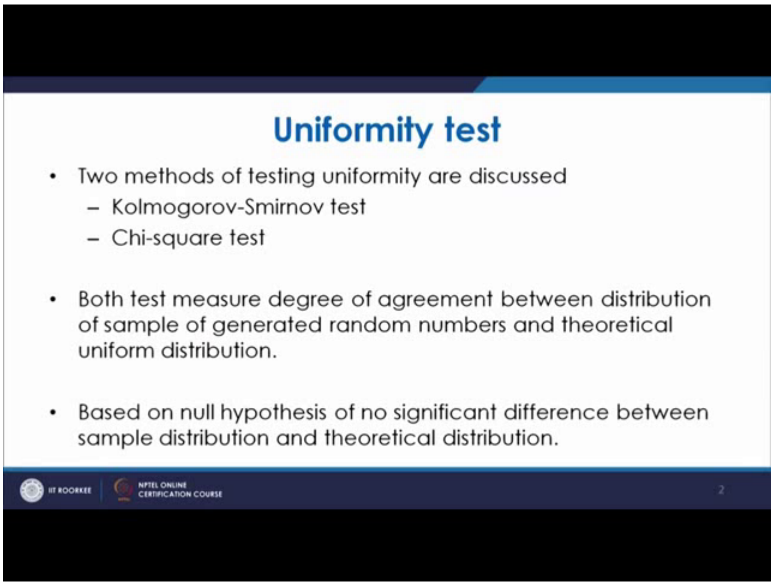
Modeling & Simulation of Discrete Event Systems
Dr. Pradeep K Jha
Department of Mechanical and Industrial Engineering
Indian Institute of Technology, Roorkee

Lecture – 18
Testing of Random Numbers

Welcome to the lecture on testing of random numbers. So, in the last lecture, we discussed about the random number generators and we discussed about the congruential generators and we had seen that how the random numbers of uniform distribution can be generated, but then it is also very important to see that these numbers should be tested for their uniformity and independence.

Because these are the 2 main criteria which a random number or the random numbers which are generated they must you know fulfill. So, the first test is the uniformity taste.

(Refer Slide Time: 01:06)



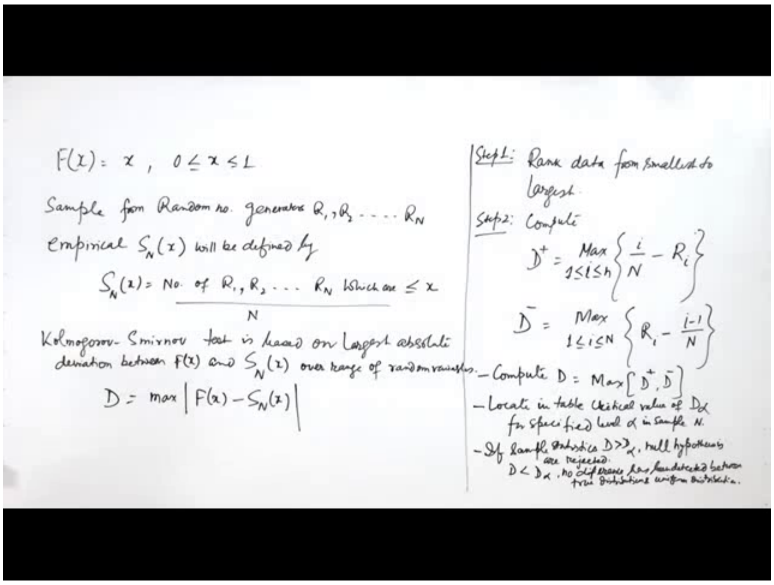
So, basically there are 2 methods of testing the uniformity and one is Kolmogorov Smirnov test and another is chi square test. So, both are the frequency test and as you know it is clear that these tests measure the degree of agreement between distribution of sample of generated random numbers I mean whatever we have generated. So, sample of generated random numbers and theoretical uniform distribution. So, basically you have to see that whatever random number you have generated they; how much they have a similarity with the theoretically uniform distribution.

So, based on the null hypothesis of no significant difference between the sample distribution and the theoretical distribution we come to the conclusion in between we have to find certain parameters we have to refer to some tables and from there we try to come to the conclusion that based on this hypothesis we can say that it cannot be rejected that they are not the uniform distribution. So, this is how these 2 tests are tested on the random numbers generated now coming to the Kolmogorov Smirnov test. So, what is the process of this Kolmogorov Smirnov test?

So, when you are given a sample of numbers suppose some 5 numbers or. So, or 10 numbers or. So, then based on that you are basically first of all arranging them in orders and then you are basically finding certain parameters D plus and D minus that we will see and then these are calculated based on these numbers and then they are compared against the tabular data which is basically for the ideal uniform distribution. So, for that particular value they are compared and then if satisfy the condition which is stated for the uniformity then we say that yes on the basis of that the hypothesis cannot be or the numbers can be said to be uniformly distributed out they are said to pass this uniformity test.

So, let us go to the principles of this Kolmogorov Smirnov test. So, what happens as we know that in the case of uniform distribution you have certain cdf that you know. So, in that F x will be x. So, for certain from up to one.

(Refer Slide Time: 04:02)



So, like for the uniform distribution what we see is we have x when x is 0 to 1. So, this is for definition of this uniform distribution that is cdf. Now what happens you have the sample from random number generators that is R 1, R 2, to R N. So, and then you have the empirical S N x. So, empirical S N x will be defined by S N x will be number of R 1 R 2 to R N which are less than or equal to x and divided by N.

So, what happens as N becomes larger you have the better approximation to the uniformity. So, if you the sample size is larger if you have large number of random numbers then the chances of getting the uniformity is higher now the Kolmogorov Smirnov test is based on largest deviation largest absolute deviation between f x that is your uniform cdf distribution cdf and S N x that which is given to you for checking that uniformity over that over range of random variables. So, for that we are finding certain parameters and these parameters are one is D equal to maximum of F x minus S N x.

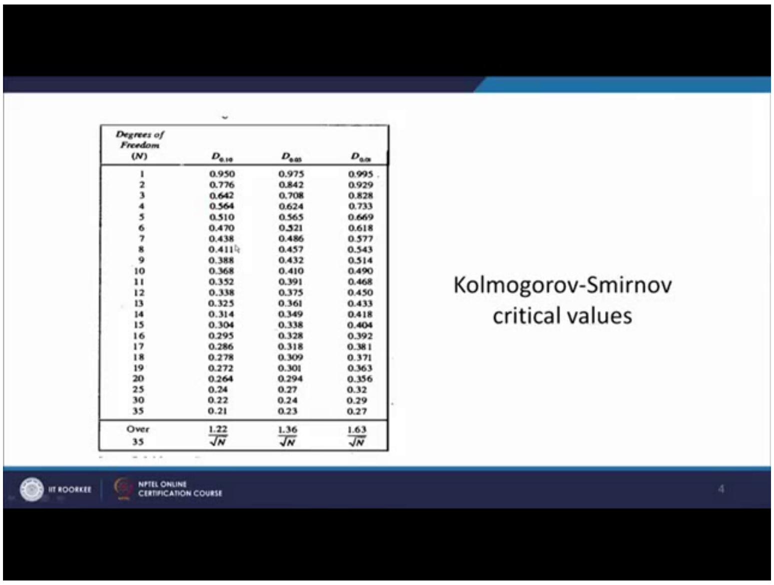
So, this D is basically calculated now the steps are like this the steps are step one.

So, you are given the data now what we do is you are ranking the data rank data from smallest to largest. So, first of all whichever data you are given you have to rank it from smallest to largest next you have to compute the parameters. So, compute; so, you will compute D plus which is defined as maximum when i is more than equal to one and then less than equal to n. So, it will be i upon N minus R i and then you have D minus that will be maximum of again i between 1 to N and it will be R i minus i minus 1 by n.

So, these 2 parameters you are going to find I mean I will vary from 1 to n. So, you will be having you will be given certain number of you know values like you are given 10 values in that case the capital N is 10. Now I will start from one to ten. So, for every number you are calculating D plus as well as D minus then next you are computing D equal to maximum of D plus and D minus. So, once you compute D plus and D minus you are going to have the value of d. So, maximum of these 2 then locate in table critical value of D alpha for a specified level alpha in sample N you we have already discussed in the last class that for such test one alpha is given one level is given which gives that probability I mean of rejection system, it must be our probability of telling that yes it is passing the uniformity test.

So, in that case alpha is given normally and normally it is taken from 0.01 to 0.05 as we had discussed. So, that value will be given to you and for this you have the table. So, ks table is given from that table you have to locate the value of D alpha and then if sample statistics D is more than D alpha null hypothesis are rejected. So, if the D is found to be more than D alpha the null hypothesis tells that the sample which has been tested it is not having the uniform distribution and if b is less than D alpha in that case we say that no difference has been detected between true distribution and uniform distribution. So, what it tells that you are getting this D which is nothing, but a maximum of the D plus and D minus and that has to be less than this D alpha that is from the table that uniform distribution table you will get this D alpha.

(Refer Slide Time: 12:30)

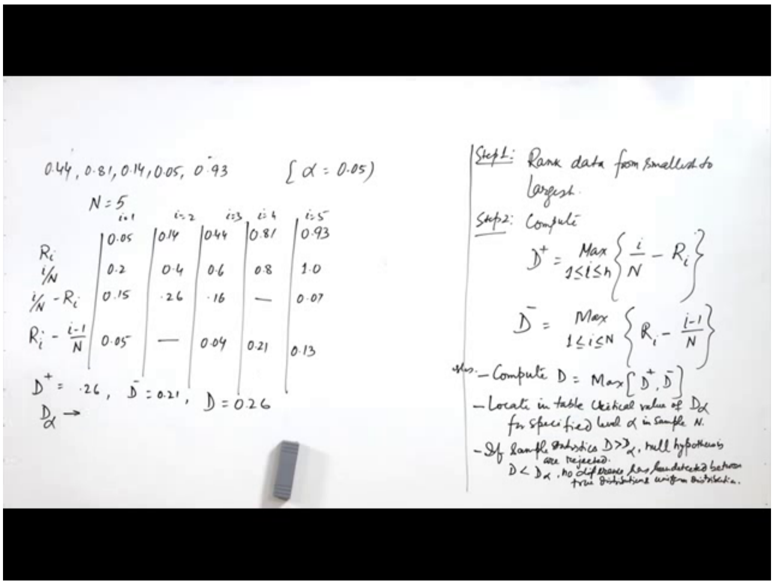


Degrees of Freedom (N)	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.828
4	0.564	0.624	0.733
5	0.510	0.565	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.410	0.490
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	0.325	0.361	0.433
14	0.314	0.349	0.418
15	0.304	0.338	0.404
16	0.295	0.328	0.392
17	0.286	0.318	0.381
18	0.278	0.309	0.371
19	0.272	0.301	0.363
20	0.264	0.294	0.356
25	0.24	0.27	0.32
30	0.22	0.24	0.29
35	0.21	0.23	0.27
Over 35	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

Now, this D alpha for a specified alpha you will have the table like this. So, in this table if you have N number of data you have the degree of freedom that is N minus 1. So, you will see against this degree of freedom value if suppose you have 10 data your degree of freedom is 9 and then I mean according to this table in front of 9, you have to see; what is the alpha value; what is the value of alpha which has been set. So, for that alpha you are going to see the value. So, suppose D alpha is 0.05. So, for 10, it will be against 9.432 or for 4 5 samples it will be against 4.6 to 4 like that and this D alpha should be more than this D.

So, in that case the sample statistics D and this if b is less than this D alpha in that case we say that there is no conclude evidence that it is not uniformly distributed. So, it is basically said to be uniformly distributed now we can check by checking the 5 random digits and then we can say we can see how this test is being carried out for any number of digits. So, if there are suppose 5 digits.

(Refer Slide Time: 13:57)



There are 5 digits given; 0.44 and then 0.81, 0.14, 0.05 and 0.93; suppose you have these 5 random numbers given and you are told to see and ensured that whether they are uniformly distributed in that case what we see here is we have and also you are given alpha value as 0.05. So, you can very well check it here you see N is 5 you have 5 data. So, N is 5.

Now, if you see you have the value R I, then you will have value i by upon i upon N you have i upon N minus R i and R i minus i minus 1 upon N that is what these parameters are required you to be you know calculated now for every i. So, you have one value is R i first numbers is point. So, first of all you are going to take them in a order in the increasing order. So, increasing order the first number will come as 0.05, second number will come 0.14, third number will come as 0.44, fourth number will come as 0.81 and fifth number will come as 0.93, then i upon N i is 1, 2, 3, 4, 5.

So, i is 1, i is 2, i is 3, i is 4 and i is 5. So, i upon N; N is 5 anyway. So, it will be 0.2, 0.6, 0.8 and 1, then i by N upon minus R i. So, i by N minus R i will be this minus this. So, that will be 0.15, this will be 0.4, minus 0.14, 0.26, this will be 0.14; 16; this will be minus 0.01. So, minus value need not be take and again this will also be minus no this is will not be minus will be 0.07, then R i minus i minus 1 upon N. So, R i minus i minus 1 upon N. So, that will be 0. So, this will be 0.05.

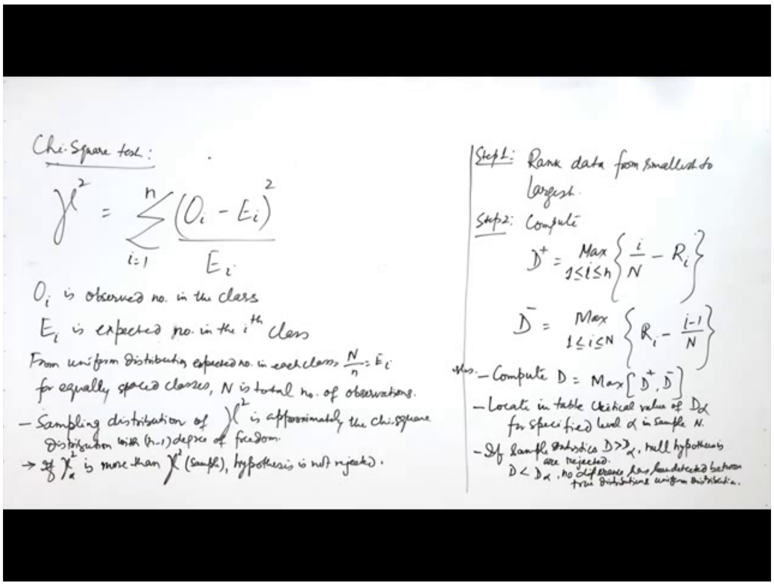
Now, i will be 2. So, R i minus; so, 0.4 minus i minus 1 upon n. So, that will be 0.05. So, that will be 0.14 minus i minus 1, 2 minus 1; one by this is 5. So, that is 0.2. So, it will be 0.14 minus 0.2. So, it is a negative number. So, we are not going to write it. Similarly 0.44 minus this will be 0.4. So, it will be 0, 4.81 minus 0.6; it will be 0.21 and this will be 0.93 minus 0.8. So, it will be 0.13. Now what we see is this is D plus. So, D plus is D plus will be maximum of i by N minus R i. So, maximum of this will be 0.26 and D minus will be maximum of R i minus i minus 1 by N. So, maximum is 0.21. So, sample statistics say that D will be equal to maximum of these 2 values that is 0.26.

Now, we have to compare it or we have to you know see that how it is faring whether it is more or less as compared to D alpha. So, D alpha we have to calculate now for D alpha calculation we must know that degree of. So, from this table we will try to see what is the degree of freedom; so, as the number of values of 5. So, degree of freedom is 4 degree freedom is N minus 1. So, against 4 against 4 we will see D 0.05 value. So, that value is given as. So, D alpha that is 0 point sorry. So, this D alpha it is given as; so, in the D alpha basically degree of area. So, this value is given 0.624. So, we have to compare against D alpha. Now in this case D alpha will be computed for m equal to 5 and the alpha is 0.05.

So, in this case we are in the case of Kolmogorov Smirnov test we are looking directly against 5 and then for the D alpha of 0.05 when we talk about the different tests in the case of chi square test there, we need to have the degree of freedom value and there we see the degree of freedom has N minus 1. So, in this case what we see is we have the D alpha value D alpha value will be taken as against 5 you see D alpha is 0.05. So, that is 0.565. Now what we see is this is 0.565 and this is D is 0.26. So, this 0.26 is less than 0.565 that is D alpha that is why we can say that this null hypothesis tells that you cannot reject this hypothesis that these numbers are from the uniform distribution or they are not having uniformity.

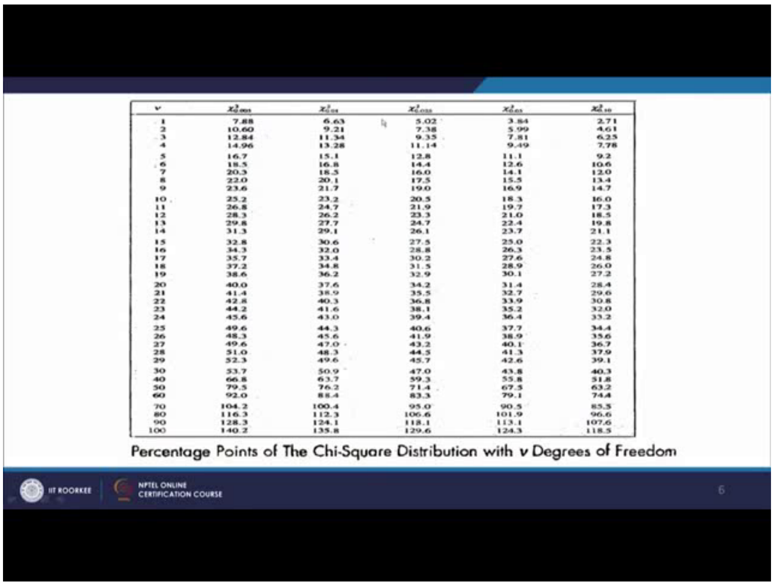
So, it means they are passing this hypothesis test of uniformity. So, this is how this Kolmogorov test; Smirnov test is carried out you may have many number of samples and as you see you have as we go more and more you can you see that you this get for any number you can have these values next type of test is the chi square test. So, in the chi square test again you find certain parameter and based on that you are checking the uniformity of the sample. So, in the chi square test what is done is you have a parameter chi square.

(Refer Slide Time: 21:11)



So, this chi square is defined as. So, this chi square test. So, this square this is defined as summation of i equal to 1 to n and then O i minus E i square upon E i.

(Refer Slide Time: 21:51).



v	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$
1	7.88	6.63	5.02	3.84	2.71
2	10.60	9.21	7.38	5.99	4.61
3	12.84	11.34	9.35	7.81	6.25
4	14.96	13.28	11.14	9.49	7.78
5	16.7	15.1	12.8	11.1	9.2
6	18.5	16.8	14.4	12.6	10.6
7	20.3	18.5	16.0	14.1	12.0
8	22.0	20.1	17.5	15.5	13.4
9	23.6	21.7	19.0	16.9	14.7
10	25.2	23.2	20.5	18.3	16.0
11	26.8	24.7	21.9	19.7	17.3
12	28.3	26.2	23.3	21.0	18.5
13	29.8	27.7	24.7	22.4	19.8
14	31.3	29.1	26.1	23.7	21.1
15	32.8	30.6	27.5	25.0	22.3
16	34.3	32.0	28.8	26.3	23.5
17	35.7	33.4	30.2	27.6	24.8
18	37.2	34.8	31.5	28.9	26.0
19	38.6	36.2	32.9	30.1	27.2
20	40.0	37.6	34.2	31.4	28.4
21	41.4	38.9	35.5	32.7	29.6
22	42.8	40.3	36.8	33.9	30.8
23	44.2	41.6	38.1	35.2	32.0
24	45.6	43.0	39.4	36.4	33.2
25	49.6	44.3	40.6	37.7	34.4
26	48.3	45.6	41.9	38.9	35.6
27	49.6	47.0	43.2	40.1	36.7
28	51.0	48.3	44.5	41.3	37.9
29	52.3	49.6	45.7	42.6	39.1
30	53.7	50.9	47.0	43.8	40.3
40	66.8	63.7	59.3	55.8	51.8
50	79.5	76.2	71.4	67.5	63.2
60	92.0	88.4	83.3	79.1	74.4
70	104.2	100.4	95.0	90.5	85.5
80	116.3	112.3	106.6	101.9	96.6
90	128.3	124.1	118.1	113.1	107.6
100	140.2	135.8	129.6	124.3	118.5

Percentage Points of The Chi-Square Distribution with v Degrees of Freedom

So, you have these values this chi square value is basically calculated and again this chi square value is compared against this standard you know table. So, in this table you have this mu that is degree of freedom. So, from here in this case the degree of freedom becomes the number of i minus 1. So, in that case you have against that you have to just see the chi square value and again same condition is applied to find the values. So, what is done is O i is observed number in the class E i is expected number in the class expected number in the ith class. So, from uniform distribution expected number in each class should be capital N upon small n that is E i. Now we have for equally spaced classes N is total number of observations.

So, what you do is in this case you have the; suppose there are many numbers which are given and you have to test its uniformity. So, you are finding the intervals and you are finding the frequencies and you know that you have in the case of uniform distribution what should be the frequency suppose you have a number and you have you have put these numbers in 10 intervals. So, in that case the interval every interval must have ideal frequency of 10 percent I mean 10 itself; if suppose there are 100 samples. So, every interval should have the frequency of 10 in the case of uniform distribution. Now when we have any sample in that sample if you are putting in the interval that will not necessarily be 10, it may be 18 or it may be 12, it may be 9 or 11 or it may be 10 even in some cases.

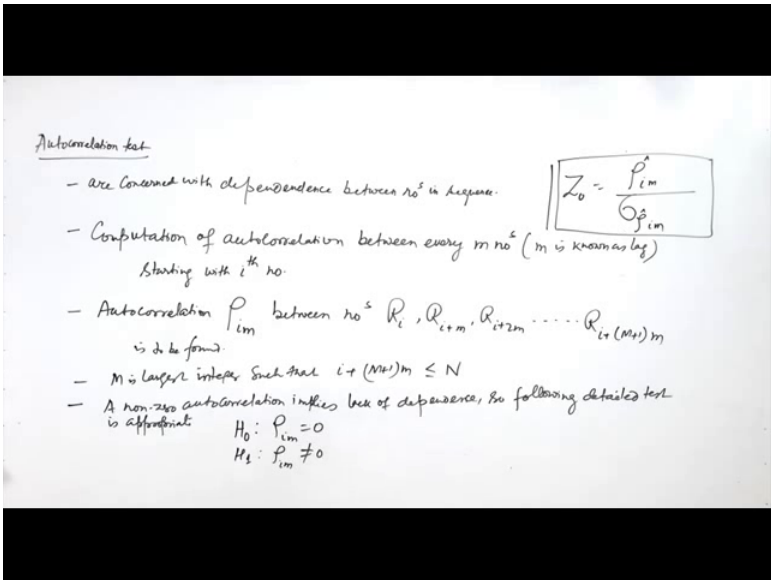
Now, in those cases all these O i and E i we are basically going to see we. So, once you have that you are going to find O i minus E i square and then that will be divided by E i. So, you will get chi square and sampling distribution of chi square is approximately the chi square distribution with N minus 1 degree of freedom. So, it is assumed that it is with N minus 1 degree of freedom. Now from there once you get this chi square value, this chi square value will be again compared with this value from this chi square distribution with a new degree of freedom now against the number of intervals. Now you are going to have the degree of freedom that is N minus 1 and against that for a particular alpha value you are going to find the chi square value.

Now, if that chi square if chi square value for a particular alpha. So, that will be if this is more than chi square of sample hypothesis is not rejected. So, you will have the chi square value from this table this value should be more than what we calculate from here if it is. So, in that case we say that I had that null hypothesis is not rejected we say that the samples for the uniform distribution if it is not. So, in that case it is will be said to be you know it will said that it may not be of uniform distribution. So, this way we can solve the examples in the coming classes and then we will see that how they are solved and how we come to find these values.

Next is the autocorrelation test. So, these 2 tests which we have seen they are basically related to the checking of uniformity. Now we will see that how they can we can come to the conclusion that they are not correlated. So, it should be seen that they are not correlated it may be. So, happening when we get this data we can observe that there is certain type of correlation between the numbers, if we find that then in that case it is more predictable. So, then that is not a completely independent type of random numbers. So, basically these test is done to see that the numbers we generate they are independent. Now in that that case you have the autocorrelation test that test tells how to check that there is no correlation.

So, for that a parameter is set and this parameter again is calculated and then based on that we are going to decide.

(Refer Slide Time: 28:20)



So, as we see this autocorrelation test this test basically are concerned with dependence between numbers in sequence. So, whatever number you may have we have it will tell us that whether there is any dependence between the number by looking at the number itself many a times we have the notion that it looks to have certain correlation may be that every fifth number every third number or after third number the first third number and then after that after every gap of 5 numbers, we see that there is a type of pattern in the numbers.

In fact, if that is known that is not going to be the example of very much an independent sample. So, that type of test is carried out using this autocorrelation test. So, computation of autocorrelation between every m numbers m is known as lag starting with ith number. So, you will have certain number of random numbers certain number of numbers which are random and whose correlation test is to be carried out and in that you are said that among these numbers whether you check whether there is any correlation or. So, so it will start with the ith number maybe 1 2 3 or. So, and from there we are seeing that between every m numbers. So, means i, i plus m i plus 2 m i plus 3 m like that. So, you have different numbers and among them if you see that there is any you try to find whether there is any correlation you try to see that.

Now, in that; so, what you do is autocorrelation that is rho i m rho i m tells the auto correlation between the number which is starting at ith number and with m as lag. So, say that is i m. So, i plus m i i plus m i plus 2 m or. So, so between numbers R i R i plus m R i plus 2 m and so it will be R i plus m plus 1 into m. Now it has to go up to m plus 1 into m is to be found. So, now, for this now this has to go maximum up to N because you are we are going to have maximum of capital N number of numbers. So, it has to go up to that. So, in that case what you get is m is largest integer such that i plus m plus 1 m is less than equal to capital M.

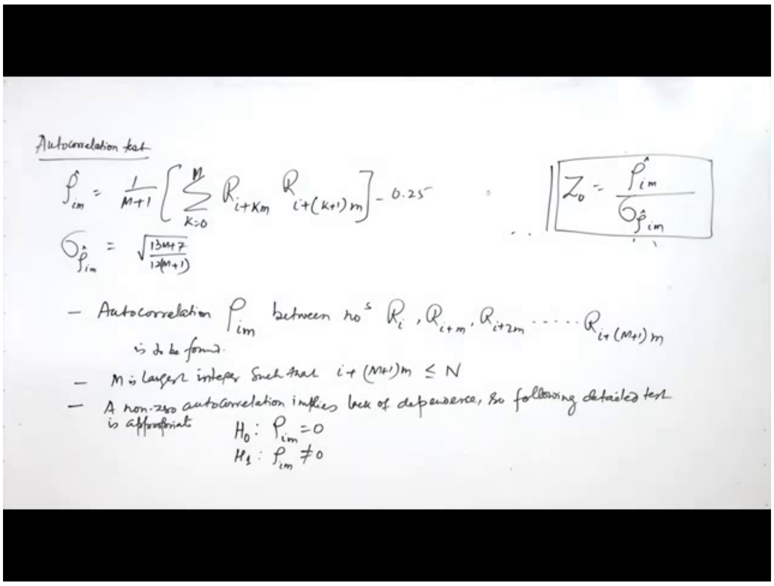
So, you have N number of samples which are to be basically tested. So, now, a nonzero autocorrelation implies lack of dependence. So, following detailed test is appropriate now we have a condition on which you have to say whether this is there is correlation or not. So, for that again you have h hypothesis. So, where rho i m is 0. So, you are telling that there is no correlation otherwise you have you are telling that since it is not equal to 0 it means they are correlated.

So, what you do is in this case now you have R i: R i plus m, R i plus 2 m or. So, in that case you further find certain statistics. So, among them you are finding the statistics z naught that is the test statistics that will be rho i m estimator value divided by sigma of rho i m. Now whatever rho i m you are getting. So, you are getting this sigma i m and then based on that you are computing it through the uniform distribution table z naught. So, this it will be the table which is given, yeah.

(Refer Slide Time: 34:10)

You have this z naught given for a particular alpha you will have be you will be given this z naught and once you know that you have to compare from there.

(Refer Slide Time: 34:35)



So, further you will have certain conditions and certain parameters to be found. So, what you see is rho i m you will be computing by 1 by m plus 1 summation of k equal to 0 to m R i plus k m and R i plus k plus 1 m. So, and minus 0.2 5 and sigma rho i m that will be again divided decided by 13 m plus 7 divided by 12 m plus 12 into m plus 1. So, that this you value you get you have to compare against the standard normal distribution table we have it mean 0 and variance 1 and then you are going to compare it and if you are getting this value to be lesser, then you say that they are not correlated you can accept this hypothesis and otherwise you are going to reject it. So, we can have some problems we can solve in the lectures of this tutorial classes which will be may be after one class or more and then we can see that how the problems of correlation can be solved and checked.

Thank you very much.