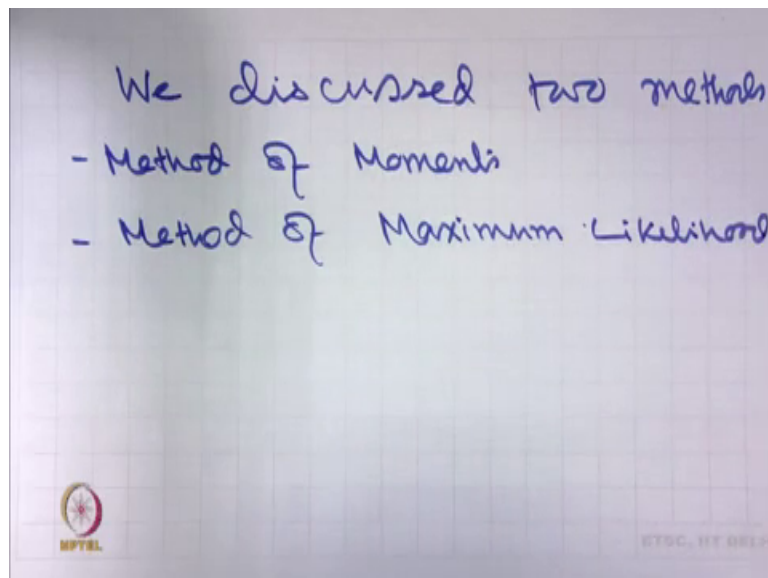


**Statistical Inference**  
**Prof. Niladri Chatterjee**  
**Department of Mathematics**  
**Indian Institute of Technology, Delhi**

**Lecture - 18**  
**Statistical Inference**

Welcome students to the MOOC's lecture series on Statistical Inference. In the last lecture, we have talked about point estimation of the parameters of the underlying distribution; from which a sample  $x_1, x_2, \dots, x_n$  is taken. In particular, we have discussed 2 methods.

(Refer Slide Time: 00:49)

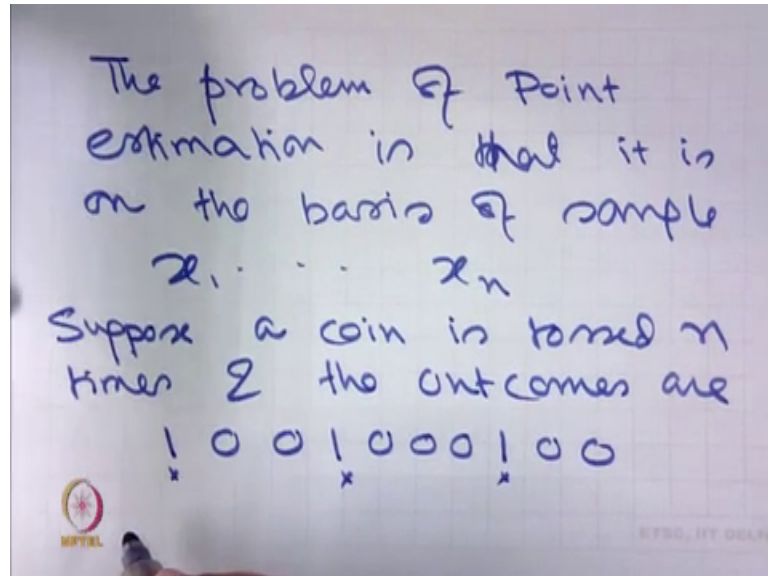


Method of moments and method of maximum likelihood; in method of moments what we have done? We have calculated sample raw moments from the sample  $x_1, x_2, \dots, x_n$  and we have equated that with the theoretical moments as given by the distribution under consideration. And by equating them we have formed equations and we obtained the estimate for different parameters by solving those equations. On the other hand, the philosophy of maximum likelihood is slightly different.

Here we try to differentiate the likelihood function of  $x_1, x_2, \dots, x_n$  under the parameter  $\theta$  by taking its derivative or partial derivative with respect to  $\theta$ , and equating it to 0, so that by solving for  $\theta$  we get a function of  $x_1, x_2, \dots, x_n$  which maximizes the likelihood function. Or in other words, which maximizes the probability of obtaining the

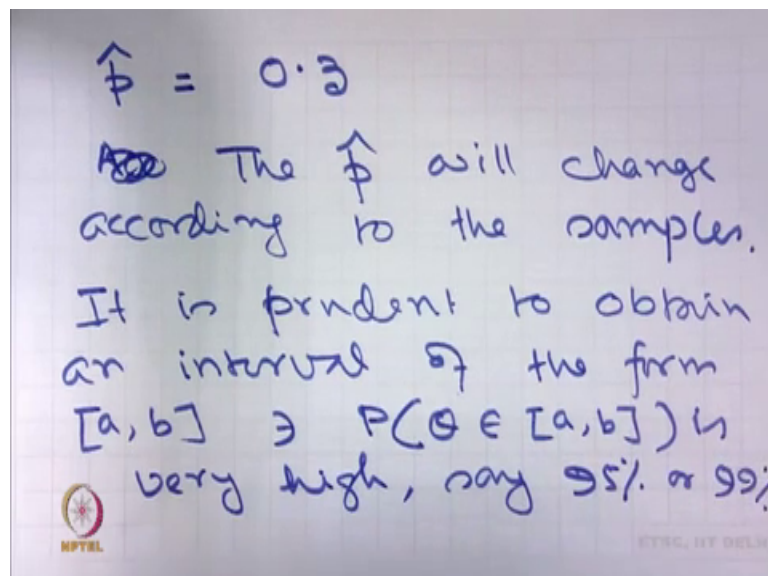
sample  $x_1, x_2, \dots, x_n$  and there theta for which that is maximized is considered to be the maximum likelihood estimated.

(Refer Slide Time: 02:51)



However, the problem of point estimation is that it is on the basis of sample  $x_1, x_2, \dots, x_n$ . And based on that we are trying to estimate the value of the parameter theta; as I have already shown you an example suppose a coin is tossed  $n$  times and the outcomes are say 1 0 0 1 0 0 0 1 0 0; that is there are 3 heads and remaining tails. Based on that, if we try to estimate  $p$  then  $\hat{p}$  comes out to be say 0.3.

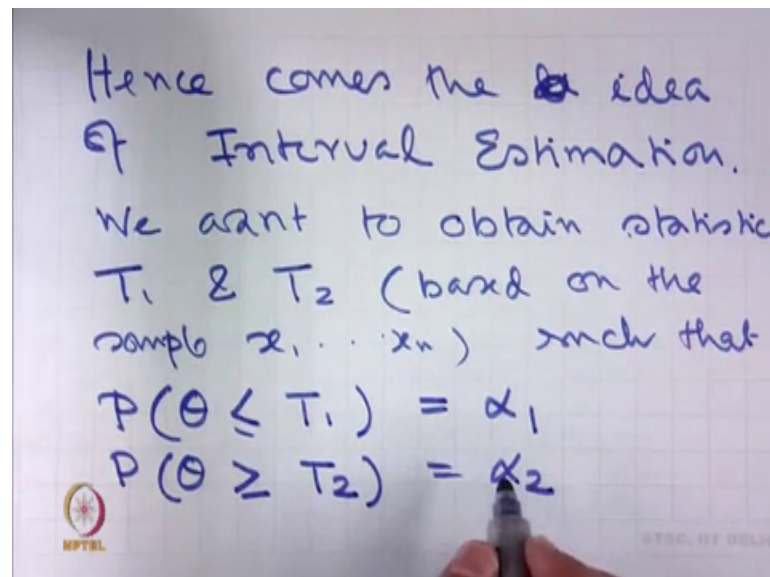
(Refer Slide Time: 04:15)



But suppose we make another experiment of the same coin, there may be 4 heads or there may be 5 heads. And accordingly the  $\hat{p}$  will change according to the number of samples.

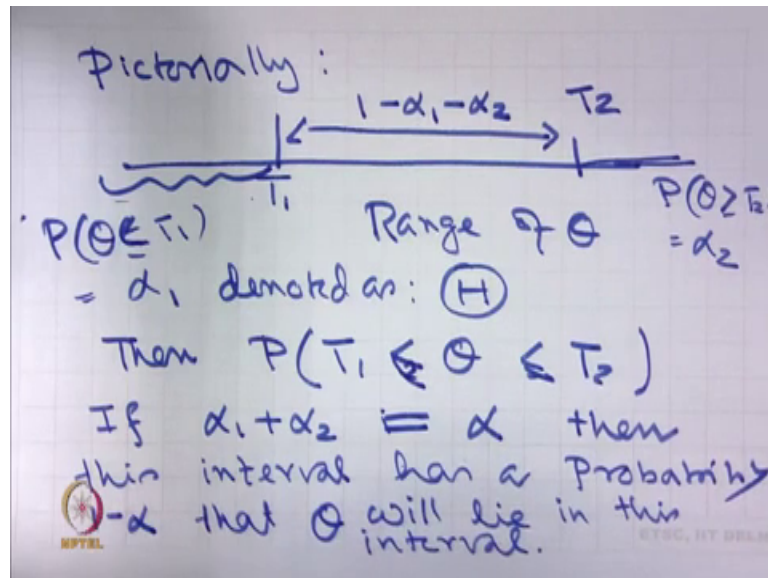
Hence, it is prudent to obtain an interval of the form say  $a$  to  $b$ ; such that probability  $\theta$  belongs to  $a$  to  $b$  is very high, say 95 percent or 99 percent. What is the advantage? The advantage is that we are very confident that based on the sample we can see that  $\theta$  is going to belong to this interval with a very high probability. That assurance we cannot give with respect to point estimation. Hence comes the idea of interval estimation.

(Refer Slide Time: 06:03)



So, what we do here? We want to obtain statistics  $T_1$  and  $T_2$  based on the sample  $x_1, x_2, \dots, x_n$  such that probability  $\theta$  less than equal to  $T_1$  is equal to say  $\alpha_1$ , and probability  $\theta$  greater than equal to  $T_2$  is equal to  $\alpha_2$ .

(Refer Slide Time: 07:25)

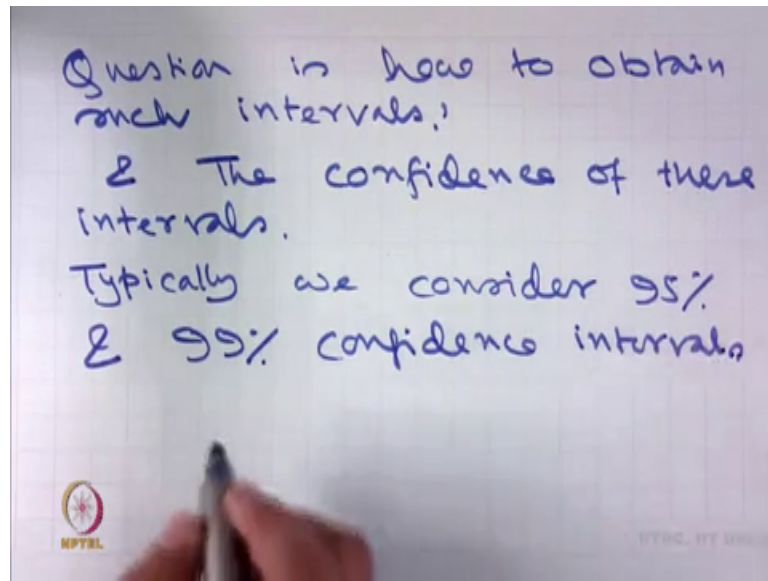


So, pictorially suppose, this is the range of theta; which we have already denoted as capital theta. Depending upon the distribution and the parameter, the theta will change, but suppose we have  $T_1$  here.

Such that probability theta belongs to this interval, or theta less than equal to  $T_1$  is equal to  $\alpha_1$ . And suppose this is the  $T_2$  such that probability theta greater than equal to  $T_2$  is  $\alpha_2$ .

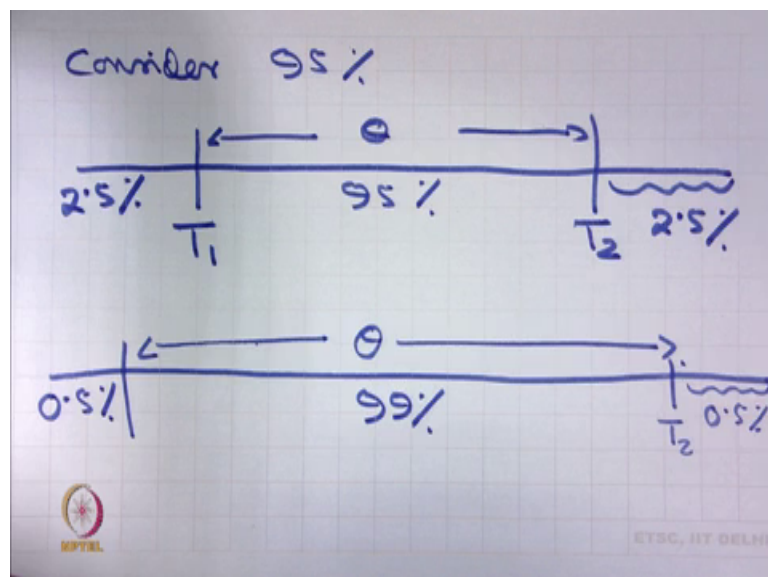
Then probability  $T_1$  less than equal to theta less than equal to  $T_2$  or say let them be strict inequality, it does not matter because it is a continuous case and this probability is therefore, going to be  $1 - \alpha_1 - \alpha_2$ . If  $\alpha_1 + \alpha_2$  is equal to  $\alpha$ , then this interval has a probability say  $1 - \alpha$ , that theta will lie in this interval.

(Refer Slide Time: 09:53)



Question is how to obtain this interval and the confidence of these intervals. By confidence we mean, the probability that the parameter will remain within that interval  $T_1$  and  $T_2$ . Typically, we consider 95 percent and 99 percent confidence intervals.

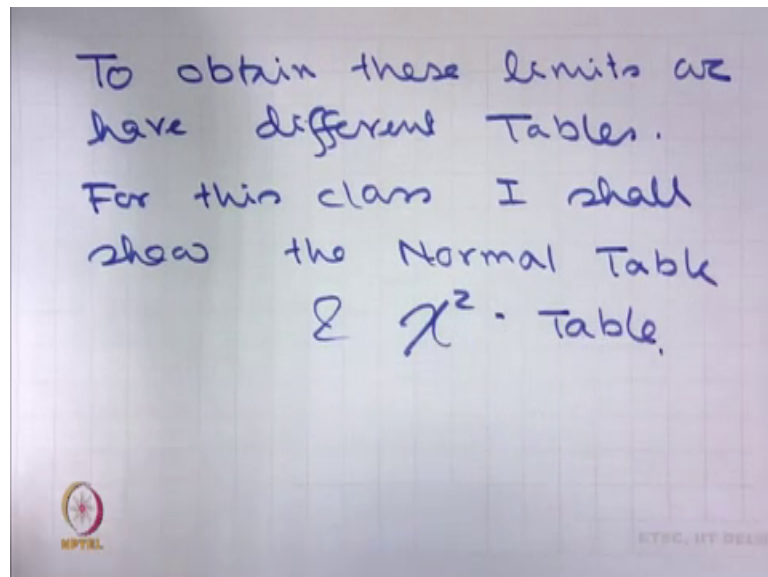
(Refer Slide Time: 11:09)



Consider 95 percent. Suppose this is the range of  $\theta$ , we want  $T_2$  such that in this there is 2.5 percent chance that the variable will occur on this side of  $T_2$ , and say  $T_1$  such that on this side also there is 2.5 percent chance that the  $\theta$  will occur in this region. So, the 95 percent confidence interval for  $\theta$  is this.

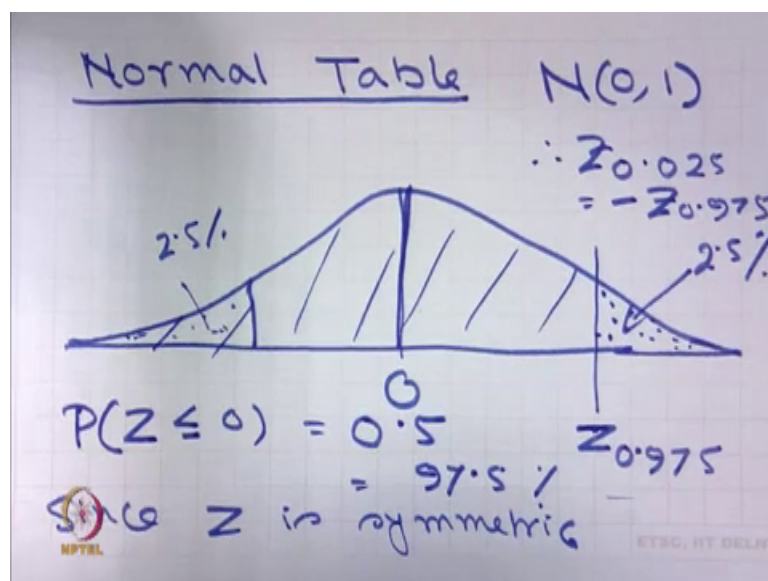
If it is 99 percent, then we look at a T 2 such that the occurrence here is only 0.5 percent and the occurrence probability of here is 0.5 percent. Therefore, that theta will occurred in this interval has the chance 99 percent. So, these are that confidence intervals that we try to obtain for the parameter theta or with the help of the sample  $x_1, x_2, \dots, x_n$ .

(Refer Slide Time: 12:50)



To obtain this limits we have different tables, for this class I shall show the normal table and chi square table. So, let us first understand how to see the normal table.

(Refer Slide Time: 13:44)



We know the shape of the normal curve is somewhat like this consider this is 0. Let us just look at normal 0 1, because for any other normal with mu and sigma square, we can convert it to the standard normal distribution. We know that probability z which is standard normal less than equal to 0 is equal to 0.5. So, for 95 percent we will look at the value say call it z 0.975.

Because the area below this is equal to 97.5 percent; that is this area is 2.5 percent. And since z is symmetric, this value such that below this is 2.5 percent is going to be minus of this. Therefore, z of 0.025 is equal to minus of z of 0.975. So, there are tables where these values are tabulated. First let me show you the normal table.

(Refer Slide Time: 16:27)

**DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z**

1	.02	.03	.04	.05	.06	.07	.08
.99	.50798	.51197	.51595	.51994	.52392	.52790	.53188
.80	.54776	.55172	.55567	.55962	.56356	.56749	.57142
.17	.58706	.59095	.59483	.59871	.60257	.60642	.61026
.72	.62552	.62930	.63307	.63683	.64058	.64431	.64803
.10	.66276	.66640	.67003	.67364	.67724	.68082	.68439
.97	.69847	.70194	.70540	.70884	.71226	.71566	.71904
.07	.73237	.73565	.73891	.74215	.74537	.74857	.75175
.15	.76424	.76730	.77035	.77337	.77637	.77935	.78230
.03	.79389	.79673	.79955	.80234	.80511	.80785	.81057
.59	.82121	.82381	.82639	.82894	.83147	.83398	.83646
.75	.84614	.84849	.85083	.85314	.85543	.85769	.85993
.50	.86864	.87076	.87286	.87493	.87698	.87900	.88100

So, if you look at that table, it is showing standard normal distribution table values represent area to the left of the Z score, as you can see that.

(Refer Slide Time: 16:52)

.5983	.54380	.54776	.55172	.55567	.55962	.56356	.5674
.7926	.58317	.58706	.59095	.59483	.59871	.60257	.6064
.791	.62172	.62552	.62930	.63307	.63683	.64058	.6443
.542	.65910	.66276	.66640	.67003	.67364	.67724	.6808
.0146	.69497	.69847	.70194	.70540	.70884	.71226	.7156
.2575	.72907	.73237	.73565	.73891	.74215	.74537	.7485
.804	.76115	.76424	.76730	.77035	.77337	.77637	.7793
.814	.79103	.79389	.79673	.79955	.80234	.80511	.8078
.594	.81859	.82121	.82381	.82639	.82894	.83147	.8339
.134	.84375	.84614	.84849	.85083	.85314	.85543	.8576
.433	.86650	.86864	.87076	.87286	.87493	.87698	.8790
.493	.88686	.88877	.89065	.89251	.89435	.89617	.8979
.320	.90490	.90658	.90824	.90988	.91149	.91309	.9146
.924	.92073	.92220	.92364	.92507	.92647	.92785	.9292
.319	.93448	.93574	.93699	.93822	.93943	.94062	.9417
.520	.94630	.94738	.94845	.94950	.95053	.95154	.9525
.543	.95637	.95728	.95818	.95907	.95994	.96080	.9616
.407	.96485	.96562	.96638	.96712	.96784	.96856	.9692

So, in this table there are all the values tabulated. Let us see the value corresponding to 0.975.

(Refer Slide Time: 17:02)

.94845	.94950	.95053	.95154	.95254	.95352
.95818	.95907	.95994	.96080	.96164	.96246
.96638	.96712	.96784	.96856	.96926	.96995
.97320	.97381	.97441	.97500	.97558	.97615
.97882	.97932	.97982	.98030	.98077	.98124
.98341	.98382	.98422	.98461	.98500	.98537
.98713	.98745	.98778	.98809	.98840	.98870
.99010	.99036	.99061	.99086	.99111	.99134
.99245	.99266	.99286	.99305	.99324	.99343
.99430	.99446	.99461	.99477	.99492	.99506
.99573	.99585	.99598	.99609	.99621	.99632
.99683	.99693	.99702	.99711	.99720	.99728
.99767	.99774	.99781	.99788	.99795	.99801
.99831	.99836	.99841	.99846	.99851	.99856
.99878	.99882	.99886	.99889	.99893	.99896

Consider this value, this is 0.975. So, we have to obtain the x value for the value on the real line such that probability a standard normal variate less than equal to this is 0.975. How to obtain this? We first look at the left side of this table and we can see the value given is 1.96, 1.9 as you can see.



(Refer Slide Time: 17:38)

0.7	.75804	.76115	.76424	.76730	.77035
0.8	.78814	.79103	.79389	.79673	.79955
0.9	.81594	.81859	.82121	.82381	.82639
1.0	.84134	.84375	.84614	.84849	.85083
1.1	.86433	.86650	.86864	.87076	.87286
1.2	.88493	.88686	.88877	.89065	.89251
1.3	.90320	.90490	.90658	.90824	.90988
1.4	.91924	.92073	.92220	.92364	.92507
1.5	.93319	.93448	.93574	.93699	.93822
1.6	.94520	.94630	.94738	.94845	.94950
1.7	.95543	.95637	.95728	.95818	.95907
1.8	.96407	.96485	.96562	.96638	.96712
1.9	.97128	.97193	.97257	.97320	.97381
2.0	.97725	.97778	.97831	.97882	.97932
2.1	.98214	.98257	.98300	.98341	.98382
2.2	.98610	.98645	.98679	.98713	.98745

So, that is 1.9, and for the second decimal place we go above and we can find 6.

(Refer Slide Time: 17:56)

TABLE VALUES REPRESENT AREA TO THE LEFT OF THE Z SCORE					
.03	.04	.05	.06	.07	.08
.51197	.51595	.51994	.52392	.52790	.53188
.55172	.55567	.55962	.56356	.56749	.57142
.59095	.59483	.59871	.60257	.60642	.61026
.62930	.63307	.63683	.64058	.64431	.64803
.66640	.67003	.67364	.67724	.68082	.68439
.70194	.70540	.70884	.71226	.71566	.71904
.73565	.73891	.74215	.74537	.74857	.75175
.76730	.77035	.77337	.77637	.77935	.78230
.79673	.79955	.80234	.80511	.80785	.81057
.82381	.82639	.82894	.83147	.83398	.83646
.84849	.85083	.85314	.85543	.85769	.85993
.87076	.87286	.87493	.87698	.87900	.88100
.89065	.89251	.89435	.89617	.89796	.89973
.90824	.90988	.91149	.91309	.91466	.91621

Therefore we get 1.96. In a similar way, suppose we want 99 percent confidence interval. Therefore, what we will be looking at? We will be looking at a value such that below that the probability of occurrence is going to be 0.995. So, again we look at the table, and we find that this is the value which is 0.99492 and if we go further we get 0.99506.

Therefore, 0.995 we would expect is somewhere at the middle. Now, if I go along the row, very slowly we can see that the value corresponding to this is 2.5. As I am dragging the my pen slowly you can see that the values are coming out to be 0.99492 and 0.99506.

So, if we go above, we can find the values are here along this column are 0.7 and 0.8. So, we can say 2.58 is the value such that a standard normal distribution will have a value less than 2.58 has the chance 0.995, what has the probability 0.995 or it is 99.5 percent chance that the standard normal value will obtain a will have a value between below or a standard normal variate will have a value below 2.58. Therefore, we need to remember at least these 2 values.

(Refer Slide Time: 21:01)

	1%	5%
Both sides	2.58	1.96

99% confidence interval in  $[-2.58, 2.58]$

95% confidence interval in  $[-1.96, 1.96]$

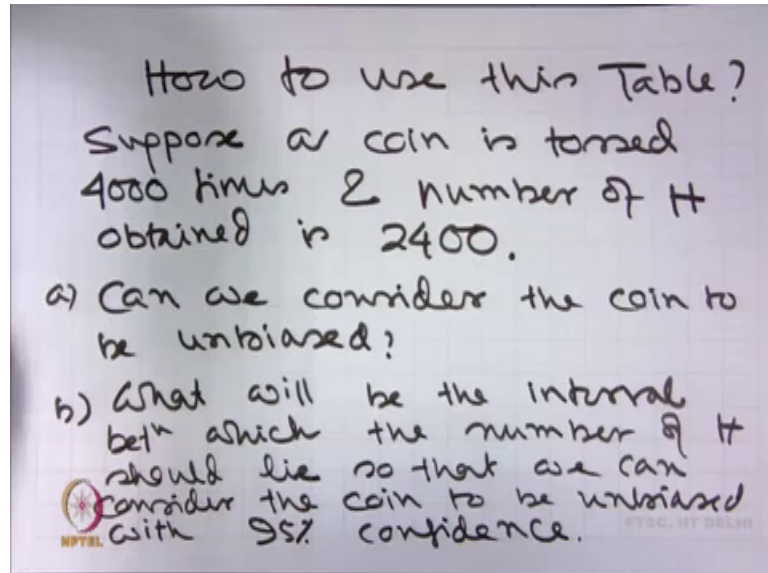
So, you remember this table, for standard normal, I write it both sided for 1 percent the value is 2.58 as we have just observed. Above 2.58 the chance is only 0.5 percent. And below minus 2.58 the chance is only 0.5 percent so, together we have 1 percent.

Or the 99 percent confidence interval is minus 2.58 to 2.58. Similarly, when we are looking at say 5 percent here then this value we have just seen is 1.96 or 95 percent confidence interval is minus 1.96 to 1.96.

So, these 2 values you need to remember because typically these are the values that we use. Of course, there are many other values as you have seen in the table it is full with values. So, for normal for different level of confidence we can actually obtain the value

from the table, but for this class we restrict ourselves altitude of 2 cases; both sided interval and 95 percent confidence and 99 percent confidence. How to use this table?

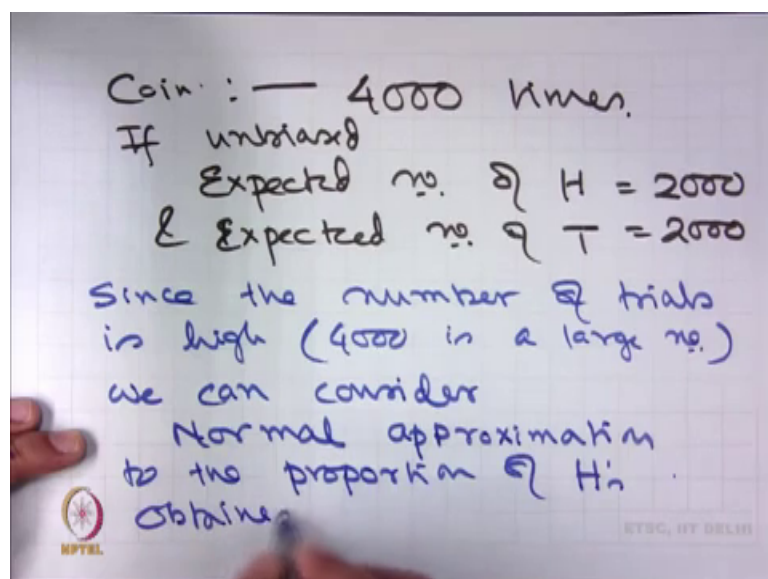
(Refer Slide Time: 23:46)



So, consider this following problem. Suppose a coin is tossed 4,000 times and number of heads obtained is 2,400. Question is can we consider the coin to be unbiased.

And second question, what will be the interval between which the number of heads should lie so that we can consider the coin to be unbiased with 95 percent confidence. So, let me explain the problem again.

(Refer Slide Time: 26:11)



So, there is a coin, we tossed it 4,000 times. If it is unbiased, expected number of heads is equal to 2,000 and expected number of tail is also 2,000. But it does not actually mean that in the result we will get exactly 2,000 heads and exactly 2,000 tails, that will not happen in generally.

So, the what you want to estimate it? That what is the value of  $p$  or the probability of success or the probability of getting a head in a toss given this result? Okay so, we solve it in the following way. Since, the number of trials is high that is 4,000 is a large number, we can consider normal approximation to the proportion of heads obtained.

(Refer Slide Time: 28:22)

Let  $P$  be the r.v denoting the proportion of  $H$ 's.

$$\therefore E(P) = \frac{1}{2} \quad V(P) = \frac{pq}{n}$$
$$= \frac{\frac{1}{2} \times \frac{1}{2}}{4000}$$
$$\therefore \frac{P - \frac{1}{2}}{\sqrt{\frac{\frac{1}{2} \times \frac{1}{2}}{4000}}} \text{ may be considered as } N(0, 1).$$

Logo of IIT Delhi is visible in the bottom left corner of the whiteboard.

Let  $p$  be the random variable denoting the proportion of heads. Therefore, expected value of  $P$  is equal to half, and variance of  $P$  as we know is equal to  $pq$  by  $n$  is equal to half into half divided by 4,000. Therefore,  $P$  minus half upon root over half into half upon 4,000 may be considered as normal 0 1.

(Refer Slide Time: 29:40)

Since we obtained 2400 H's  
 $\therefore$  Obtained proportion =  
 $\frac{2400}{4000} = \frac{6}{10} = \frac{3}{5}$   
 $\therefore \frac{\frac{3}{5} - \frac{1}{2}}{\sqrt{\frac{\frac{1}{2} \times \frac{1}{2}}{4000}}} \sim N(0, 1)$   
Denominator =  $\frac{1}{2} \times \sqrt{\frac{1}{4000}}$   
 $= \frac{1}{2} \times \frac{1}{20} \times \frac{1}{\sqrt{10}}$

Since, we obtained 2,400 number of heads therefore, obtained proportion is equal to 2,400 divided by 4,000 is equal to 6 upon 10 is equal to 3 upon 5. Therefore, 3 upon 5 minus half divided by root over half into half into 4,000 can be assumed to be normal 0 1.

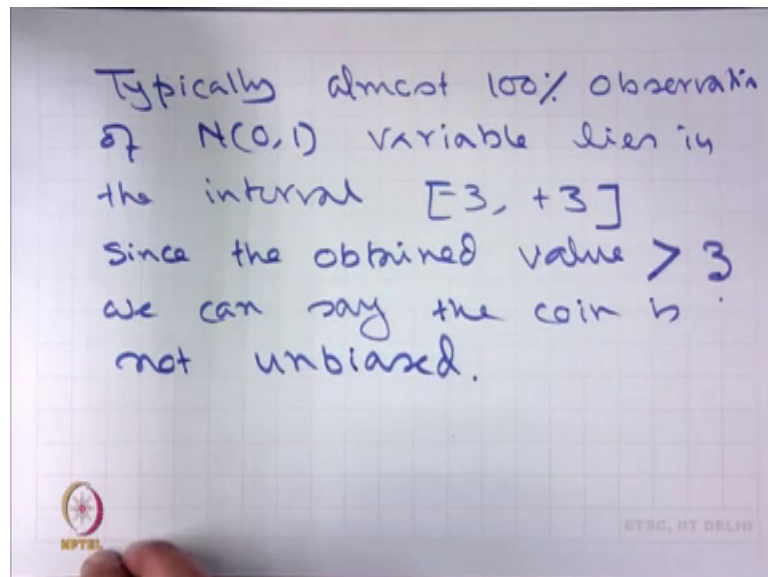
Now this denominator is half into root over 1 upon 4,000 is equal to half into 1 by 20 into 1 upon root 10 is equal to 1 upon 40 into 1 upon root 10.

(Refer Slide Time: 31:11)

$= \frac{1}{40} \times \frac{1}{\sqrt{10}}$   
 $\therefore$  The obtained value of the  $N(0, 1)$  variable is  
 $\frac{\frac{3}{5} - \frac{1}{2}}{\frac{1}{40} \times \frac{1}{\sqrt{10}}} = \frac{\frac{6-5}{10}}{\frac{1}{40} \times \frac{1}{\sqrt{10}}}$   
 $= \frac{1}{10} \times 40 \times \sqrt{10} = 4 \times \sqrt{10} > 3$

Therefore, the obtained value of the normal 0 1 variable is  $3 \text{ by } 5 \text{ minus } \frac{1}{2} \text{ upon } 1 \text{ by } 40 \text{ into } 1 \text{ by } \sqrt{10}$  is equal to  $6 \text{ minus } 5 \text{ upon } 10 \text{ divided by } 1 \text{ upon } 40 \text{ into } 1 \text{ upon } \sqrt{10}$ ; is equal to  $1 \text{ upon } 10 \text{ into } 40 \text{ into } \sqrt{10}$ , which is equal to  $4 \text{ into } \sqrt{10}$  and which is much greater than 3.

(Refer Slide Time: 32:36)



Typically, almost 100 percent observation of normal 0 1 variable lies in the interval minus 3 to plus 3. Since, the obtained value is much greater than 3 we can say the coin is not unbiased. If the coin were unbiased, then it is the value obtained should lie in the interval minus 3 to plus 3.

The second part of the question says that what should be the number of heads? Part b is what should be the number of heads to consider the coin to be unbiased with 95 percent confidence.

(Refer Slide Time: 33:59)

(b) What should be the number of H's to consider the coin to be unbiased with 95% confidence.

# of Heads  $\sim \text{Bin}(n, p)$

$\therefore$  Here  $n = 4000$

$p = \frac{1}{2}$

$\therefore E(\# \text{ of H's}) = \frac{N}{2} = 2000$

$V(nPq) = 4000 \times \frac{1}{2} \times \frac{1}{2} = 1000$

We know number of heads is follow binomial  $n, p$ . So, here  $n$  is equal to 4,000,  $p$  is equal to half. Therefore expected number of heads is equal to  $N$  by 2 is equal to 2,000, and variance is equal to  $npq$  is equal to 4000 into half into half is equal to 1,000.

(Refer Slide Time: 35:37)

$\therefore \frac{|N - 2000|}{\sqrt{1000}} \leq 1.96$

If  $N$  is the no. of H's, the

$\frac{|N - 2000|}{10\sqrt{10}} \leq 1.96$

as we are looking for 95% confidence.

Therefore,  $N$ , minus 2,000 divided by root over 1,000 has to be less than equal to 1.96, this we have already obtained as we are looking for 95 percent confidence. Or in other words, if  $N$  is the number of heads, then modulus of  $N$  minus 2,000 divided by 10 root 10 has to be less than equal to 1.96.

(Refer Slide Time: 36:46)

or  $|N - 2000| \leq 1.96 \times 10\sqrt{10}$   
or  $2000 - 1.96 \times 10\sqrt{10} \leq N \leq 2000 + 1.96 \times 10\sqrt{10}$   
Now  $1.96 \times 10\sqrt{10} \approx 62$ .  
 $\therefore$  The 95% confidence interval for Number of  $H$ 's to consider the coin to be unbiased is  $[2000 - 62, 2000 + 62]$   
 $= [1938, 2062]$

NPTL ETDC, IIT DELHI

Or modulus of  $N$  minus 2,000 is less than equal to 1.96 into 10 root 10 or 2,000 minus 1.96 into 10 root 10 has to be less than equal to  $N$ , less than equal to 2,000 plus 1.96 into 10 root 10. Now 1.96 into 10 root 10 is approximately 62.

Therefore, the 95 percent confidence interval for number of heads to consider the coin to be unbiased is 2,000 minus 62 to 2,000 plus 62 is equal to 1938 comma 2,062. Or in other words, if we toss the coin 4,000 times and the number of heads is obtained in this range between these 2 values, then we are 95 percent confident that this coin is unbiased. Let us look at another problem.



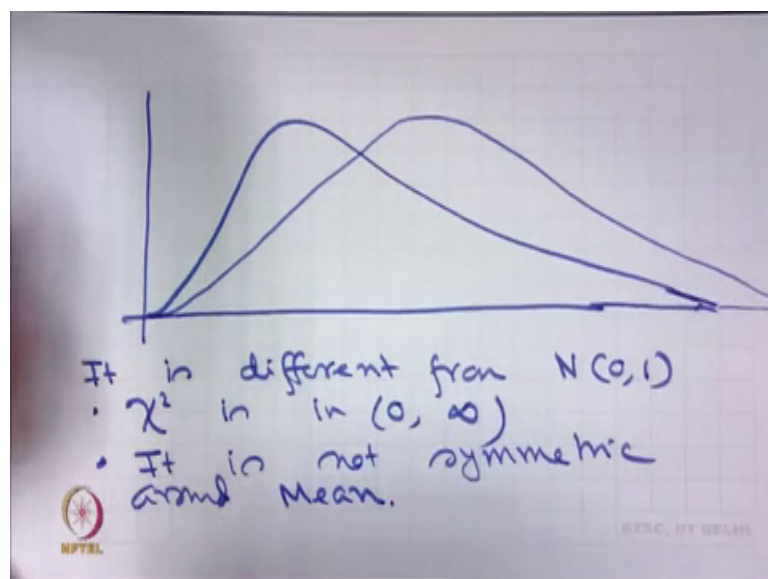
(Refer Slide Time: 39:11)

Ex Consider  $N(\mu, \sigma^2)$   
And we want to find confidence interval for  $\sigma^2$   
Let us consider the case that  $\mu$  is known.  
In that case:  
$$\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$$
  
where  $x_1, \dots, x_n$  in the sample obtained

Consider normal  $\mu$  sigma square and we want to find confidence interval for sigma square. Let us consider the case that  $\mu$  is known.

In that case,  $\sum_{i=1}^n (x_i - \mu)^2 / \sigma^2$  is equal to 1 to  $n$  divided by sigma square. We know that, it is a chi square with  $n$  degrees of freedom where  $x_1, x_2, \dots, x_n$  is the sample obtained.

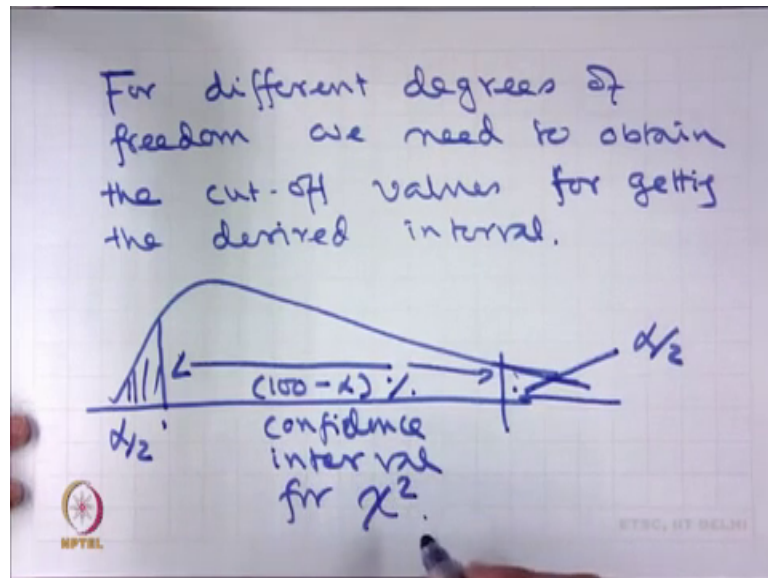
(Refer Slide Time: 40:55)



In general, the chi square distribution will have a shape like this. So, it is different from normal  $0, 1$  as chi square is in  $0$  to infinity. It is a positive random variable therefore; it

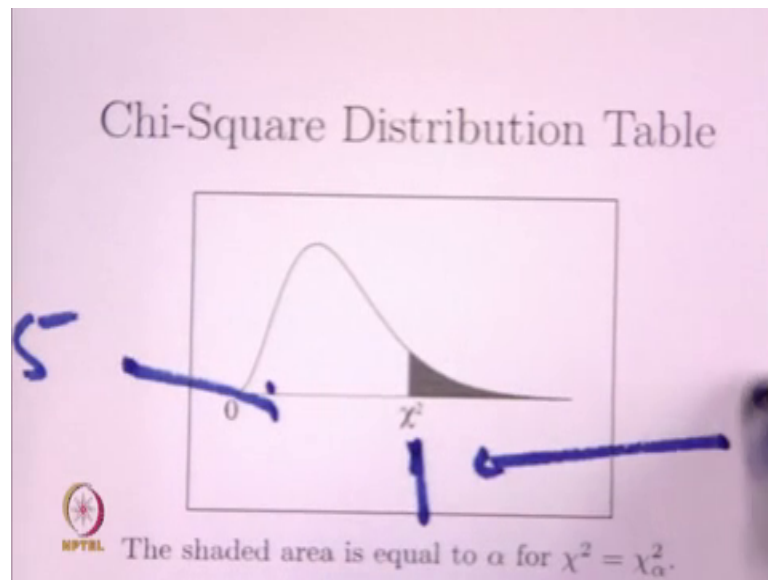
ranges over 0 to infinity. And secondly, it is not symmetric around mean. Also the shape of the chi square changes with that degrees of freedom, as degrees of freedom increases our chi-square may take a shape like this.

(Refer Slide Time: 42:12)



Therefore for different degrees of freedom, we need to obtain the cut off values for getting the desired interval. Therefore, we will be looking at a value here such that this probability is alpha by 2 will be of looking at a value here, such that this probability is also alpha by 2. So, that this is the 100 minus alpha percent confidence interval for chi square. So, let me show the chi square table.

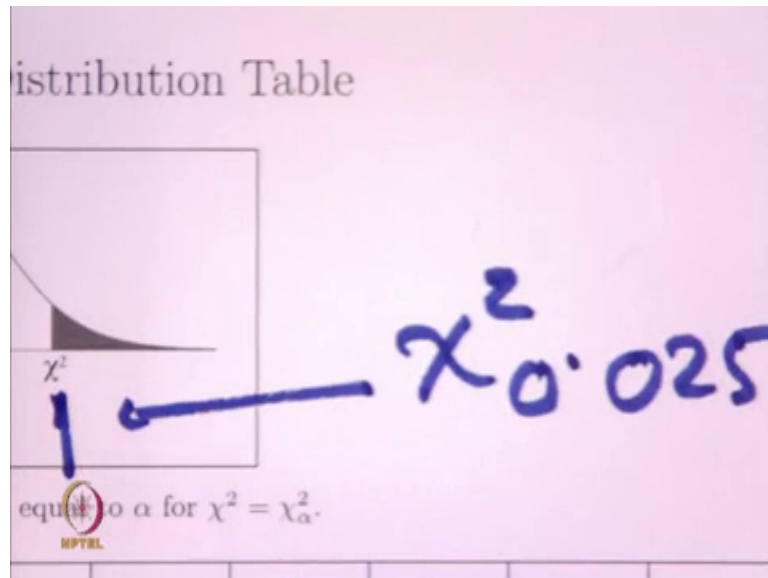
(Refer Slide Time: 43:46)



So, if you look at you can see that it is a chi square distribution table. The values are tabulated for different degrees of freedom and, the values are given for this point such that the shaded area is equal to alpha.

That means the probability the chi square distribution is greater than this value, that probability is alpha. And therefore, when we are looking at a 95 percent confidence interval, we look at this value such that the above this the probability is only 2.5 percent. Also we look at a value here such that, above that the probability is 97.5 percent. Or in decimal form, this value will give corresponding to chi square with 0.975. 0.975 and for this, we are looking at the value for chi square with 0.025.

(Refer Slide Time: 45:22)



How do we get the values?

(Refer Slide Time: 45:31)

$df$	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$
1	0.000	0.000	0.001	0.004
2	0.010	0.020	0.051	0.103
3	0.072	0.115	0.216	0.352
4	0.207	0.297	0.484	0.711
5	0.412	0.554	0.831	1.145
6	0.676	0.872	1.237	1.635
7	0.989	1.239	1.690	2.167
8	1.344	1.646	2.180	2.733
9	1.735	2.088	2.700	3.325
10	2.156	2.558	3.247	3.940
11	2.603	3.053	3.816	4.575

If we look at that table, we will find that the values are given for different degrees of freedom; 1, 2, 3, 4 and here you can see the values tabulated are for 0.975, and if I go further to the right, we can see the values are tabulated for 0.025.

(Refer Slide Time: 45:58)

shaded area is equal to  $\alpha$  for  $\chi^2 = \chi^2_{\alpha}$ .

$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$
0.004	0.016	2.706	3.841	5.024
0.103	0.211	4.605	5.991	7.378
0.352	0.584	6.251	7.815	9.348
0.711	1.064	7.779	9.488	11.143
1.145	1.610	9.236	11.070	12.833
1.635	2.204	10.645	12.592	14.449
2.167	2.833	12.017	14.067	16.013
2.733	3.490	13.362	15.507	17.535
3.325	4.168	14.684	16.919	19.023
3.940	4.865	15.987	18.307	20.483
4.575	5.578	17.275	19.675	21.920

(Refer Slide Time: 46:05)

to  $\alpha$  for  $\chi^2 = \chi^2_{\alpha}$ .

$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
2.706	3.841	5.024	6.635	7.879
4.605	5.991	7.378	8.910	10.597
6.251	7.815	9.348	10.445	12.838
7.779	9.488	11.143	11.977	14.860
9.236	11.070	12.833	13.566	16.750
10.645	12.592	14.449	15.202	18.548
12.017	14.067	16.013	16.816	20.278
13.362	15.507	17.535	18.409	21.955

In fact, it is given for 0.005, 0.010, like that for a fixed set of alpha the values have been computed. In particular, let us look at chi square with 100 degrees of freedom.

(Refer Slide Time: 46:27)

25	10.520	11.524	13.120	14.611
26	11.160	12.198	13.844	15.379
27	11.808	12.879	14.573	16.151
28	12.461	13.565	15.308	16.928
29	13.121	14.256	16.047	17.708
30	13.787	14.953	16.791	18.493
40	20.707	22.164	24.433	26.509
50	27.991	29.707	32.357	34.764
60	35.534	37.485	40.482	43.188
70	43.275	45.442	48.758	51.739
80	51.172	53.540	57.153	60.391
90	59.196	61.754	65.647	69.126
100	67.328	70.065	74.222	77.929

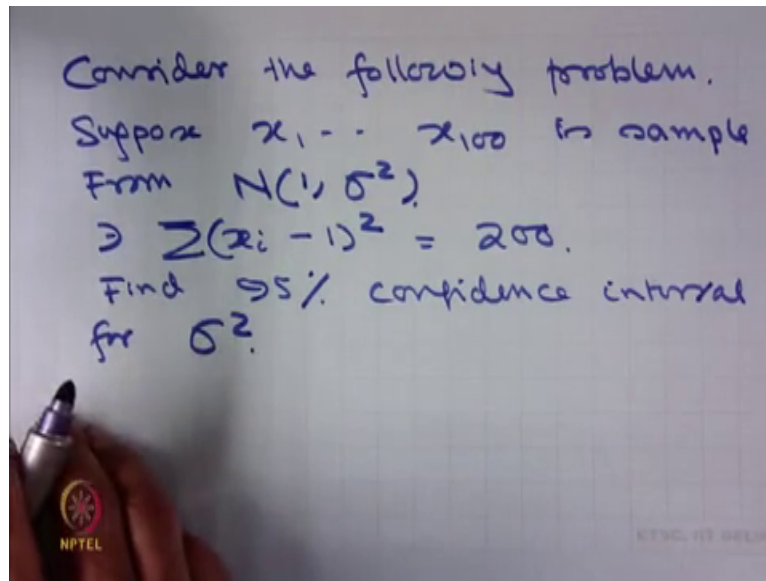
Look at the value it is 100 degrees of freedom, and corresponding to 0.975, this value is 74.22; that means, a chi square 100 degrees of freedom will take a value greater than equal to 74.222 with probability 9.975 or 97.5 cases out of 100.

(Refer Slide Time: 47:09)

13.848	15.659	33.196	36.415	39.364
14.611	16.473	34.382	37.652	40.646
15.379	17.292	35.563	38.885	41.923
16.151	18.114	36.741	40.113	43.195
16.928	18.939	37.916	41.337	44.461
17.708	19.768	39.087	42.557	45.722
18.493	20.599	40.256	43.773	46.979
26.509	29.051	51.805	55.758	59.342
34.764	37.689	63.167	67.505	71.420
43.188	46.459	74.397	79.082	83.298
51.739	55.329	85.527	90.531	95.023
60.391	64.278	96.578	101.879	106.629
69.126	73.291	107.565	113.145	118.136
77.929	82.358	118.498	124.342	129.561

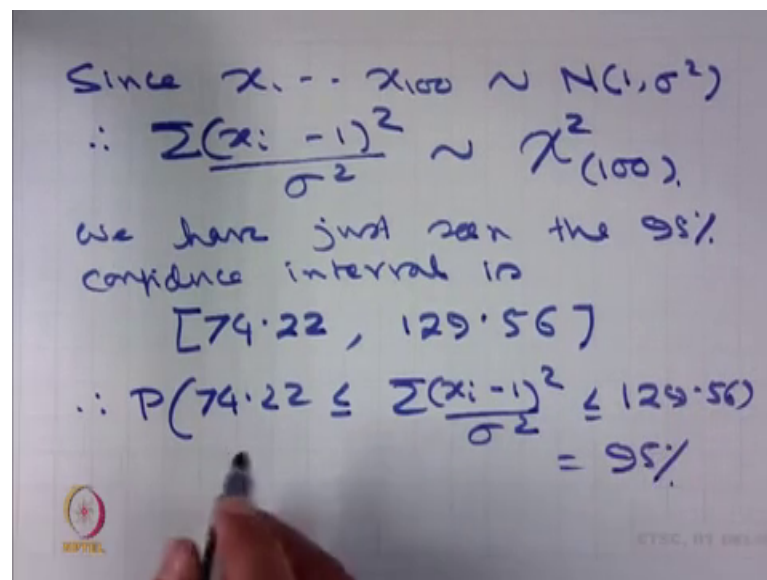
And, if we go further, we can see that for 0 to 5 the values are 129.561. So, let us consider one problem.

(Refer Slide Time: 47:31)



Suppose  $x_1, x_2, \dots, x_{100}$  is a sample from normal  $1$  comma  $\sigma^2$ ; such that  $\sum (x_i - 1)^2 = 200$ . Find 95 percent confidence interval for  $\sigma^2$ .

(Refer Slide Time: 48:53)



Since,  $x_1, x_2, \dots, x_{100}$  are from normal with  $1$  comma  $\sigma^2$ ; therefore,  $\sum (x_i - 1)^2$  upon  $\sigma^2$  is chi square with 100 degrees of freedom. We have just seen the 95 percent confidence interval is 74.22 to 129.56. Therefore,

probability  $\frac{1}{74.22}$  less than equal to  $\frac{\sigma^2}{\sum(x_i - \bar{x})^2} \leq \frac{1}{129.56}$  is equal to 95 percent.

(Refer Slide Time: 50:19)

The image shows a handwritten derivation for a 95% confidence interval for the population variance  $\sigma^2$ . The steps are as follows:

$$\text{or } P\left(\frac{1}{129.56} \leq \frac{\sigma^2}{\sum(x_i - \bar{x})^2} \leq \frac{1}{74.22}\right)$$

$$\text{or } P\left(\frac{\sum(x_i - \bar{x})^2}{129.56} \leq \sigma^2 \leq \frac{\sum(x_i - \bar{x})^2}{74.22}\right) = 95\%$$

$$\text{or } P\left(\frac{200}{129.56} \leq \sigma^2 \leq \frac{200}{74.22}\right) = 95\%$$

$$\text{or } P(1.54 \leq \sigma^2 \leq 2.69) = 95\%$$

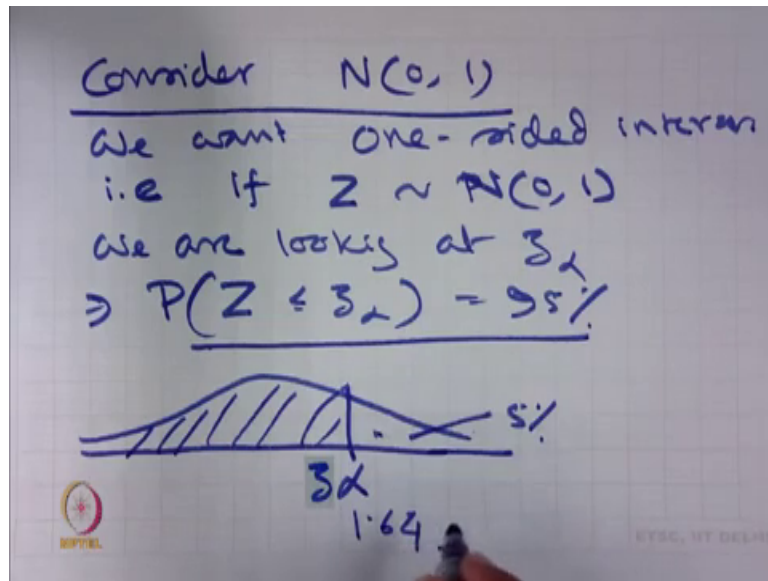
At the bottom, it states: 95% confidence interval = 95% for  $\sigma^2 = [1.54, 2.69]$ .

Or probability  $\frac{1}{129.56}$  less than equal to  $\frac{\sigma^2}{\sum(x_i - \bar{x})^2} \leq \frac{1}{74.22}$  is equal to 95 percent, or probability  $\frac{\sum(x_i - \bar{x})^2}{129.56} \leq \sigma^2 \leq \frac{\sum(x_i - \bar{x})^2}{74.22}$  is equal to 95 percent, or probability  $\frac{200}{129.56} \leq \sigma^2 \leq \frac{200}{74.22}$  is equal to 95 percent, or probability  $1.54 \leq \sigma^2 \leq 2.69$  is equal to 95 percent.

Since, this is given to be 200 or probability  $\frac{200}{129.56} \leq \sigma^2 \leq \frac{200}{74.22}$  is equal to 95 percent. Or probability  $1.54 \leq \sigma^2 \leq 2.69$  is equal to 95 percent. Therefore, 95 percent confidence interval for  $\sigma^2$  is equal to 1.54 comma 2.69. So, the problems that we solved, they are all both sided we have taken.



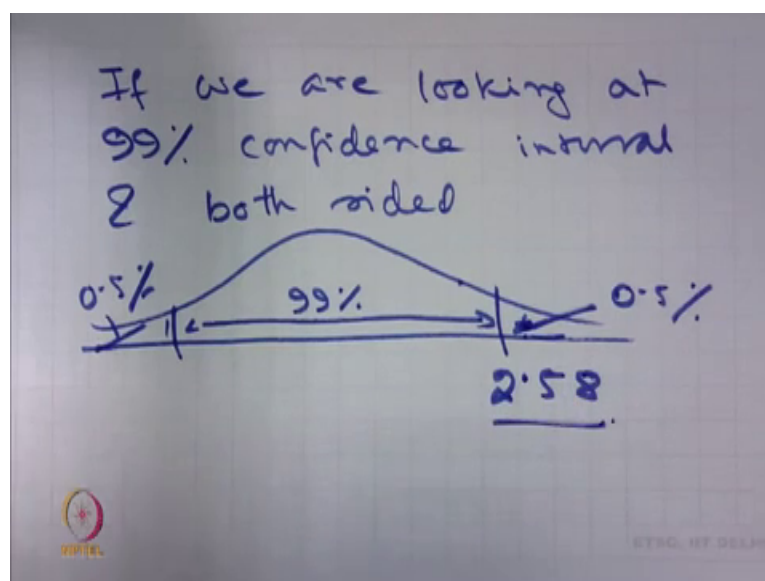
(Refer Slide Time: 52:51)



If we want to take one sided interval, say consider normal 0, 1 and we want one sided interval; that is, if  $Z$  follows normal 0 1, we are looking at  $z_\alpha$  such that probability  $z$  less than equal to  $Z_\alpha$  is equal to say 95 percent.

Then it is one sided interval; or in other words if it is a normal curve, we are looking at  $z_\alpha$  such that this area is 95 percent. That is, this area is 5 percent, these are called one sided intervals. And for normal 0, 1 at 95 percent, these value is 1.64.

(Refer Slide Time: 53:39)



If we are looking at 99 percent confidence interval and both sided therefore we are looking at this is only half percent. And this is 0.5 percent; that means, only 1 percent are beyond this range; therefore, this is 99 percent then for standard normal these value is 2.58. You should remember these values for solving problems with respect to interval estimation. Okay students, with that I conclude my lectures on theory of estimation.

We have learnt the properties of estimators. We have learnt how to estimate the value of a parameter from a sample, when you are looking at point estimation. Also today, we have seen how to look at an interval estimation for a parameter given a sample  $x_1, x_2, \dots, x_n$ . In the next class, I shall start the last chapter of statistical inference namely, testing of hypothesis.

Thank you.