minimum value of k, such that

$$(k-6)^2/6 + (100-k-94)^2/94$$

That has to be greater than equal to 3.84, if we have to reject the null hypothesis. That is

$$(k-6)^2 * (1/6 + 1/94)$$

has to be greater than equal to 3.84. Now, that boils down to k minus 6 whole square multiplied by 0.1 double seven has to be greater than equal to 3.84.

That is

$$(k-6)^2 \geq 3.84 / 0.177$$

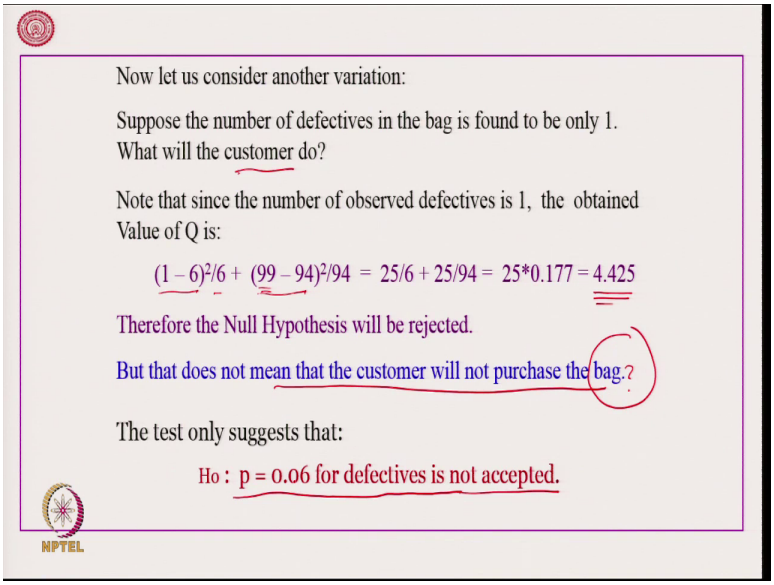
which is coming out to be 21.69. Therefore, we shall reject the null hypothesis, if k is such that

$$(k-6)^2 \geq 3.84 / 0.177 = 21.69$$

Therefore we can easily see that the value of k can be 11. Hence, we can see that the maximum allowable number of defective is 10. Because, in that case k minus 6 whole square is coming out to be 4 square that is 16, which is less than 221.69 and we are going to accept the null hypothesis.

But, if k is equal to 11, then this is coming out to be 25 and therefore we are going to reject the null hypothesis; therefore, the maximum allowable number of defective is 10.

(Refer Slide Time: 34:43)

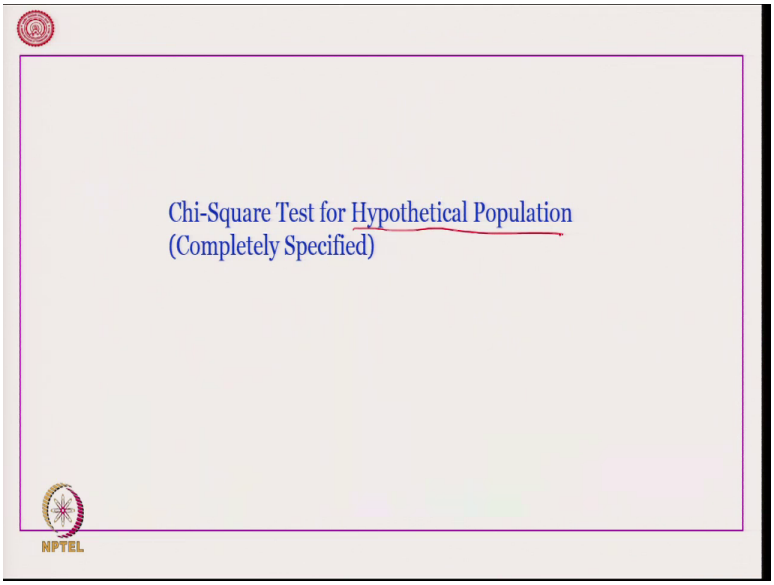


Now, let us consider another variation: Suppose the number of defectives in the bag is found to be only 1. Question is what will the customer do? Note that since the number of observed defective is 1; the obtained value of Q is going to be 1 minus 6 whole square by 6. Then 99 minus 94 whole square divided by 94, this is coming out to be 99; because since there is only one defective, 99 of the other elements or items are non-defective. This value is coming out to be 4.425, which is greater than the α that is 3.84 which we have seen sometime back.

What does it mean? Does it mean that the customer will not purchase the bag because we are rejecting the null hypothesis? No. The customer will go for purchasing the bag because number of defective has been found to be very small. Then what does it mean that we are going to reject

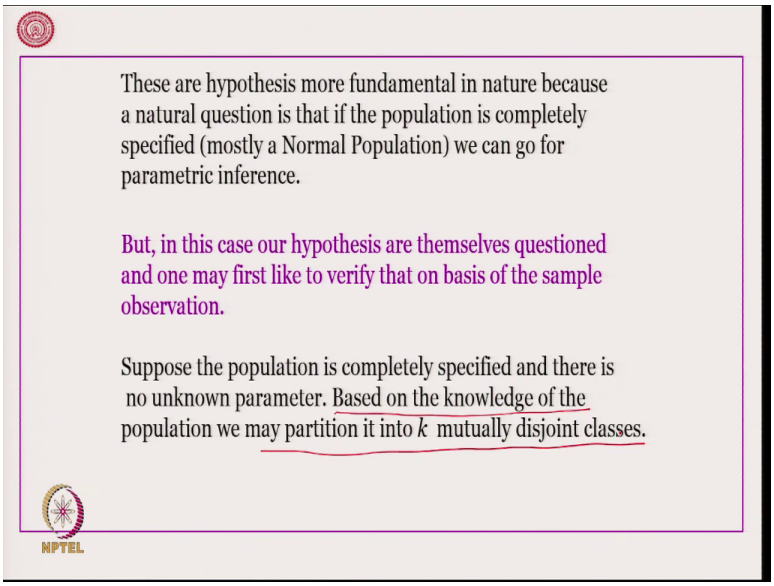
the null hypothesis; it means that we are rejecting that p is equal to 0.06. That is the proportion of defectives is not accepted, because the test suggest that the proportion of defective is quite different from 6 percent.

(Refer Slide Time: 36:17)



Now, we can use Chi-Square test in a slightly different way as follows. What I am saying Chi-Square test for hypothetical population which is completely specified.

(Refer Slide Time: 36:34)

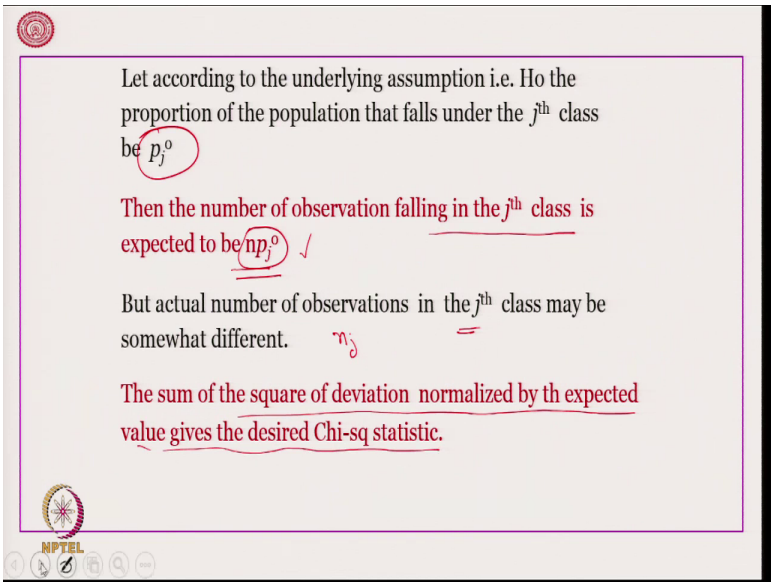


So, these are hypothesis more fundamental in nature because a natural question is that, if the population is completely specified; mostly is a normal population, then we can go for parametric inference. So, what does it mean? I stated at the very beginning that parametric inference is more robust or powerful than nonparametric; but, parametric inference works under normality assumption.

So, if we are given a sample and we can test, that it is actually coming from a normal population, then we are really happy because we can use all our parametric inference techniques. So, that is why sometimes we try to check if the given sample is coming from a standard distribution. And that is why I am saying that the hypothesis that it is coming from a normal population itself can be questioned, and we need to make sure that is correct.

That is that assumption is correct and therefore we like to verify that on the basis of a sample observation. So, how to do that? Suppose the population is completely specified and there is no unknown parameter. We know that if there is no unknown parameter, then the density is completely known to us; and therefore the distribution is known to us. But, based on that knowledge what we do? We try to partition the entire sample space into k mutually disjoint classes.

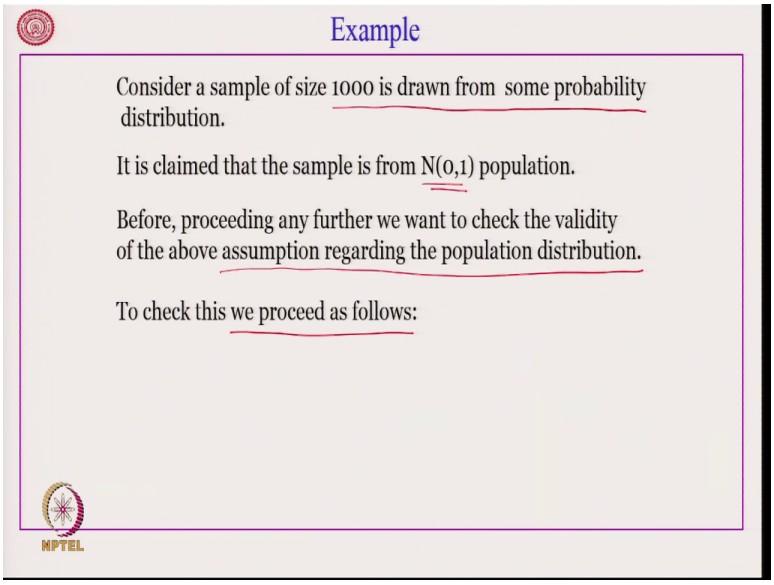
(Refer Slide Time: 39:04)



So, let according to the underlying assumption that is Ho is the proportion of population that falls under the jth class. Then the number of observations falling in the jth class is expected to be npj^0. Because this is the population proportion, n is the sample therefore, expected number of sample in the jth class is npj^0. But, actual number of observations in the jth class maybe somewhat different as we have already seen.

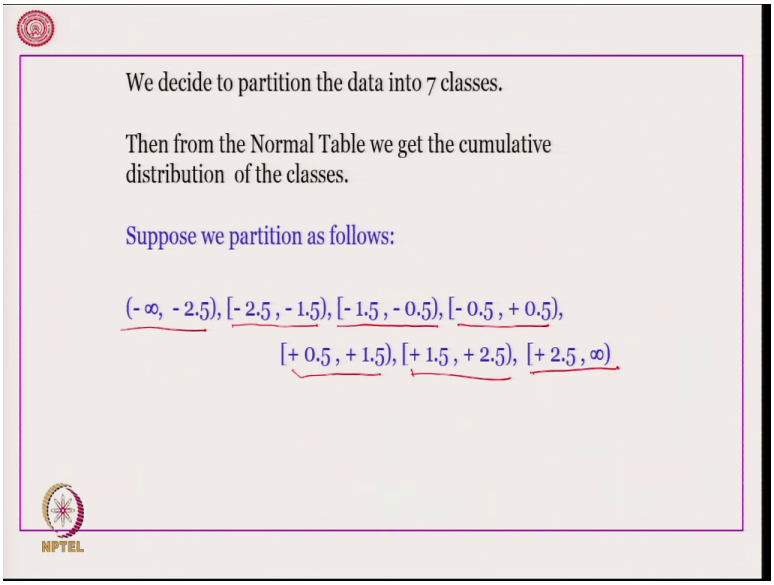
Thus, there may be a gap between the actual observations from the jth class, which if we call nj; between nj and npj^0. The sum of the square of the deviation normalized by the expected value gives the desired Chi-Square statistic. This we have already explained.

(Refer Slide Time: 40:02)



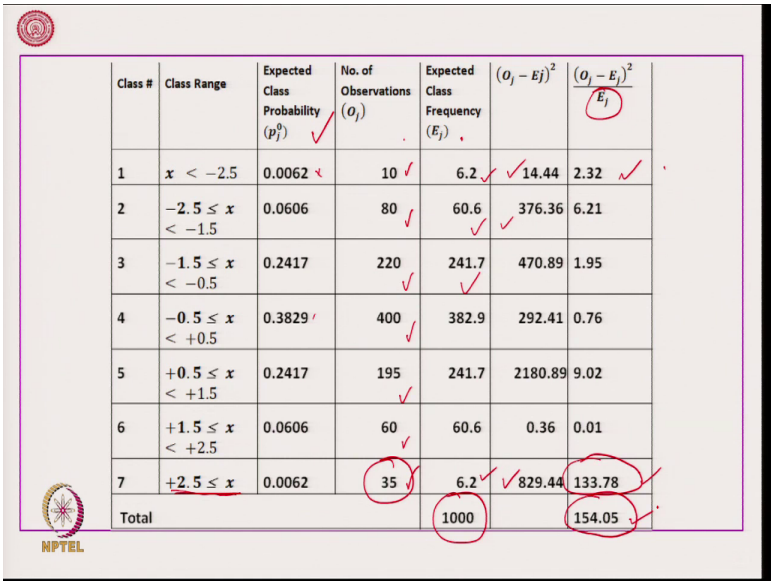
So, to consider a sample of size 1000 that is drawn from a probability distribution. We claimed that the sample is coming from N(0,)1 population; but we need to verify it. Therefore, before proceeding any further we want to check the validity of the above assumption regarding the population distribution. So, how to do that that is the question; so we proceed as follows.

(Refer Slide Time: 40:33)

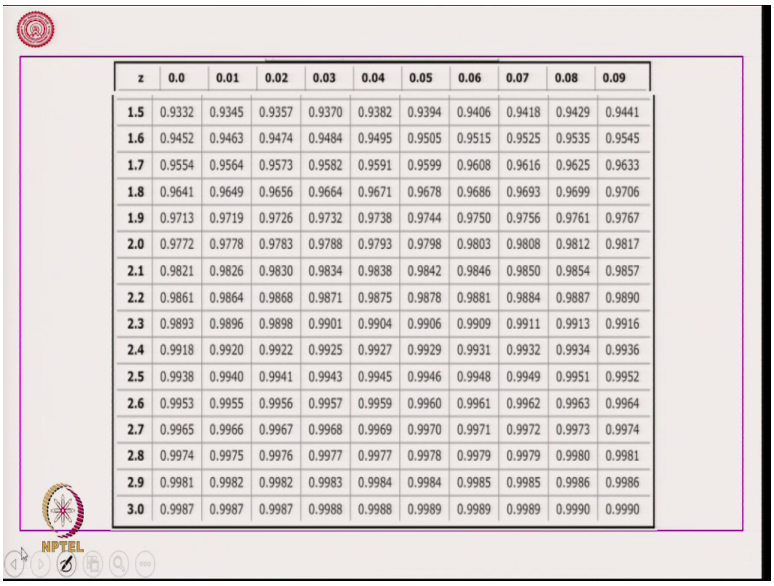


What we are doing? We decide to partition the data into 7 classes; then from normal table we get the cumulative distribution of the classes. Suppose we partition it as follows:- ∞ to - 2.5, - 2.5 to -1.5, - 1.5 to - 0.5, -0.5 to +0.5; similarly, + 0.5 to 1.5, 1.5 to 2.5 and 2.5 to ∞. Thus, there a 7 classes.

(Refer Slide Time: 41:13)



Class #	Class Range	Expected Class Probability (p_j^0)	No. of Observations (O_j)	Expected Class Frequency (E_j)	$(O_j - Ej)^2$	$\frac{(O_j - E_j)^2}{E_j}$
1	$x < -2.5$	0.0062	10	6.2	14.44	2.32
2	$-2.5 \le x < -1.5$	0.0606	80	60.6	376.36	6.21
3	$-1.5 \le x < -0.5$	0.2417	220	241.7	470.89	1.95
4	$-0.5 \le x < +0.5$	0.3829	400	382.9	292.41	0.76
5	$+0.5 \le x < +1.5$	0.2417	195	241.7	2180.89	9.02
6	$+1.5 \le x < +2.5$	0.0606	60	60.6	0.36	0.01
7	$+2.5 \le x$	0.0062	35	6.2	829.44	133.78
Total				1000		154.05



z	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

So, these are the 7 class; x is less than minus 2.5, x is between minus 2.5 to minus 1.5 up to x is greater than plus 2.5. Now, the question is what is the expected class probability Pj0? So, x less than minus 2.5 is 0.0062. How to get that? We get it from the normal table. So, this is minus 2.5, this is 0. So, this is the cumulative probability of a standard normal distribution taking value less than minus 2.5 that is 0.0062. Therefore, that we have obtained. Now, let me do it one or two more for your understanding. What is the probability that a standard normal random variable is between minus 2.5 and minus 1.5.

So, again we go back less than minus 1.5 is 0.0668 0.0668 and less than minus 2.5 is 0.0062. Therefore, we get the difference is 6060. So, the probability is 0.0606 and that is what we have written here. In a similar way we can compute for all I have given the entire normal table for your help. You can compute from here the class probabilities for all the 7 classes which I am showing here, like this in this column. Suppose the actual observations belonging to this 7 classes are as follows: 10, 80, 220, 400, 195, 60 and 35.

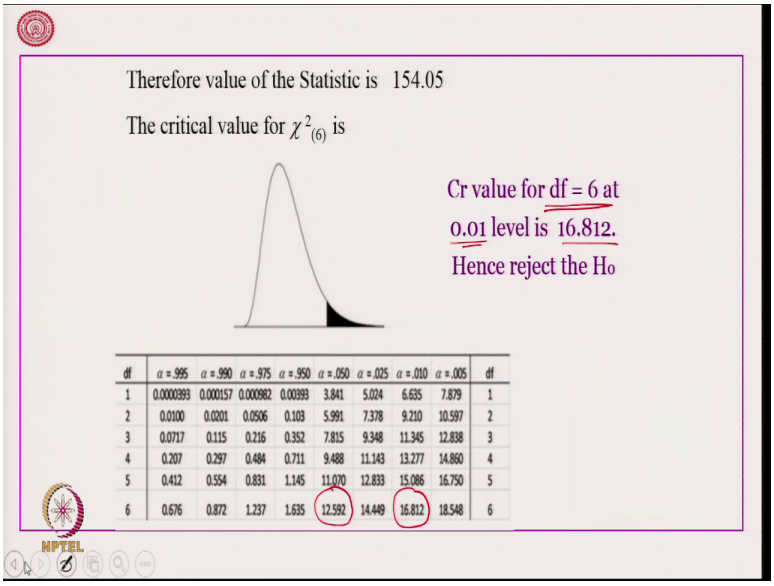
Therefore expected class frequency how do we get? We multiply these values by 1000; therefore we get 6.2, 60.6, 241.7 like that up to 6.2, and that should sum up to 1000. So, this is the expected class frequency, this is the observed values. Therefore we need to compute the square of their difference, which is coming out to be like this 14.44, 376.36 up to 829.44. Therefore, for each one of them we calculate this normalized thing by dividing by Ej, and these are the values that we obtained.

And the sum of this values

$$\sum_j \frac{(O_j - E_j)^2}{E_j},$$

summing over j is equal to 1 to 7, we get 154.05. Is it too high? We do not know but looking at it; I can itself see that this cannot be from normal 0, 1. Because x greater than 2.5 that should number should be very small; but here we find that number is pretty large which is 35. And if we look at this sum of square, we can see that that is giving us a huge value, making the sum to be too high. But, still it can be from normal 0, 1.

(Refer Slide Time: 45:15)



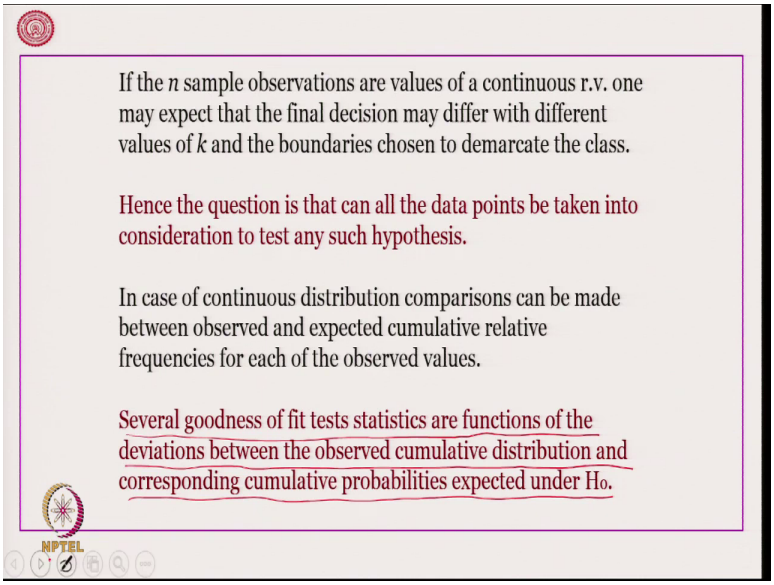
df	α = .995	α = .990	α = .975	α = .950	α = .050	α = .025	α = .010	α = .005	df
1	0.0000393	0.000157	0.000982	0.00393	3.841	5.024	6.635	7.879	1
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597	2
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	11.070	12.833	15.086	16.750	5
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548	6

So, let us check with the chi square value; so, again we look at the chi square table. This is going to be chi square with 6 degrees of freedom; because there are 7 classes. And therefore k is equal to 7 and therefore k minus 1 is equal to 6 is the degrees of freedom; therefore, we look at chi square with 6and for 0.05, the critical value is 12.592. Therefore, so first we look at what is the critical value at 1% level of confidence that is 16.812 and at 5 % level of confidence, it is 12.592.

Therefore, since for degrees of freedom at 6 at 1 percent level; the value is 16.812. Therefore, we reject the null hypothesis at 1% level of significance; therefore quite naturally it has to be

rejected at 5% level of significance as well. Or, in other words we are not going to accept that this data is coming from a standard normal distribution.

(Refer Slide Time: 46:46)

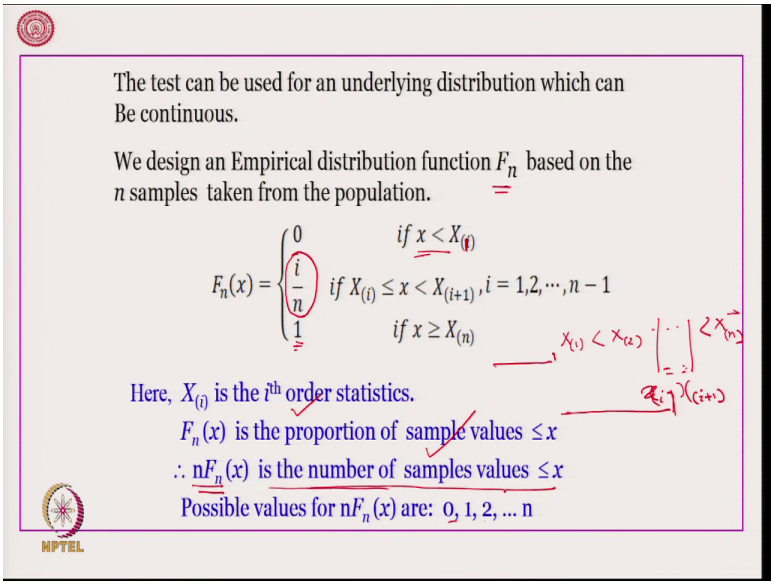


However, one can think of that if n is the sample observations and they are coming from a continuous random variable. One can expect that the final decision may differ with the different values of k and the boundaries chosen to demarcate the class. I have just given you an example with only one way of choosing the class boundaries, and the number of classes is equal to 7. One can think of making different class boundaries, different number of classes and test. I will leave it as an exercise for you to check with a few different classes and computing the corresponding chi square statistic.

Therefore, one important problem is that here we have considered elements belonging to a particular class. Therefore, the points are losing their individuality, can we make a test where we look all data points into consideration to test such a hypothesis. In case of continuous distribution comparisons can be made between observed and expected cumulative relative frequencies, for each of the observed values.

So, instead of putting them into several classes, we look at each point in a sorted order, in an increasing order and check up to that point what is the observed frequency, and what could have been the actual frequency, if you are looking at certain underlying distribution. And then we try to see how much difference it is actually generating; so let us go and illustrate. So, several goodness of fit tests statistics are functions of the deviations between the observed cumulative distribution and the corresponding cumulative probabilities expected under Ho.

(Refer Slide Time: 49:00)

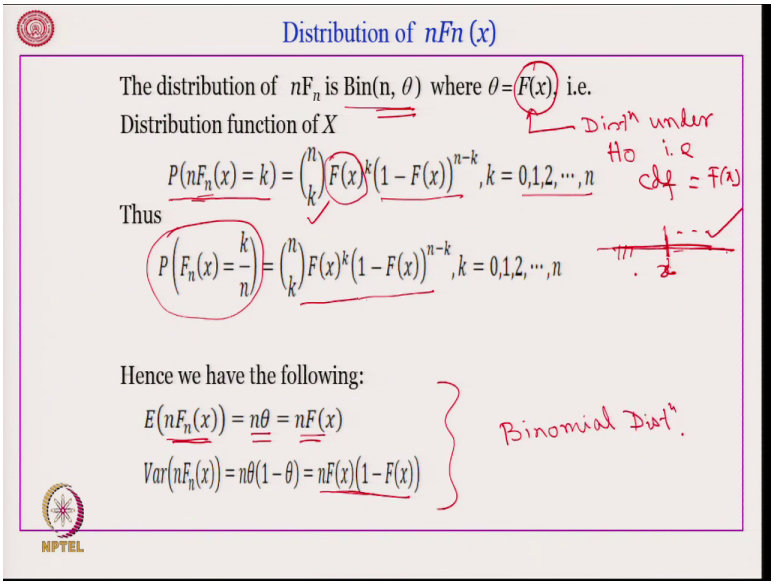


So, we shall study one major algorithm in this regard which is called Kolmogorov-Smirnov one sample test. The test goes as follows: we design an Empirical distribution function Fn based on the n samples taken from the population. So, we have the ordered as I have said that we have to take them in sorted order; so, x1 less than x2 less than xn is the ordered sample. What does it mean? Below this point there was no observation; that is Fn x is equal to 0, if x is less than x1. And xn has to be 1, if x is greater than xn.

So, for any x more than this we know that the entire sample is less than that. Therefore, the cumulative Empirical distribution function Fn will take the value 1. In between for xi and xi plus 1, if these are the observed statistic; any x between this interval means that there are i many observations out of n less than this. Therefore, what we are getting the cumulative probability to be i by n. So, that is how we design the Empirical distribution function Fn, where xi is the ith order statistic.

Fn (x) is the proportion of sample values less than equal to x, n Fn x is the number of sample value less than equal to x, possible values for n Fn Xn are 0, 1, 2 up to n. As you can understand that if I multiply this by n, we get that number of sample values less than equal to x.

(Refer Slide Time: 51:02)

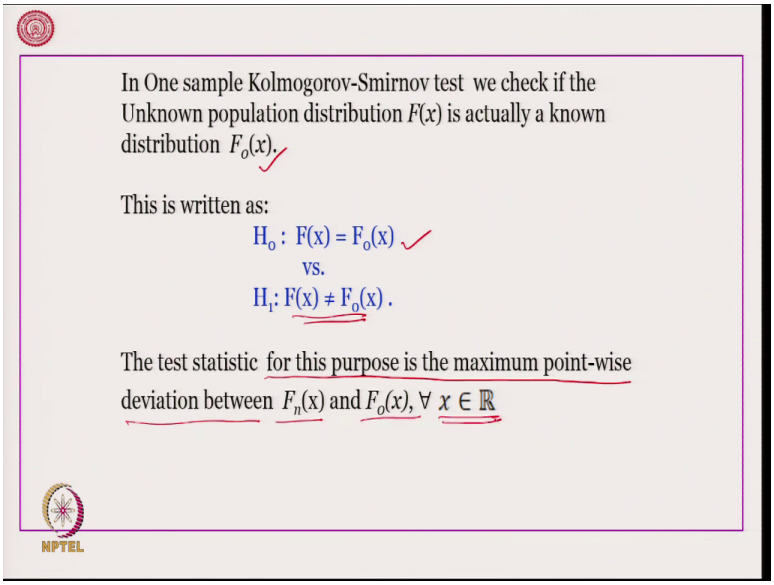


What is the distribution of n Fn x? The distribution of n Fn is binomial with n theta; where theta is equal to Fx. What is Fx? Fx is the distribution under null hypothesis; that is the cdf is equal to Fx. Therefore, under the null hypothesis, the distribution function Fx means that probability and observed value is less than equal to x is Fx. Therefore we are considering it to be a binomial distribution with Fx is the probability of falling in this side; and 1 minus fx is the probability that it will fall on the other side of x.

Therefore, probability n Fn x is equal to k is equal to out of n we are choosing k that is ${}^{n}c_{k}$; they are getting the probability Fx to the power k because they are less than equal to x. 1 minus Fx to the power n minus k because they are falling here; where k is equal to 0, 1, 2 up to n. Therefore, probability Fn x is equal to k by n, I am just dividing both side by n; is equal to nck Fx to the power k, 1 minus Fx to the power n minus k.

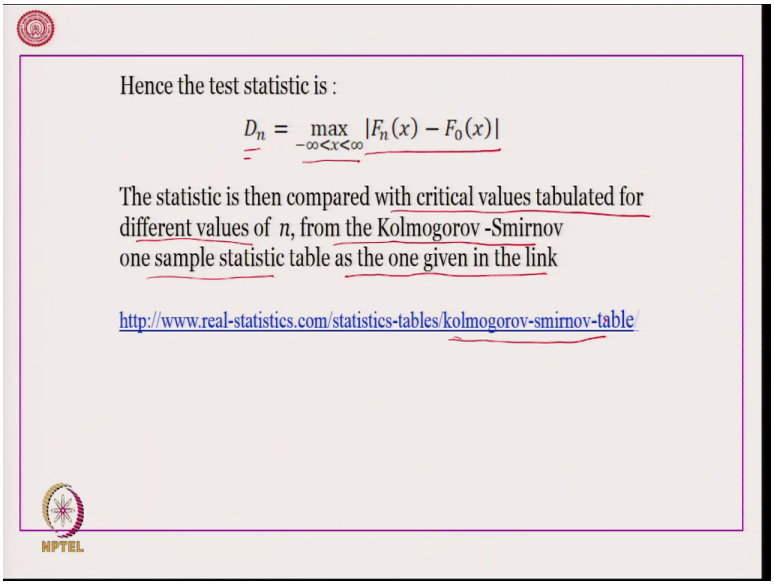
So, if that is a binomial distribution, then expected value of n Fn x is equal to n theta is equal to nFx; where Fx is the distribution. We are comparing with and its variance is going to be nFx into 1 minus. This is coming from binomial distribution.

(Refer Slide Time: 53:05)



What does it imply? It implies that the expected value of Fn x is equal to Fx, and variance of Fn x is equal to Fx into 1 minus Fx by n. In one sample Kolmogorov-Smirnov test, we check if the unknown population distribution Fx is actually a known distribution F0 x; that is you are checking Fx is equal to F0 x versus Fx not equal to F0 x. This test statistic for this purpose is the maximum point wise deviation between Fn x and F0 x; for all x belonging to R.

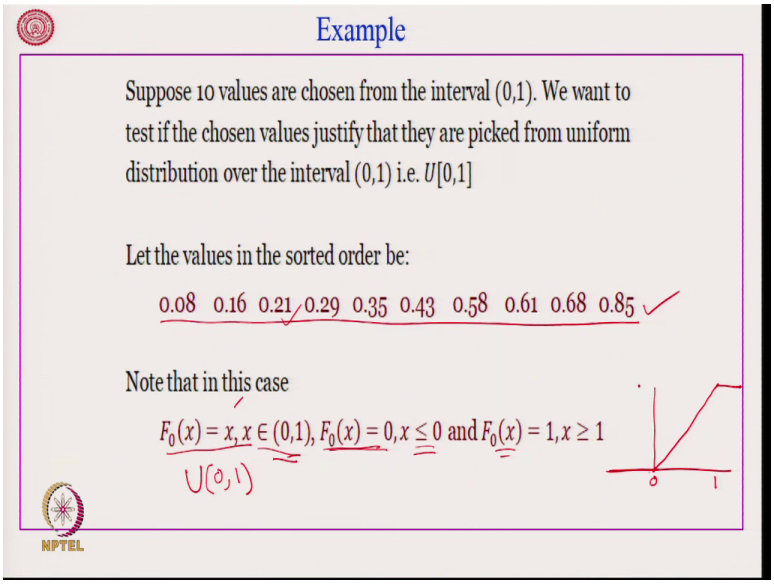
(Refer Slide Time: 53:52)



Therefore, the test statistic is going to be Dn is equal to maximum of Fn x minus F0 x; for x between minus infinity to infinity. This statistic is then compared with critical values of tabulated

for different values of n, from the Kolmogorov-Smirnov one sample statistic table as the one given in the following link. So, I am giving you here a link to check Kolmogorov-Smirnov cut off values or critical values for one sample test.

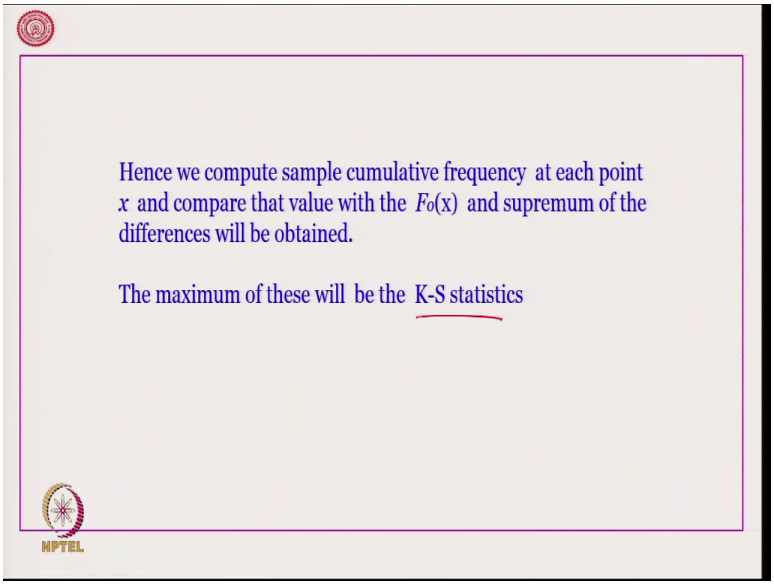
(Refer Slide Time: 54:31)



So, let me illustrate it with an example, suppose we have taken ten observations from uniform 0, 1 distribution. Let these values be like this 0.8. 0.16, 0.21, up to 0.85. Therefore, what F0 x, F0 x is uniform 0, 1; therefore F0 x is equal to x, for x belonging to 0, 1. Because uniform 0, 1 distribution that means it is ranging only from 0 to 1 that means it takes values only from 0 to 1.

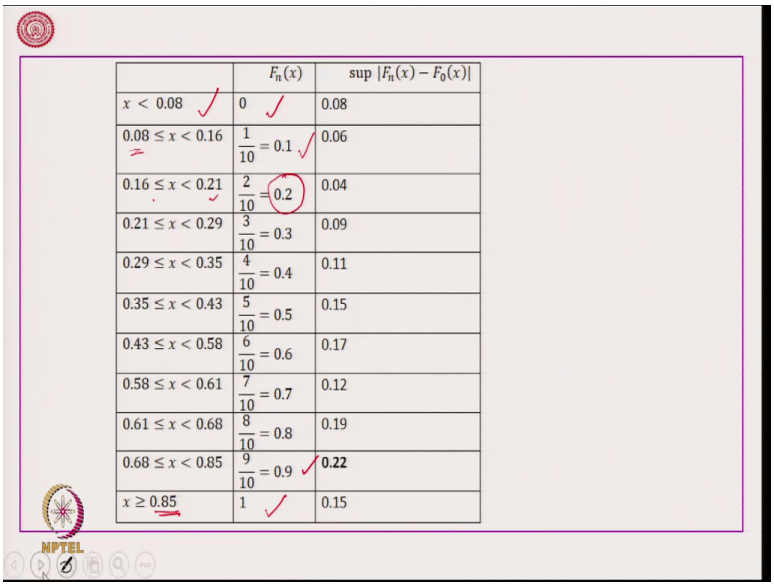
What is $F_0(x) = 0$, for $x \leq 0$; and $F_0(x) = 1$, for $x \geq 1$. Therefore, $F_0(x)$ will have a following shape it is 0 up to 0 and then it goes uniformly like this. At 1, 1 and then it takes the, it becomes horizontal. So, that is how this uniform distribution will look like.

(Refer Slide Time: 55:47)



Hence, we compute the sample cumulative frequency at each point x, and compare that value with the F0 x and we consider the supremum of the differences and that is what is going to give me the Kolmogorov-Smirnov of statistic.

(Refer Slide Time: 56:05)

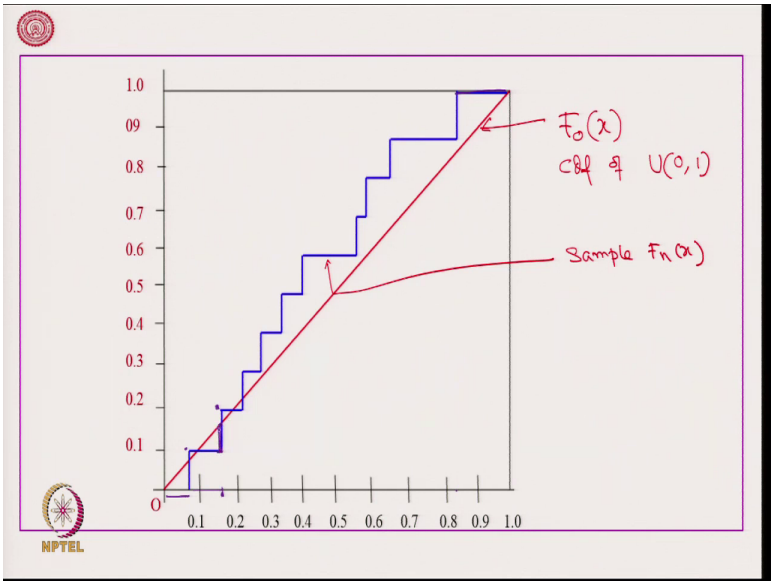


	$F_n(x)$	$\sup \lvert F_n(x) - F_0(x) \rvert$
$x < 0.08$	0	0.08
$0.08 \le x < 0.16$	$\frac{1}{10} = 0.1$	0.06
$0.16 \le x < 0.21$	$\frac{2}{10} = 0.2$	0.04
$0.21 \le x < 0.29$	$\frac{3}{10} = 0.3$	0.09
$0.29 \le x < 0.35$	$\frac{4}{10} = 0.4$	0.11
$0.35 \le x < 0.43$	$\frac{5}{10} = 0.5$	0.15
$0.43 \le x < 0.58$	$\frac{6}{10} = 0.6$	0.17
$0.58 \le x < 0.61$	$\frac{7}{10} = 0.7$	0.12
$0.61 \le x < 0.68$	$\frac{8}{10} = 0.8$	0.19
$0.68 \le x < 0.85$	$\frac{9}{10} = 0.9$	0.22
$x \ge 0.85$	1	0.15

So, what we are doing we had these values, so x less than 0.08; we did not get any observation. So, therefore that is 0 for between 0.8 to less than 0.16; we have got only one observation 0.08. Therefore, Fn x is equal to 0.1, point 16 to just less than point 21; we have got two observations point eight and point one six therefore, Fn x is equal to point 2. Similarly, for up to point 85, we

got 9 observations and at tenth observation in the ordered sample is point 85. Therefore, for x greater than equal to point 85, we get the value 1. Now, what is the value of the supremum?

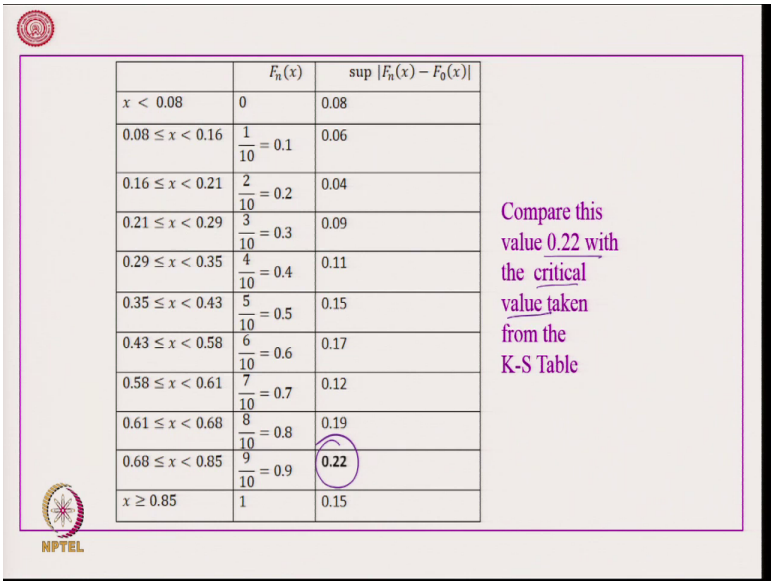
(Refer Slide Time: 57:03)



To do that let us first look at a graph. So, this is the graph that we have drawn for our understanding. So, this red line is the F0 x that is the cdf of uniform 0, 1. The blue line which is the sample Fn x. So, what is happening till point 8 it is 0, at point 8, we get one observation; therefore it goes up to point 1. Then the next observation is at point 16, therefore it gets a jump up to point 2 and like that the last one is at point 85 and that gives a jump up to this point that is 1.

Therefore, for each of this interval we can always look at what is the maximum of the difference for any x belonging to that interval. Say for example, in this interval the maximum difference coming out between the red line and the blue line is this much; that is the supremum.

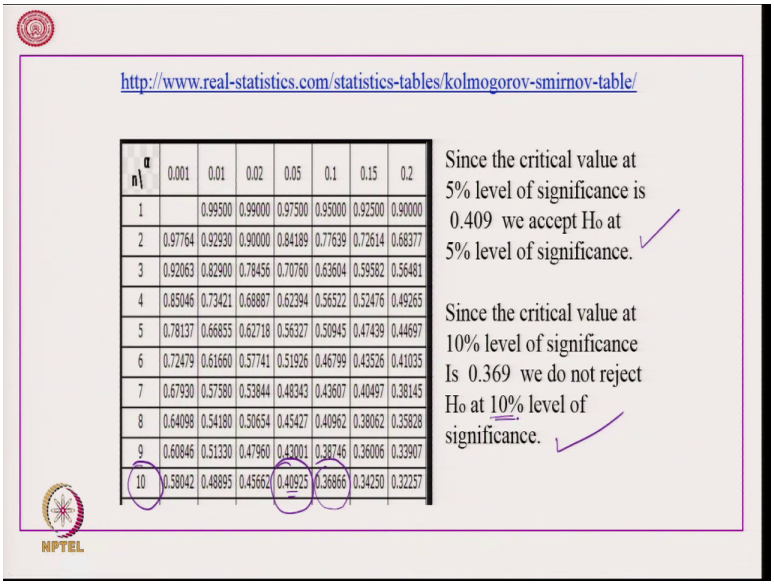
(Refer Slide Time: 58:24)



Therefore, we have computed those values when Fn x is 0 and x is less than 0.8; then the supremum is coming out to be 0.08. Because F0 x at 0.08 is equal to the same value F0 x is equal to x. Similarly, for point 08 to point 16, Fn x is point 1; therefore the supremum is coming out to be the difference between point 1 and point 16 that is point 016. Point 16 to point 21, the Fn x is 0.2; therefore the maximum difference is coming out is from point 16 and point 2 that is 0.04. Point 21 to point 29, the value is point 3; therefore the supremum of the difference is coming out to be point 21.

0.3 – 0.21 that is 0.09, like that we have computed all of them and we have got the maximum value to be 0.22. Now, if we look at the graph, we will find that the maximum value is coming here; this is the maximum difference which is coming out to be 0.22. So, the statistic for us is going to be that 0.22, therefore we need to compare this value with the critical value; that we can take from the Kolmogorov-Smirnov of table.

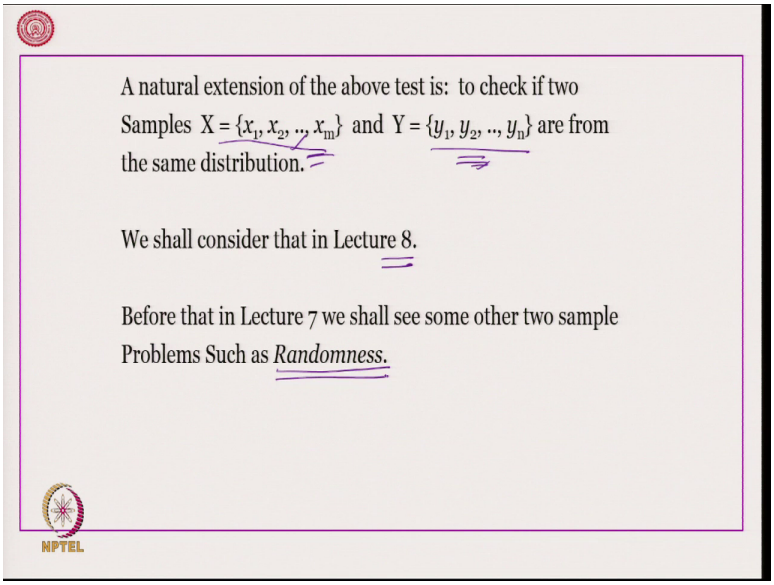
(Refer Slide Time: 60:08)



n \ α	0.001	0.01	0.02	0.05	0.1	0.15	0.2
1		0.99500	0.99000	0.97500	0.95000	0.92500	0.90000
2	0.97764	0.92930	0.90000	0.84189	0.77639	0.72614	0.68377
3	0.92063	0.82900	0.78456	0.70760	0.63604	0.59582	0.56481
4	0.85046	0.73421	0.68887	0.62394	0.56522	0.52476	0.49265
5	0.78137	0.66855	0.62718	0.56327	0.50945	0.47439	0.44697
6	0.72479	0.61660	0.57741	0.51926	0.46799	0.43526	0.41035
7	0.67930	0.57580	0.53844	0.48343	0.43607	0.40497	0.38145
8	0.64098	0.54180	0.50654	0.45427	0.40962	0.38062	0.35828
9	0.60846	0.51330	0.47960	0.43001	0.38746	0.36006	0.33907
10	0.58042	0.48895	0.45662	0.40925	0.36866	0.34250	0.32257

So, let us look at the table, we have value of n is equal to 10 at 5 percent level of significance; it is 0.409, we are looking at only 3 decimal spaces. At point 1, we are getting 0.369; so we are comparing our obtained value with these values. What we have obtained is 0.22 therefore, we accept the Ho at 5% level of significance; and also we cannot reject it at 10% level of significance. That means we are going to accept it also at 10 percent level of significance.

(Refer Slide Time: 60:52)



Now, a natural extension of the above test is to check, if we take 2 different samples x1, x2, xn; and y1, y2, yn are they coming from the same distribution. In that case what is going to happen?

We are not going to compare the sample frequency distribution with a standard hypothetical distribution. Rather we shall look at their sample cumulative functions for both x and y and we will check how much they differ. So, if the maximum value is greater than some threshold, then we are going to reject that they are coming from the same distribution; otherwise we are going to accept that.

So, that we will consider in lecture 8, but before that in the next lecture we shall see some other two sample tests. In particular our focus is going to on Randomness testing; that is whether the sample taken by us is actually random. Okay friends I stop here today, in the next class we shall do these problems. Till then thank you.