Essentials of Data Science with R Software – 1
Professor. Shalabh
Department of Mathematics & Statistics
Indian Institute of Technology Kanpur
Lecture No. 56
Unbiased Estimators

Hello friends welcome to the course essentials of Data Science with R Software – I in which we are trying to understand the basic concepts of probability theory and statistical inference. So, you can recall that in the last lecture, we initiated a discussion on the estimation of parameters. And towards the end I had to explain you that there are a couple of properties which are expected in a good estimator because if these properties are present in those estimators, then we try to say that estimators are good.

So, in this lecture, I am going to talk about the property of unbiasedness right well out of the properties biasedness, consistency, sufficiency efficiency etc there is no ordering, but I am simply trying to take the biasedness, as the first property not because of any ordering because I have to choose one topic for this lecture. The first question comes, what is this unbiasedness? Forget about statistics. Think about in a very simple way. Have you ever used this word unbiased in your life day to day life?

Answer is yes, many times you say that, this teacher is unbiased or sometime you say no, no, this teacher is biased towards a particular student or sometimes in our family also we say, my parents are biased towards my younger brother or towards my elder brother or my sister. These types of sentences are very common. Now, in case if you try to understand and try to explain yourself, what do you really want to convey when you try to use this word unbiased or biased?

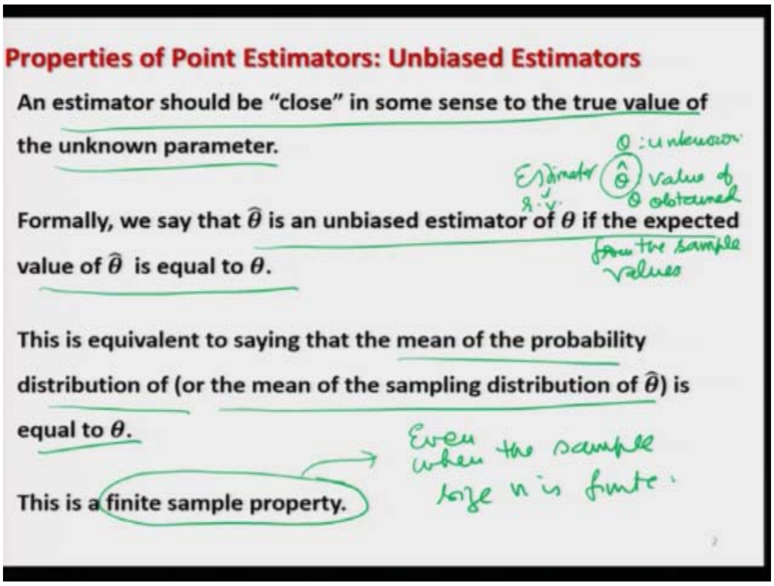
For example, if there are two students in a class and suppose me as a teacher, what I try to do, that both the students are suppose making the same mistake in the examination in a question and suppose out of 10 to one student, I gave suppose 7 marks and I deduct 3 marks for the mistake and to other student I give only 5 marks and I deduct 5 marks for the mistake. Now, what will be the comment of the students for me? They will say, I am biased towards a student that means, what would I have not done so, that they cannot call me biased.

The only thing is this, if I had given the same marks to both the students. Means either 7 to both or 5 to both, then there should not be any problem. So, that means, whenever I am trying to deviate from the true value, I am trying to introduce a bias in the marks off my students. So now, if you try to think in these lines and try to think about your estimator that is going to estimate the value of an unknown parameter of the population what do you expect?

You expect that the parameter which you want to estimate, which is actually really unknown that you do not know, you are trying to know its value through estimator which is a function of a random variable, that is a statistic, and you want this value to be close to the true value. You do not want µch deviation from the true value. So, this is basically the idea about this biasedness character of an estimator.

So, let us try to begin our lecture and try to understand from the statistical point of view, how do we define this biasedness, and how do we check that whether an estimator can be set to be an unbiased estimator of the unknown parameter or it is a biased estimator of the unknown parameter? So, let us begin our lecture.

(Refer Slide Time: 04:31)

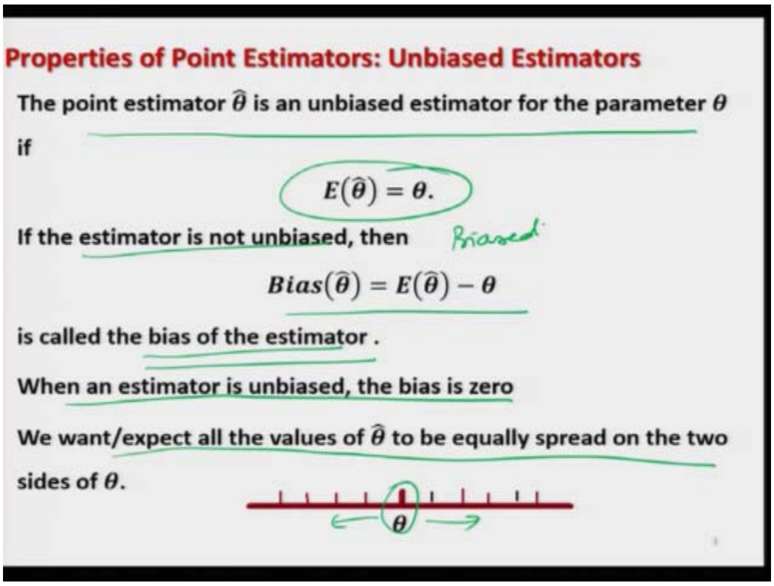


So, now, in case if you try to recall that you were trying to estimate an unknown parameter θ. And we had indicated by this a $\hat{\theta}$, this is a value of θ obtained from the sample values that you know. So now, and this is and this $\hat{\theta}$ is called as an estimator. And $\hat{\theta}$ is also a statistic, so this is a random variable. So, what we expect that an estimator should be close in some sense to the true value of the unknown parameter.

And formally and statistically speaking, we say that $\hat{\theta}$ is an unbiased estimator of θ if the E($\hat{\theta}$) = θ. And if you try to understand what is the meaning of this. This is trying to say that the mean of the probability distribution or the mean of the sampling distribution or $\hat{\theta}$ is equal to θ. And remember one thing this property the property of this unbiasedness, this is a finite sample property.

Why I am calling it a finite sample property, because in the next lecture, I am going to talk about the property of consistency, and consistency will be an asymptotic property, means, when the sample size become large. But when I say there is a finite sample property that means, it will hold true even when the sample size n is finite. This is what I mean.

(Refer Slide Time: 06:16)

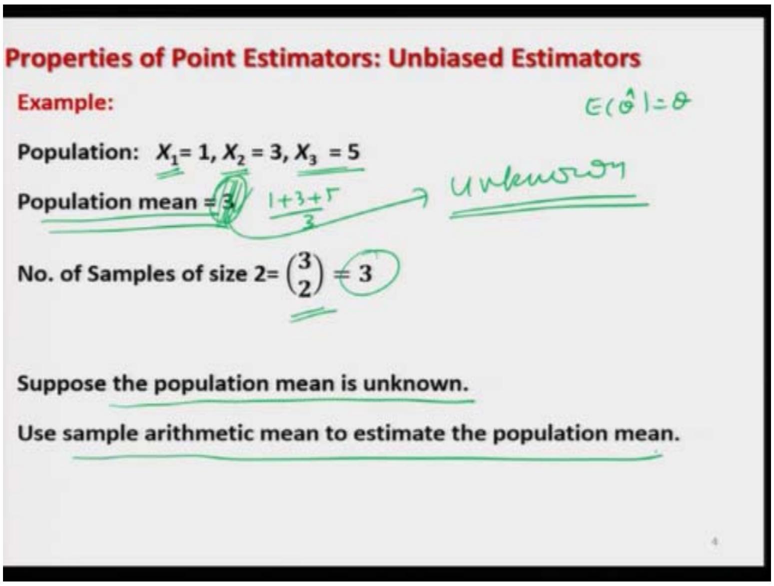


So, now, I can introduce here the formal definition once again that the point estimator $\hat{\theta}$ is an unbiased estimator of the parameter θ, if $E(\hat{\theta}) = \theta$. So, and in case if this condition does not hold true, then we say that the estimator is not unbiased, but it is biased. And in that case, the bias of $\hat{\theta}$ is defined as like this $E(\hat{\theta}) - \theta$, this is called the bias of the estimator.

And we will always be interested in knowing whether the estimator that we are using for estimating an unknown parameter of the population is an unbiased estimator of the parameter or not and if it is a biased estimator, then what is the amount of bias. Because if the bias is large, then possibly the estimator may not give us the good value of the unknown parameter θ.

So, obviously, when an estimator is unbiased, obviously, the bias is zero. And the interpretation of this unbiased character is that we want or respect that all the values of θ had to be equally spread around the two sides of θ like this. If this is my here θ, then this value of these are the values of θ here on these two sides.

(Refer Slide Time: 07:39)

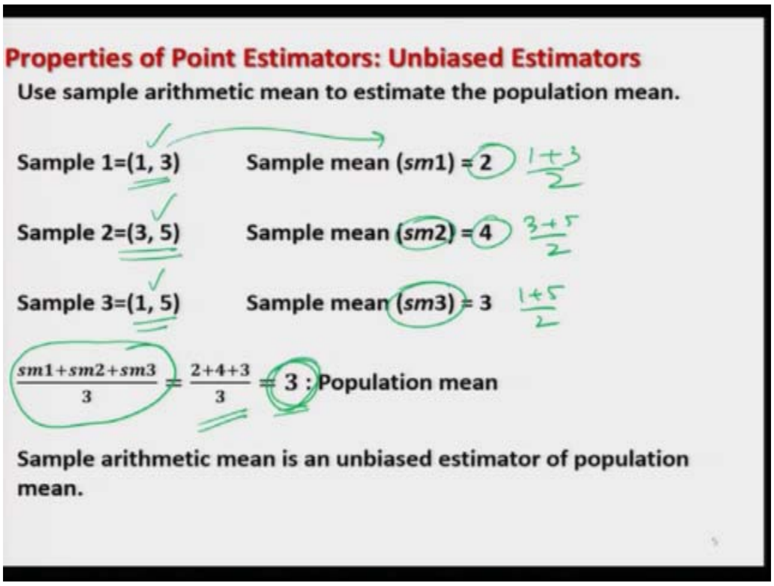


And let me try to explain you that what is the meaning of $E(\hat{\theta}) = \theta$. And I am trying to take a very simple example. This simple example will make you understand that what are we really going to do when we are trying to deal with the probability distribution. Suppose I have a population of only 3 values, say capital X_1 equal to 1 capital X_2 equal to 3 and capital X_3 equal to 5. And now if you try to see what is the arithmetic mean, of these 3 values, which is here, 1 plus 3 plus 5 upon 3, X is equal to here 3, so the population mean here is 3. And yeah, this you can say that this is unknown to us.

Just for the sake of example, I am just taking these values but definitely in practice, this value will be unknown, so I can just remove this value from here. Means I cannot hide this value so that will indicate as if you do not know. Now, suppose you want to know this value and you try to draw a sample of size 2.

Now, there are $\binom{3}{2}$ or say 3 chose 2 possible ways in which you can draw a sample of size 2 from a population of size 2. So, this is three choose two is equal to 3. So, now, we want to estimate the unknown population mean by the sample arithmetic mean. And so, for that, we try to choose here all the three possible sample and we try to find out the sample mean.

(Refer Slide Time: 09:06)



So, if you try to take the first sample, suppose you are getting the values say 1, 3 then the sample mean in this case will come out to be 1 plus 3 divided by 2, which is actually 2. And in case if you draw one more sample, so there are actually 3 possible samples. The value is 1, 3, 3, 5, and 1, 5. So, suppose if you try to draw a second sample, and suppose you get the value 3,5 , then in this case the sample mean will come out to be 3 plus 5 upon 2 which is equal to here 4 and that is supposed to be the indicated by SM 2, sample mean 2.

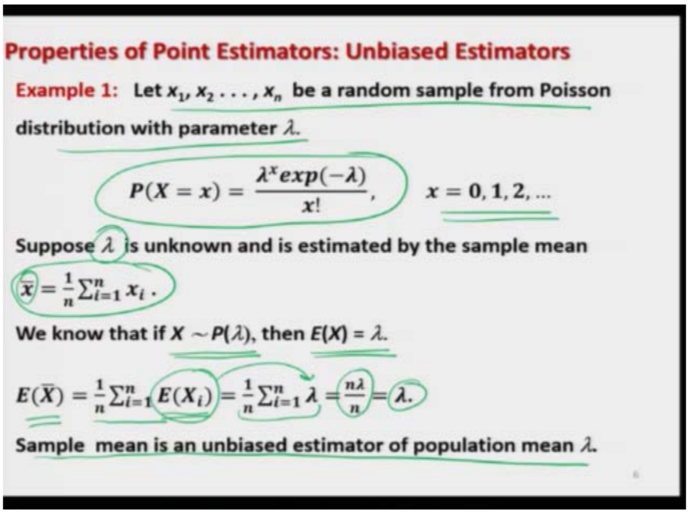
And similarly, if you try to take the units 1 and 5 in the sample, then the sample mean is going to be computed by 1 plus 5 upon 2 which is equal to 3 and that is going to be indicated by the indicator or the symbol SM 3 sample mean 3. Now, try do one thing. Try to take the arithmetic mean of these sample means. So, we try to compute here $\frac{sm1+sm2+sm3}{3} = \frac{2+4+3}{3}$, and this will come out to be here 3. So, now, if you try to see what was this 3?

This was you are here, this population mean that was unknown to us. So, you can see here that if you are trying to take all possible samples from the population. And in case if you are trying to find out the mean of all the means, and if this value comes out to be the same as the value of the population mean then we say that our estimator is an unbiased estimator of the population mean. And if you try to see here, here you are trying to estimate the value of this population mean, which is actually here 3, you are trying to estimate it by the value 2 or 4 or by 3.

So, you can see here that the difference between the estimated value here. These are your here estimated value and population mean it is not going to be large. Well, in practice you are going to have only one sample. So, if you get this sample you get here a sample mean 2 so that division is only for 2 difference 3, and if you are getting this suppose sample whose sample mean comes out to be 4.

So, we can see here that the difference between 3 and 4 is only one unit. And similarly, in case if you get a sample here whose sample mean comes out to be a 3 then the difference between 3 and 3 is exactly going to be 0. So, this is what we mean by the unbiasedness character or the property of the estimator.

(Refer Slide Time: 11:44)



So, now, let me try to take here a couple of examples and I try to show you that how you can establish that whether an estimator is unbiased estimator or not. We will see after some time that do you really need to know the R software or do you really want to know that how to compute them in the R software, let us try to understand these things.

So, suppose let me take the first example, where we said that let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ. So, we know the probability mass function Poisson is given by this, $\frac{\lambda^x exp(-\lambda)}{x!}$, $x = 0,1,2, \ldots$.

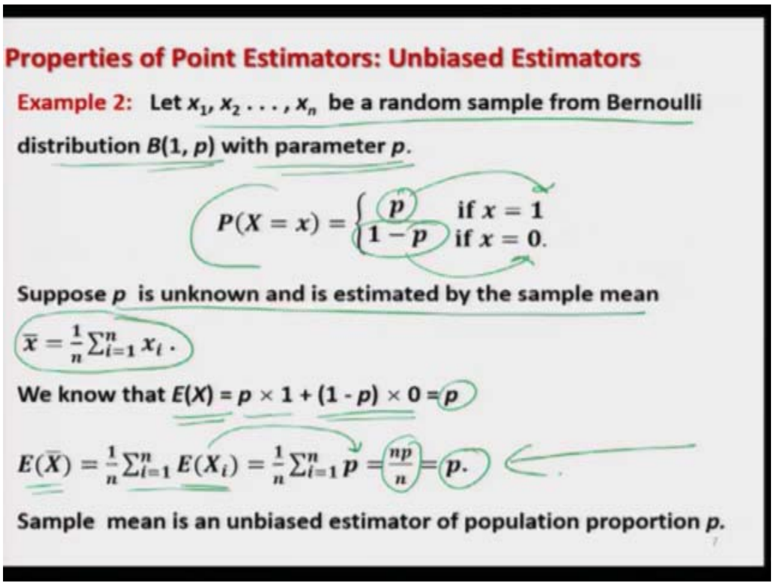
Now, you want to know the value of parameter λ that is unknown to us. And suppose, we propose that we try to estimate the value of λ by the sample arithmetic mean, Well, that is another question that how can you take such a call that whether λ can be estimated by arithmetic mean or some other quantity?

So, that part, I will skip at this moment and I will try to address this issue, when I am trying to introduce the estimation methods. So, at this moment you can just assume that suppose somebody has informed you that you can estimate the unknown parameter λ by the arithmetic mean, say this one $\frac{1}{n}\sum_{i=1}^{n} x_i$. So, now, we want to know that if I am trying to know the value of λ by the sample mean $\bar{X}$ then is it going to be an unbiased estimator or not.

So, now, I try to use the properties that we have learned earlier. We know that if a random variable has got a Poisson distribution with λ, then the E(X) is coming out to be λ. Means if you cannot recall, you can just go to the lecture on the Poisson Distribution and there you will find this result. Now, I try to find out here the E($\bar{X}$), and now you know how to compute this value. So, that is going to be simply here $\frac{1}{n}\sum_{i=1}^{n} E(x_i)$.

Now, this $E(X_i)$ is here λ. So, now, if you try to solve this quantity, this will come out to be nλ/n, which is here λ. So, you can see here that E($\bar{X}$) is coming out to be same as the value of the population parameter λ. So, I can say that sample mean is an unbiased estimator of the population mean. So, now, you can see here this is how we try to establish whether an estimator is an unbiased estimator of the population parameter or not.

(Refer Slide Time: 14:36)



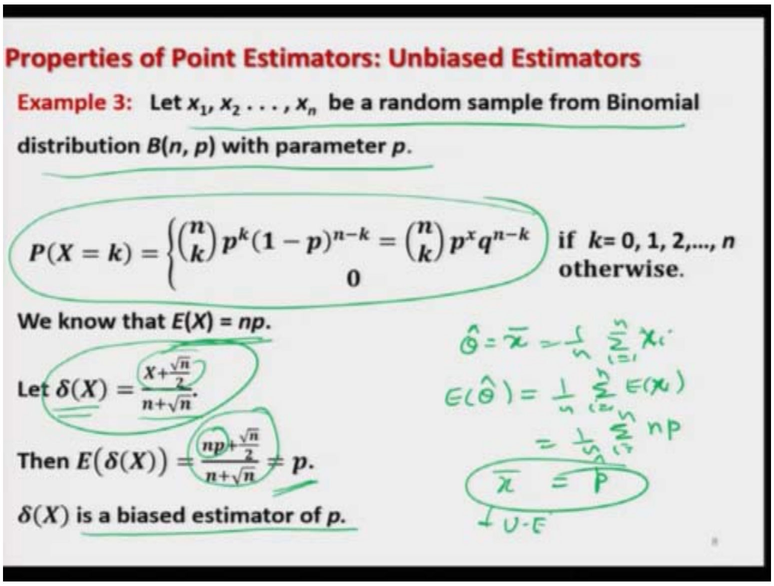
Now, let me try to take one more example. Suppose, I try to consider that let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution, which was indicated by B(1, p) with unknown parameter small p. So, this is the probability mass function of Bernoulli distribution. So now, since, the parameter p is unknown to us suppose we propose or somebody inform us that a p can be estimated by the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Now, we want to know whether this estimator is an unbiased estimator or not. So, we know from the properties of Bernoulli distribution and that we also had understood earlier also that E(X) =p X 1 + (1 – p) X 0.

Now, you can see here it takes value p when X equal to 1 it takes value 1 minus p, when X is equal to 0, so, this value will come out to be here p. So, now, if you try to find out the E($\overline{X}$), this will come out to be $\frac{1}{n}\sum_{i=1}^{n} E(X_i)$ and then you try to replace E(X_i) by p and this will come out to be np/n that is equal to here p.

So, in this case also you can conclude that sample mean is an unbiased estimator of the population proportion or the population parameter is small p. Well, in case if you try to take something else possibly this may not hold true, but that you can know only after doing this type of algebra.

(Refer Slide Time: 16:15)

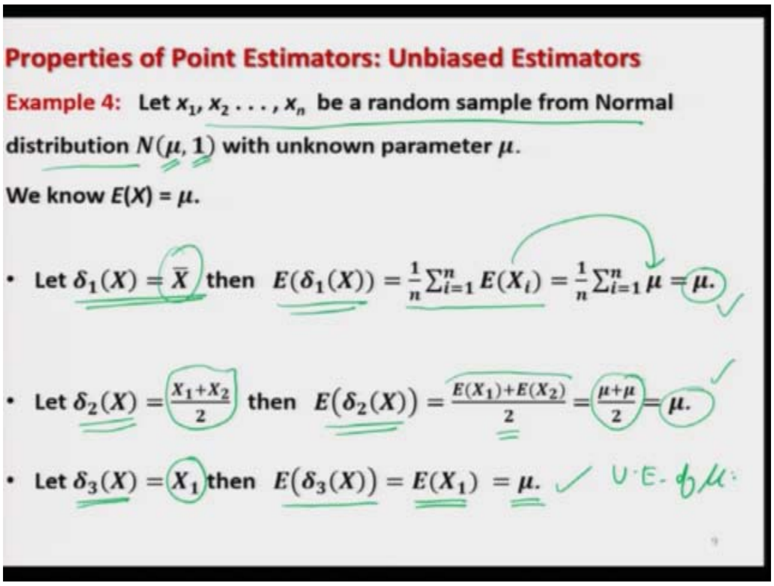


Now, just to give you an idea that, that not all estimators are always an unbiased estimator of the population mean. Suppose I try to take one more example, where I assumed that the random sample $X_1, X_2, \ldots, X_n$ is drawn from a binomial distribution with parameter p. So, we know that the probability mass function is given here like this that we had understood earlier. And in this case E(X) = np.

So, now, in case if you try to take an estimator here say $\delta(X) = \frac{X+\frac{\sqrt{n}}{2}}{n+\sqrt{n}}$, and if you try to take its expectation, this will come out to be E(X) is going to be replaced by np and all other part is constant so it will appear as such. Whatever you do, means, you cannot say that in general this is equal to p, this quantity is not equal to p. So, in general means, I can say that this estimator δ(X) what you have considered here this is a biased estimator of p.

Well, in case if you try to take it here another estimator here say $\hat{\theta}$ is equal to $\bar{X}$. You can see here, this is the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ and if you try to take its expectation it will come out to be 1 upon n, summation i goes from here 1 to n and expected value of Xi expected value of Xi is going to be here, say here np, so this will come here and be here p. So, you can see here in this case, this sample mean will come out to be an unbiased estimator of p, but this estimator δ(X) is a biased estimator of p.

(Refer Slide Time: 18:05)

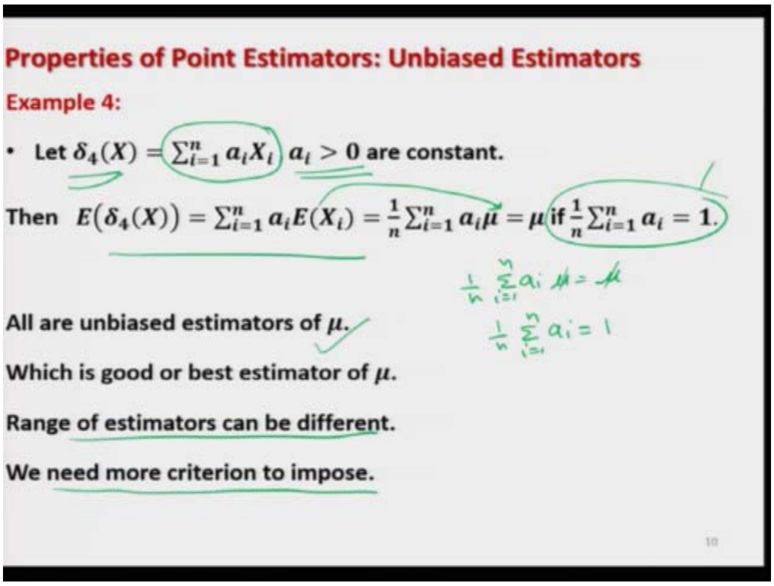


Similarly, I try to take here one more example. And this sample $X_1, X_2, \ldots, X_n$ is coming from a normal distribution, and whose mean μ is unknown, but variance is known as 1. Now, I tried to consider here suppose a different estimator and then I try to see for their unbiasedness. So, I try to consider here this estimator $\delta_1(X) = \bar{X}$ and then I try to see the $(\delta_1(X)) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$. So, this δ 1x equal to $\bar{X}$ this is an unbiased estimator of the population mean μ.

Now, I try to propose that instead of using the sample mean, I tried to use the arithmetic mean of the first two observation like an X_1 plus X_2 upon 2 and I define here another estimator $\delta_2(X)$. Now, in case if you try to take the expectation of $\delta_2(X)$ this comes out to be $(E(X_1) + E(X_2))/2$. This is μ plus μ divided by 2 which is equal to here μ. So, you can see here, that is X_1 plus X_2 upon 2 as an estimator of μ is also an unbiased estimator of μ. So, now, you have seen that there are two estimators that are unbiased.

Now, let me try to take one more option. And I decided that I will use only the first observation to estimate my μ, and I denote this by $\delta_3(X)$. Then you can see here that expected value of $\delta_3(X)$ will come out to be as expected value of X_1 this is also actually μ. So, now you can see here not one, not two but all these three estimator they are all our unbiased estimator of μ.

(Refer Slide Time: 20:07)



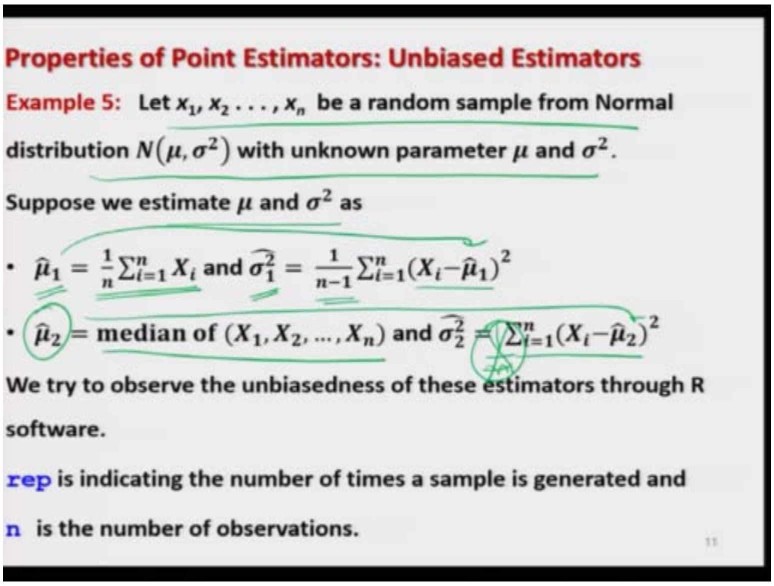
And similarly, if I try to take one more option that suppose I say that suppose I try to consider the $\sum_{i=1}^{n} a_i X_i$ as an estimator of μ, where this a_i greater than zero or some constant. Now, I want to know that under what type of condition this $\delta_4(X)$ that a $\sum_{i=1}^{n} a_i X_i$ is going to be an unbiased estimator of the population mean. So, I try to take his expectation and I find out here that this $\sum_{i=1}^{n} a_i E(X_i)$ and which can be replaced by here μ.

So, now, I can say that $\frac{1}{n}\sum_{i=1}^{n} a_i \mu = \mu$ that means μ here is a constant on both sides. So, I can cancel this here μ on both the sides and I can write down here $\frac{1}{n}\sum_{i=1}^{n} a_i = 1$. So, this is the condition that if this condition holds true then I get an unbiased estimator of μ here. So, now, you can see here suppose, if this condition is true that we are trying to choose this constant in such a way such that $\sum_{i=1}^{n} a_i = n$ then you can see here all these estimator $\delta_1(X)$, $\delta_2(X)$, $\delta_3(X)$, and $\delta_4(X)$ they all are unbiased estimators of μ

Now, the question comes, which of them is a good estimator or the best estimator of μ? What would you suggest to a user that in order to know the value of an unknown parameter μ, which of the estimator out of these four estimator can be used that is going to give us a good result. So, that means, we have to think something more.

And besides this thing you can see here that the range of the estimators of δ 1x del,ta 2x, δ 3x and δ 4x, that is also different. And in general, that may be different for actually is. So, that means, we need some more information or we need to more criteria to be imposed on such estimation procedure that only one quality that is one property that is the property of unbiasedness is not going to help us. We need some more qualities.

(Refer Slide Time: 22:27)



So, now, let me try to take here a very simple example, and I try to show you on the basis of R software that how it does things, these things will look like in case if you are trying to establish that whether the property of unbiasedness holds for an estimator or not. Because in data science, it is possible that you are trying to use an estimator to estimate any parameter or a function of parameters, and you really do not know or it is very difficult to establish mathematically that whether the estimator is unbiased or not.

So, in these cases, we can use the simulation techniques, and we can have a fair idea that whether this estimator is going to be an unbiased estimator or not. Well, these results are based on the simulation, and these results will change from simulation to simulation and every time you conduct this experiment, but more or less, the results are not going to change mch so, that you cannot be confident in saying whether this estimator is biased or unbiased.

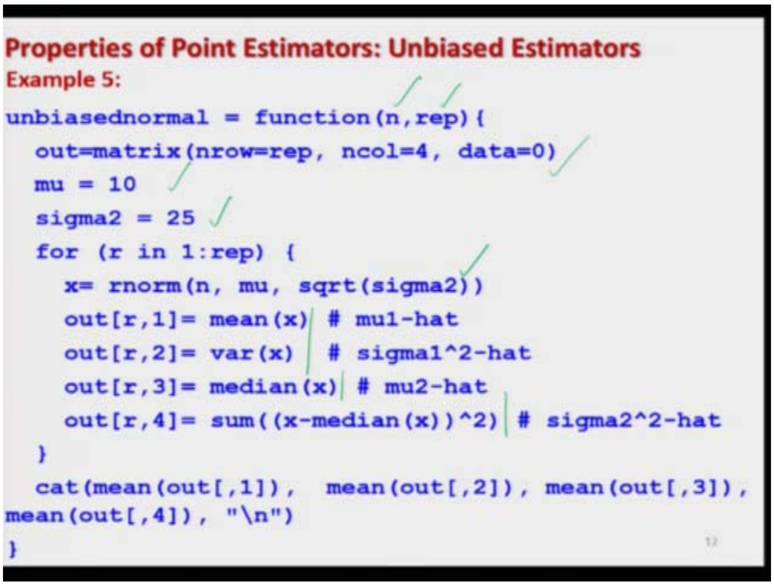
So, what we try to do here that we try to take a take a sample from the normal population with mean, μ and σ^2. And suppose this both μ and σ^2 are unknown parameters. Well, how to estimate a σ^2 that we have not yet discussed, but I will discuss it a soon. But here because for the sake of understanding, suppose, we propose that we are going to estimate μ and σ^2, and there are two possible choices.

The first choice is this I will try to take here $\hat{\mu}_1$ to be here simply this sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ and the estimator for the variance is $\widehat{\sigma_1^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_1)^2$. And similarly, the other option I can say that instead of computing the mean I try to compute the median of $X_1, X_2, \ldots, X_n$ and let this be my second estimator $\hat{\mu}_2$. And based on that we try to estimate the value of σ^2, as here says $\widehat{\sigma_2^2} = \sum_{i=1}^{n}(X_i - \hat{\mu}_2)^2$.

So, notice one thing that here I have not used any division here like this 1/n or 1/(n – 1) this is not here. It is simply the sum of the squares. So, now, we try to observe between these two estimator of μ and σ^2 how we can get an idea that whether these estimators are unbiased or not? Definitely I am saying once again the 100 percent confidence can be achieved only when we can prove mathematically theoretically that the estimators are unbiased or biased, but in complicated situation, this siμlation will help us at getting a reasonably fair idea.

So, what I try to do? Now you know how to generate the random numbers from the normal distribution. So, I try to generate a large number of samples and then I try to compute the values of here $\hat{\mu}_1$, $\hat{\mu}_2$, $\widehat{\sigma_1^2}$, $\widehat{\sigma_2^2}$ and then I try to compute the mean of those $\hat{\mu}_1$, $\hat{\mu}_2$, $\widehat{\sigma_1^2}$, and σ 2 square.

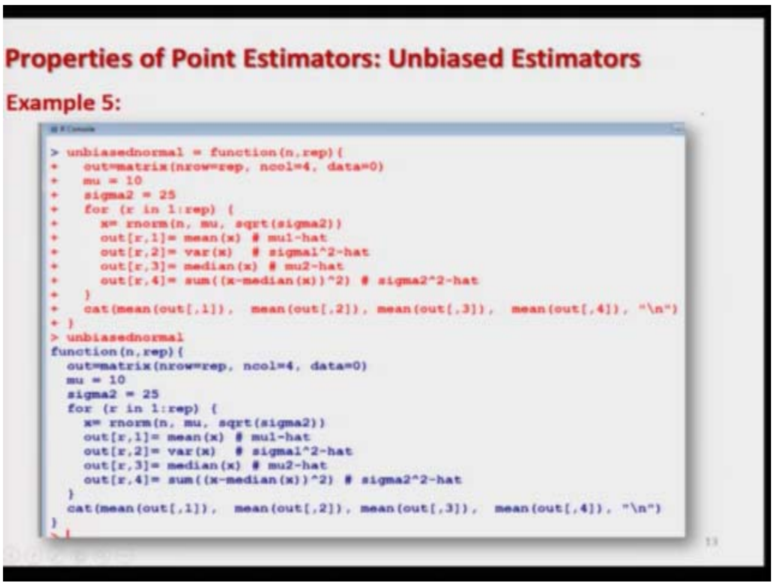
(Refer Slide Time: 25:48)

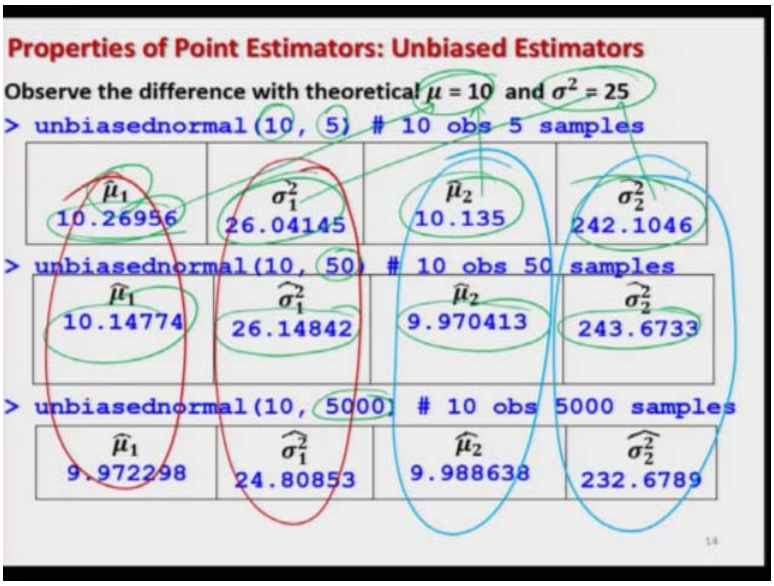


So, here I have written this small program, you can also conduct it on your computer, but anyway. I will try to show you it towards the end, but at this moment, I will try to show you the outcome. So, here I am simply trying to generate say n number of observation and we are going to repeat them for say some number of times and whatever are the outcomes, they are stored in a matrix.

And here I am considering μ equal to 10 and σ^2 as 25. And then I try to generate the normal the random observation from the normal distribution and I try to compute the mean and variances using the estimators $\hat{\mu}_1$, $\hat{\mu}_2$, $\widehat{\sigma_1^2}$, $\widehat{\sigma_2^2}$. And then I try to compile the information.

(Refer Slide Time: 26:32)





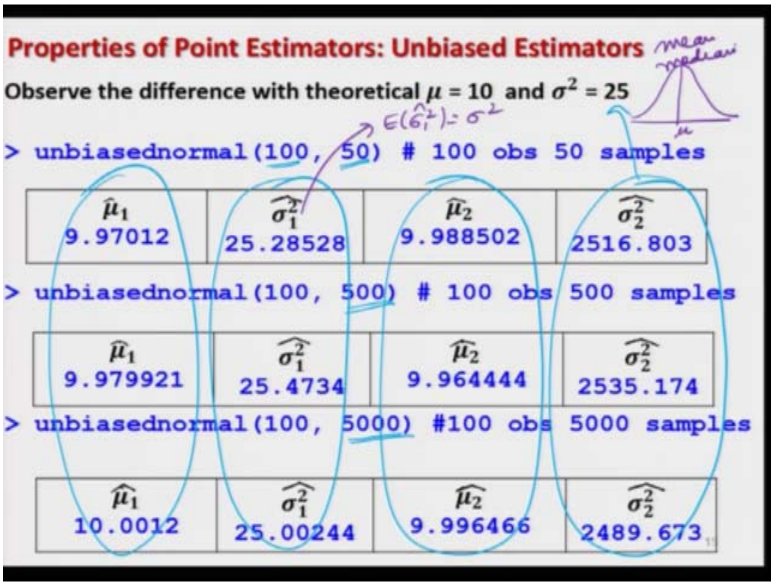
So, you can see here, this is the screenshot of the same program, which I just showed you. And now if you try to see here, that when I am trying to use here, this program, the first value here is n and second value here is rep. So, when I tried to just a user the sample of size 10 and I generated only four or five times then I get here the values of here $\hat{\mu}_1$ like this $\hat{\mu}_2$ like this $\widehat{\sigma_1^2}$ like this and $\widehat{\sigma_2^2}$ here like this.

And you can see here that these values are close to μ equal to 10 or and σ is equal to 25 or not. So, you can see here for this $\hat{\mu}_1$ this value here is 10.26, which is pretty close to 10 and the value of $\widehat{\sigma_1^2}$ this is close to 25, but for the median case, the estimated μ_2 is pretty close to 10, and but the variance σ^2 is very different from this 25.

And if you try to repeat this observation, and instead of 5 times getting the sample you try to take 50 samples, and you try to compute the same thing, you can see here the values in the mean. They are not changing μch, the values in the variance estimator $\widehat{\sigma_1^2}$ they are not changing μch. The values of the median that is also not changing μch, but the variance σ^2 is still very different from the true value.

And similarly, if you try to generate 5,000 samples from the same population, and you try to compute the same thing, you can see here that the values of here $\hat{\mu}_1$ they are not changing μch. The values of here, $\widehat{\sigma_1^2}$ they are also not changing μch or they are not deviating μch from the true value. Whereas, if you try to compute the estimate the $\hat{\mu}_2$, these values are also close to 10 but the values of $\widehat{\sigma_2^2}$, they are very far away from the true value 25.

(Refer Slide Time: 28:38)



And similarly, if you try to increase the sample size, and you try to make the sample size to be here 100 and if you try to repeat it for say 50 times that mean 50 sample 500 samples and 5,000 samples still you can see here that the values of this $\hat{\mu}_1$, they are pretty close to 10, the values of σ^2 they are also pretty close to the true value 25, the values of $\hat{\mu}_2$ is also close to the true value 10, but the values of σ^2 they are very different from the true value 25.

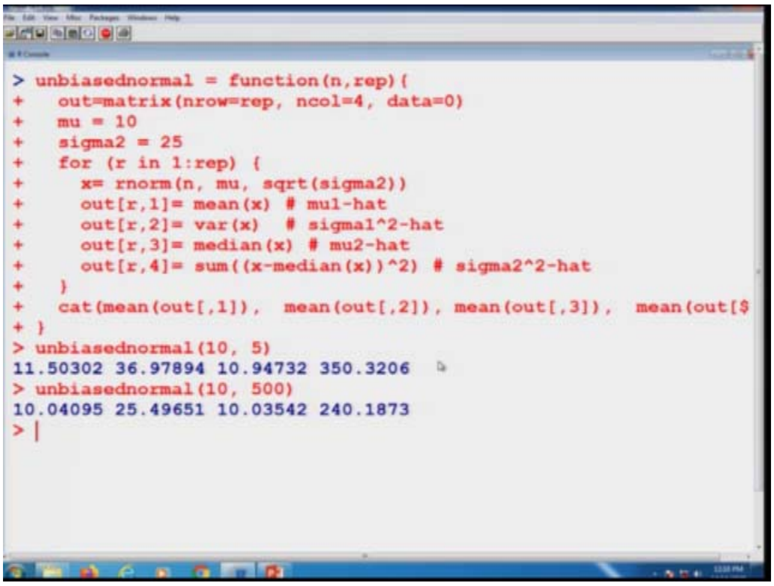
The reason for this I can explain you theoretically also. The reason here is that in a normal distribution, the values of the mean and median they are actually at μ. And the estimator, this σ^2 actually this is an unbiased estimator of σ^2. That is why you can see here that these values are very close to the true value.

Now, the confusion here is between this sample mean and sample median, because μ is the value of sample mean, as well as the indicating the value of sample median. Because in a symmetric curve mean and median are in the normal curve they are lying at the same point. So, μ is also indicating the mean and μ is also indicating the median. And since the distribution is symmetric, that is why the value of mean and median they are coming out to be nearly the same.

And it is difficult to judge from the simulation that whether mean is an unbiased estimator of μ or median is an unbiased estimator of μ. The fact is this that mean is an unbiased estimator of μ whereas, median is not an unbiased estimator of μ, but sample median is the biased estimator of μ. But by looking at this numerical results, you cannot really see with 100 percent confidence unless and until you have done the algebra or the mathematical proof.

So, that is why these mathematical proofs are required. But definitely when you are trying to deal with a complicated condition, complicated situation, then definitely this siμlation technique also give us a fair idea. So, means, I just want to show you that whether this program is really working or not.

(Refer Slide Time: 31:09)



So, I tried to copy and paste this program you can do it from the slides yourself into the R console, and I try to paste it here. So, and then I try to execute this program, which I have shown you the outcome over here. For example, you can see here if I try to make it here, this program here.

So, you can see here once you try to execute it, this is trying to give you here these values. Yes, you can see here the first value is here the means the $\hat{\mu}_1$ second value here is the $\widehat{\sigma_1^2}$ third value is here the here $\hat{\mu}_2$ and fourth value here is the $\widehat{\sigma_2^2}$. And similarly, if you try to repeat it for say 500 times, then you can see here these values are operating here like this.

So, well indicating or getting these values from the execution of the program is not a difficult job right. But, my idea was very simple, that I have introduced you the theoretical construct of our statistical property. Well, in case if you want to be 100 percent confident you have to depend only on the mathematical theory. And this situation can be seen in real life, very often.

Suppose, some of your colleague comes to you and that person does not know statistics µch, but that person is good in siµlation. And suppose, that fellow suggests that, the population is normal and one can estimate the mean by arithmetic mean, median or say mode that means, get a sample, try to compute the either the automatic means of the sample values or the median of the sample values or the mode of the sample values.

Now, when he is trying to do the siµlation with this type of programming, then he will get the same values, nearly the same values. Some value will be 10 some value will be 9.9 or 10.1. So, practically it is very difficult for us to judge whether these three estimators are unbiased or not. So, now, under these type of conditions when you really want to be confident, you have to take the help of theory, you have to find out the expected value of sample means, sample median and sample mode.

And in case if you are getting the value as µ, then you can tell your friend well, these three are unbiased, but this is not going to happen. Only sample mean will come out to be an unbiased limited and sample median and sample mode that are going to be biased. That is a different thing that the amount of bias is going to be less and this is a very common thing which is happening in the data science.

And that is the basic difference between the two. And that is why we are here to understand that what is the connection between the statistics theory and data science and its application in the data science both have to go along, both have to go together, they cannot be separated. And if they are separated as I said in the beginning, that you will not become a doctor but you will become only a compounder.

So, that was my idea. This is how I wanted to present this topic. But definitely, there are many more distribution, there are many estimators. I would request you that you try to pick up a book, try to look into the problem, try to take the problems from your assignment and try to solve them. And I am sure that now you should be convinced that why theoretical foundation of statistics is needed to understand the intricacies involved in the data science.

So, you try to think about a try to practice try to practice. And I will see you in the next lecture with more details on consistency of the estimator. Till then, goodbye