Data Analysis and Decision Making – I
Prof. Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 15
Estimators

Welcome back my dear friends, very good morning, good afternoon and good evening to all of you. And this is the DADM which is Data Analysis and Decision Making I, this NPTEL MOOC series course. And this total course if you, if you know, which I do mention in the beginning of each class in general and then I basically wrap up whatever we have considered in the last class. So, this is a 12 week course for 30 hours. And each week we have 5 lectures each lecture being for half an hour, and we are in the 15 lecture; that means, you are going to end our third week lecture series.

So, if you remember and my name is Raghu Nandan Sengupta from IME Department, IIT Kanpur; If you remember we were discussing we did discuss in the last 2 lecture which was a 13 and 14 a little bit detailed, because these would be coming up time and again later on; about the chi square distribution, about the t distribution, about the F distribution. For the chi square distribution, it was basically chi square was being formed by the sum of the squares of n number of standard normal deviate. And we also saw that chi square distribution could be of degrees freedom of n and degrees of freedom n minus 1.

Now, for n we can have the degrees of freedom provided that the in set of information is not being lost for one case. What is that case? I am going to come to that later on, I will just repeat it. And for degrees of freedom n minus 1 we have lost one degrees of freedom, based on that we will proceed. Let me tell you another thing; obviously, I did not mention that, but it will be coming up later on also. I will basically wrap up this 4 or 5 important points about all the 3 distributions at one go once I basically. Give the wrap up of the revise what we considered in the 13 and the 14 lecture.

Then we went to the t distribution, t distribution was being formed by the ratio of a z divided by the square root of a chi square divided by it is degrees of freedom. I am just mentioning it because they were being utilized this concept of sum of the squares or division all these things would come later on. Then for the t distribution also we had at a tds distribution is n degrees of freedom t distribution with n minus 1 degrees of freedom. And I did mention that time and again that the t distribution has in the long run, if the with the sample size basically becomes very large; that means, degrees of freedom becomes very large. Then the mean value and the standard deviation value becomes, exactly equal to 0 and 1 which is basically the case for the z distribution also.

So, t distribution can mimic the z distribution or the other way around; that means, in case if you have a t distribution within the larger degrees of freedom, we can use safely use the z distribution in place of t distribution and continue with our work for the tables. I will come to the tables later on also. Even though I only discussed the tables for the z distribution. Then later on we came to the F distribution; so, in the F distribution in to you comparing 2 populations with the corresponding samples. Sample size being m as in mango n as in Nagpur. And, we considered their distributions with m comma n degrees of freedom or m minus 1 n minus 1 degrees of freedom depending, on what the loss of information is. Now remember one thing we also did mention, and I also did harp on the point that there are 2 s.

S is basically the standard error of for the sample which is basically the square root of the sample variance. So, in the standard error you basically have s dash square and s without the dash square. So, the s dash was the case where you do not lose any degrees of freedom, the mean value of the population remains as mu, and we divide by n in order to basically find out what is the square of the standard error for the sample. Now, when we are losing one degrees of freedom; that means, the population mean value mu is not known we replace it by the sample mean which is X bar n. And then we divided by n minus 1.

So, we will lose one degrees of freedom, because you are utilizing the first set of observations for the first time X 1 to X n to find out the best estimate for the population mean with this is the sample mean. And then we also showed that how the different ratios; that means, if you remember that I did mention that how the z chi square distribution was formed, how the t distribution was formed, and the F distribution was formed by a chi square divided by it is degrees of freedom in the numerator. And in the denominator another chi square with these degrees of freedom. So, these 3 ideas of the ratios and how the formation of these 3 distributions were formed, basically helps us in trying to understand what are the different rules based on which you can utilize these 3 distributions.

Now, one thing which I would like to mention here before we proceed further on, and later on we will see that in the case of interval estimation point estimation, and especially in the case of hypothesis testing this would be very important. That, remember the this chi square distribution is used for the case when we have something to do with the variance of a population and you want to test. It the z distribution so that means, only one population we want to basically we have some information from the past, we want to basically verify that fact is true or not.

Now, the F distribution is used for the case when we have something to do with the comparison of the 2 sample means variances, sorry, 2 sample variances or standard deviation of the population; such that we use the chi square with m comma n or m minus 1 n minus 1 depending on the case it as it is. Now, the question would; obviously, arise that when we do use the z distribution and the t distribution. So, as you can guess, the z distribution, the t distribution would be used in order to find out something related to the mean of the population. Z being the case when the when the information said related to the standard division is known. And t being the distribution when the standard deviation information; that means, for the population is not known. And when you want to find out the differences of 2 means also then you use either the 0 the t as the case may be.

So, this last few minutes or the last few seconds which I mentioned about when z and t would be used, and when chi and F would be used, I would repeat that time and again. So, please be rest assured that these things would become very clear to you as you proceed and discuss the ideas. Now, we will consider I am just giving an example I will come to the details later on most probably in the 16, 17 lecture. So, when if you remember we also discuss that if you have a population. So, your basic idea was to find out the parameters.

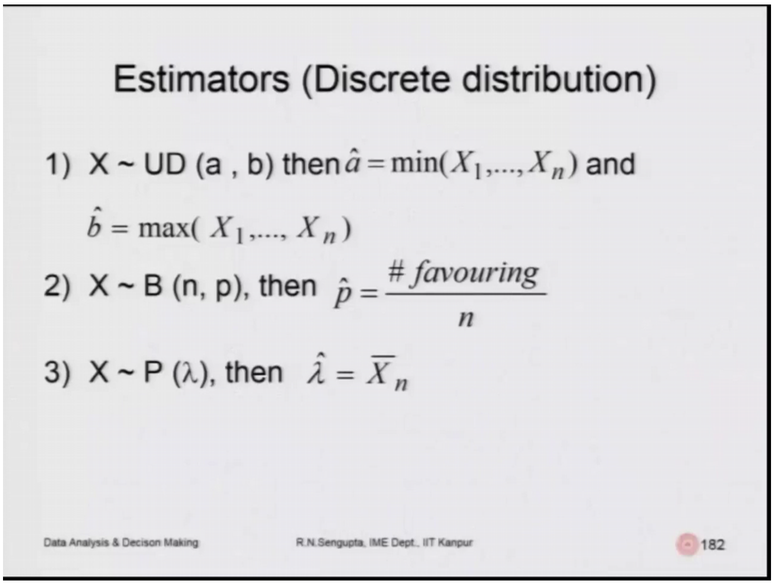
So, parameters for the normal was mu and sigma square parameters was for the binominal were n, np, exponential loss lambda, or a n lambda operation was lambda, or theta whichever you express. And using this parameters location shape and scale which generally in general statistics sense, the location shape and scale are given by alpha beta gamma, in order to express that how what we mean by the location shape and scale parameter.

So, using this location shape and scale parameter we find out the first moment second moment or the combination of the moments. First moment being mean value, second moment be the variance and so on and so forth now if we know about the population moments or population parameters, then trying to find out the parametric distribution of this of the sample becomes easier. In case if the population mean values or population variances of the modes of populations are not known what we do is that we replace by the characteristics from the sample, which is if you remember it is small t suffix n. So, this t is nothing to do with the t distribution remember that, the small t suffix n is basically the characteristics of the sample which we mix the population parameter or population characteristic to the best possible extent and we did also mention that the 2 important properties.

Being unbiasedness and consistency and if you remember I drew that diagram with a graph with 4 quadrant corresponding in the normal distribution in order to basically portray that what we mean by unbiasedness and consistency. So, generally considering the properties or unbiasness and consistency, we would basically have the estimators for different type of distribution. I will assume without the proofs what are the estimators. I am not going to go into the details of unbiasness and consistency. May be many of the cases the consistory property may not hold, but we will try to utilize those estimated meters in order to basically give a feel that how they can be utilized.

Now, consider this case. And this is very heavily used in the case of extreme value distribution even though extreme value distribution discussion is not part of this course. Now, in extreme value distributions or if they are skewed value distributions; say, for example, you want to do this experiment and the before I start the experiment let me give you the background, and what is written on the slide would make sense and then you will understand that what I was going to say.

(Refer Slide Time: 09:50)



Consider the, you have a discrete distribution, actually from the population and the minimum and the maximum values are a and b. And now a and b are not known to you. Now I would ask you the question that how can you find out a and b is it possible. So, there would be different ways all you can basically come out, but let me give you a very simple example. Say for example, you pick up a sample from that uniform discrete case.

The first case of n observations you pick it up, note down the minimum and keep it. Again and you either you replace this observation into the box or you do not replace. So, that would hardly matter because you are considering that the whole population is infinite I mean you are picking up observation samples of n number in size. You do the second time, pick up the mean from that second set of observations 1 to 2 n, note it down and keep it aside, you continue doing it.

So now what you have is basically for each case you have our sample where you have noted down the mean. And if you continue doing it technically it is possible that at one case when you do it in finite number of times, the mean a minimum of the minimum would basically be the best so called best predictor the word best I am using, not in the sense of unbiasness not in the sense of consistency is basically from a very layman's point of view how you can understand.

So, that minimum or the minimum would be the best predictor for a, and if you do the same thing for the maximum then the maximum and the maximum basically give you the best predictor for b. So, then based on that we can basically continue the work. So, this is basically what it says in point number 1 if X is the uniform discrete with a and b, then a hat now the word the hat is being utilized to basically denote that what is the estimator from the sample, and how it is able to basically give us the best prediction about the population.

So, a hat would be the minimum from X 1 to X n, and if I continue doing it I will find out minimum many minimums. And then I find out the minimum or the minimum and continue. While b hat would be the best predictor b would be the maximum from X 1 to X n. And again as I said you will continue taking the maximum many maximums find out the maximum the maximum and basically continue.

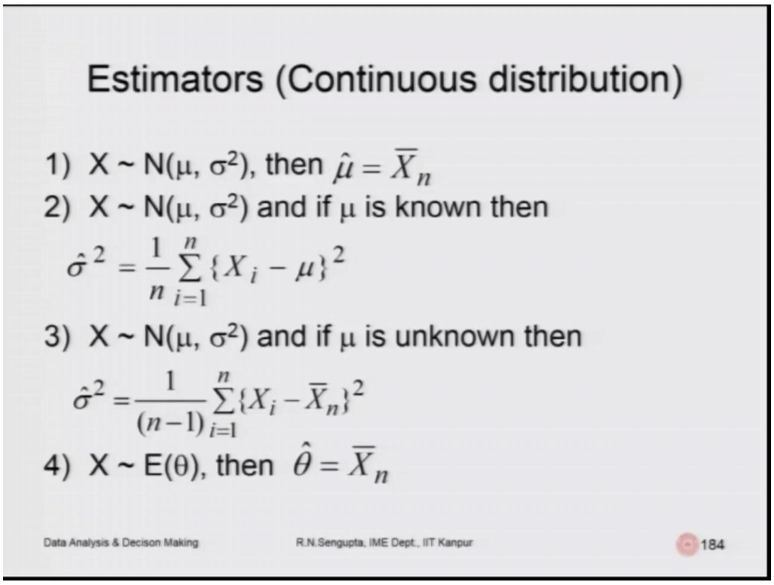
So, these technically and what I am going to do is discuss would be basically with the many of the estimates from the sample which are the best predictors for the population parameters based on which you can do the experiment. That means, based on experience in the sense, that we pick up the sample estimate, which are the best predictors for the population parameters, and use the sample estimate to basically do one experiment or do a studies whatever it is.

For the binomial distribution X with the parameters n and P, then if I want to find out p hat which is the probability of success or q hat which is the probability of failures and if you remember that 1 plus, 1 is equal to P plus q. Then p hat would basically be the number of things which are favorable divided by n, which is the set of observations we have, which you are picking up.

So, say for example, if you want to find out an experiment, I do not want to find out: what are the number of red balls. And or what are the number of white balls in a box which has red and white only. So, we will pick up observation find out the number of reds divided by the number of observations picked up and basically try to estimate what is the probability of finding out the red ball similarly due for the white balls also.

Now, consider we have a Poisson distribution. Again without proofs all these things which I am saying without proof. So, for the Poisson distribution that best if you remember the lambda is the mean value say, and it can be proved mathematically the best estimate for the lambda which is the mean of the population is the menials sample mean. So, hence if I pick up observations one to n add them up divide by n that will give me the best estimate for lambda. Hence it is given as lambda hat is equal to X bar n.

(Refer Slide Time: 13:56)



So, this is done, let me come to the exponential case and for the continuous case. So, few are the estimators for the continuous case. Pay of very special attention to them for the normal case, other things would be discussed as and when we proceed. Consider the normal distribution. Mu and sigma square are the parameters we need to find then the best estimators for the sample for the population mean mu it would basically be the sample mean which is X bar n.

Now, the question would rise that arise what the, what is the best estimator for the case when we have the want to find out something to the variance, to do with the variance of the population which is the normal distribution. So, in this case they would be 2 outcomes. And as I mentioned that in the 14th class or 14th lecture; so, consider the population mean is not known. So, in that case you will use if you remember I did mention about s dash and the s without the dash.

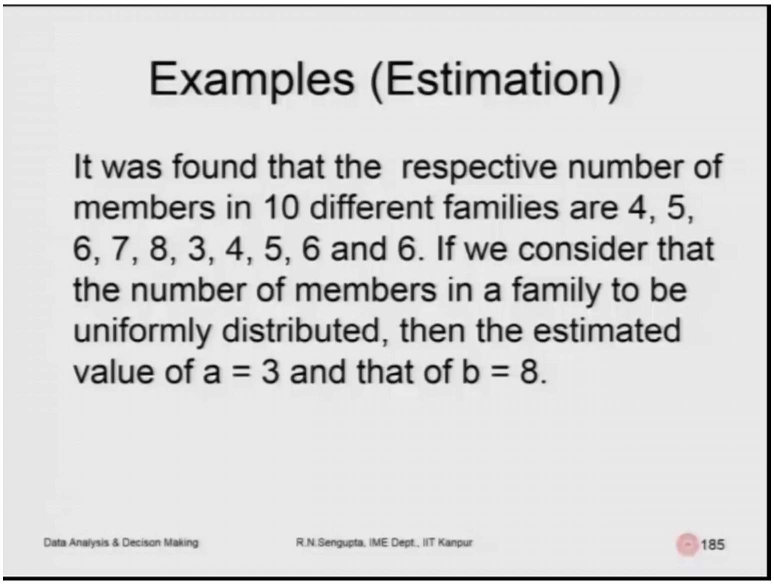
So, consider the population mean, let me make come start from the easiest one, consider the population mean is known. So, in that case the population variance if it is not known we want to find it out, then the best estimate from the sample would be 1 by n summation of X i minus mu whole square; that means, we are trying to find out the difference between the set of observations from the population mean square them up and add them up and divide by n. But in this case you should remember that we are dividing by n because we are not losing one set of information considering the sample of set of absorptions and been picked up.

Now, this would basically be the s dash. Now, when we do not have the population mean so, hence you want to replace the population mean by the best estimate from the sample. In that case, the best estimate from the sample which will mimic the population mean which is mu to the best possible extent would basically with the sample mean. So, if you are replacing mu by X bar n, in that case you have utilized the first set of observations from X 1 to X n. The first time and try to find out the sample mean, hence the degrees of freedom would be lost by 1, hence that value now would be s dash would be summation of X i minus X bar n whole square sum them up divided by n minus 1. So, this is what is given.

In case if you have say for example, and this will be utilized time and again. Please bear with me in case we have basically the exponential distribution with only one parameter which is lambda or theta; that means, the a value is 0. In that case, we will try to utilize the actual sample mean to be the best estimate for the population mean and proceed accordingly. Hence, lambda hat or theta hat would be given by X bar n.

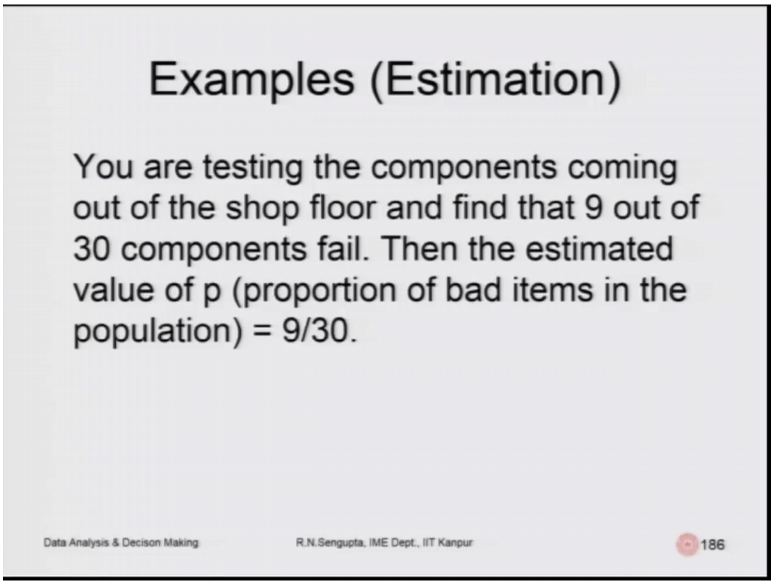
Now, somebody may be thinking that what if f a is given; is it not possible that if we pick up observations find out the mean and then continue with the world. So, obviously, those type of a layman ways of trying to handle them are intuitive and they give may give us results, but theoretically the results may not be sufficient to prove about the property of unbiasness and consistency. So, I will give you the results as we proceed and try to utilize them as and when the case arises for our studies.

(Refer Slide Time: 17:15)



So, let us consider an example. It was found that the respective number of members in 10 different families are respectively as 4, 5, 6, 7, 8, 3 and so on and so forth. So, if you consider that the number of members in a family to be uniformly distributed, then the estimated value of a; that means, a would be what in this case that we will find out the minimum of all the set of observations which is basically 3. And the maximum of that set of observations which will give us b; that means, a hat and b hat would be the best predictors of a and b would corresponding the values b. 3 for a which is a hat, and 8 for b which is b hat.

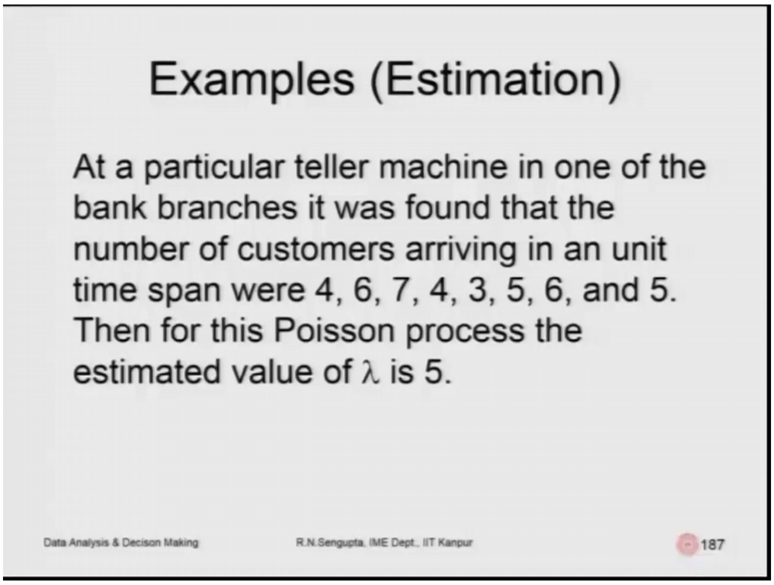
(Refer Slide Time: 17:45)



Now, consider your testing the components coming out of the shop floor, and find out that 9 out of the 30 components failed, then the estimated value of p which is the proportions of the objects which are bad or good. Whatever we say they would be given by 9 by 30. So, say for example, that if you find out that the 9 of them failed which means the number of objects that actual proportions of the probability of that total lot to have defective and non-defective one would be in the following way calculated.

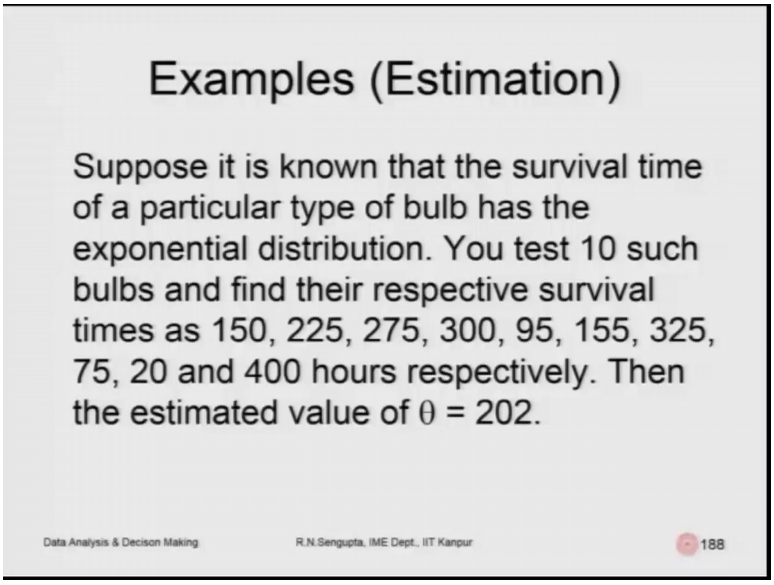
So, the probability of the defective one would be 9 by 30, because 9 on the bad ones out of the 30 pickings we are doing that. And a corresponding value of p which is the good items would be in the case, if the actual value of q was given it will be 1 minus q would give your p in this case as q has been replaced by q hat which is the best estimate you will find out 1 minus 9 by 30 which will be 20 1 by 30 will give us the best value of p which is p hat.

(Refer Slide Time: 18:43)



Let us consider the next example. At a particular tailor machine in one of the bank branches it was found that the number of customers arriving in an unit time span were given by 4, 6, 7, 4, 3, 5, 6 and 5. So, they are arriving and obviously, we want to find out the rate would be given by the mean value, and what is the expected value based on which you want to do the calculation. So, this for a Poisson process is the best estimate for the population mean would be the sample mean. So, for to find out the sample mean, what we do? We add the set of observations divided by the number of observation that value of lambda hat or mu hat or theta hat whatever you are saying would be given by 5.

(Refer Slide Time: 19:32)



Suppose it is known, the next example, suppose it is known that the survival time of a particular type of bulb has the exponential distribution, you test 10 of them. So, in sample size is 10, and the bulbs were found to be and have a survival time of 150, 225, 275, 300 and 95, 155, 325, 75, 20 and 400 hours. But remember here there in this case very into degree and rightly. So, we are considering a to be 0 because you put on the bulb, and it can fail immediately. So, obviously in that case a is 0. And then the best estimate would be as I have already discussed for lambda would be lambda hat, lambda hat which is basically given by the sample mean. So, for the sample mean, what we do? You add up all the values starting from 150 to 400 divided by 10 and that value comes out to be 202, ok.

Now, before going into, I will I am giving the flavor and then I will switch over to the concept of estimation all these things. Now, whenever we want to do some studies, and I will give you this background using the multiple linear regression. Whenever you do some studies, our main idea is to basically to find out an estimate from the for the population values, and do those studies in such a way that they mimic the population characteristics to the best possible extent. And I, for that I gave you the ideas of consistency ideas of unbiasness and proceeded accordingly.

Now, the task is there are 3 type of things we will try to cover in the in the univariate case and the multivariate case. The first one we being the case, when you want to find out the actual point value of that particular population parameter 0.1. Another would be basically we want to find out the interval range between which the population parameter would basically be point number 2. And point number 3 would be the case when you want to basically hypothesize or basically pass the judgment based on some set of information which you have. And you want to be you want to either disprove of proof depending on whatever set of observations which you have, and what is your conclusion based on those set of observations on which you do the statistical test.

Now, that for the case of the point distribution, our main task is to find out unbiasness consistency of those sample estimates which I just discussed like that for the case for the lambda for the case for the say. For example, for the Poisson distribution theta or lambda for the case for the normal distribution is mu and sigma. So, obviously, we replace them the way they had values corresponding to what the information we get from the sample. Now, what is important for us further on is to find out that; obviously, when we want to basically predict or estimate that population parameter, we would also be interested to give the idea that what is the interval between which it will lie lying. So, when we are trying to basically do that so, our main task would be to what level of significance we are able to meet.

So, level of significance means that what is the accuracy level; considered very significant very simply. Now, when you are mentioning the point of level of accuracy or level of significance, what we will our level of confidence what we mean is that, if my level of confidence is 90 percent what I want to portray is that if I keep doing this experiment 100 number of times 90 of those times or successful based on the fact which we have produced. And 10 percent in the cases the case is not successful. Hence, the probability of success or the level of confidence is given by 990 by 100 which is 0.9 and in the case of not being successful will given by 10 by 100 which will be 0.1.

Now, one thing should be remembered, that whenever you are trying to basically understand the level of significance or level of confidence, they would be technically 3 ways. 3 ways are not distinct ways, but I am giving you a flavor that 3 ways of trying to analyze the problem. In one case it can be off the type greater than, greater than whatever value we have and based on that the interval or the hypothesis would be framed accordingly. I will come to the exact things of hypothesis and an interval later on. And in the second case it would second case it would be basically either less than time. So, given one value we want to basically hypothesize or basically pass in judgment that it will be less than time. And in the third case it will be within certain range.

So, when we basically mean the level of confidence or the or say for example, 90 percent 10 percent whatever it is, we will frame the question in such a way that the total probability will be divided accordingly. So, let us remember that. So, when I am using the word, that the mean value of the say for example, some tie dots which are coming of from the shop floor is; say for example, 1.2 meters, and if I am hypothesizing that in on an average the values are greater than.

Then obviously, it would mean that the greater than type would have a certain probability depending on my level of significance or level of confidence. When I am a mentioning the point that they would be less than 1.2 then; obviously, the corresponding level of confidence would be framed accordingly. When I am saying that they are in they would be somewhere between 1.2 and say for example, 1.25 in that case, that interval would basically determine and give me some information what is the level of confidence which you want to have.

Now, in all these cases what will consider technically that, whatever the distribution is it is a binomial distribution, it is basically a Poisson distribution, if it is an exponential distribution, hyper geometric distribution; obviously, our life would be best if we knew the exact values of the sample estimate and proceeded accordingly. But the problem is that in many of the cases trying to find out the exact values of the parameters for the population given the samples estimates is not possible. So, what we do is that we consider the central limit theorem to be true. So, if you remember that it did mention about the central limit theorem and that will now become very important as we are discussing.

The central limit theorem to be true will utilize the central limit theorem, replace the actual distribution by the normal distribution, but be rest assured that corresponding mean value and the sample and the standard deviation of the variance of that particular normal distribution would be some property of the norm of the mean value and the standard deviation of the actual distribution.

So, when we are taking a binomial distribution, then the corresponding central limit theorem will give us a normal distribution; where the mean value and the standard deviation of the variance of the normal distribution would have something to do with np and npq. Where, if you remember np and npq are the corresponding mean values and the variances of the binomial distribution.

So, we will consider those things in such a way such that they are trying to solve the problems would be easier. Now remember on one thing that the case of whatever you will discussing the interval estimation, the point estimation the hypothesis testing, would be given from the viewpoint of the univariate case. And later on it will be given in such a way that you can easily consider the multivariate case very simply. If this has been more of at least in this 15th lecture, I have been discussing more about the concepts we are going to come across. And I will come to this actual utilization of the theorems and utilization, remember me I am not going to discuss the theorems in detail the utilization of the theorem the concept through diagrams and so, 3 simple problems.

So, if you read those type of books which I have suggested, it will be it will give you a lot of information's about how we are going to analyze the problems. The reason being that for when you are trying to basically analyze the problems from the point with the multivariate cases, many of the cases may become a little bit complicated. I am not saying they are absolutely difficult they may become completed, but the idea of trying to basically understand the multivariate distribution with the case of the univariate case would give you a lot of confidence if you are able to handle the simple problems in the univariate case.

With this I will end this 15 lecture. And the slide which is in front of you I will come to this to please bear with me, I will come to these discussions later on. So, with this I will end this lecture. And thank you very much for your attention and thank have a nice day.

Thank you.