Essentials of Data Science with R Software – 1
Professor. Shalabh
Department of Mathematics & Statistics
Indian Institute of Technology Kanpur
Lecture No. 57
Efficiency of Estimators

Hello friends welcome to the course, Essentials of Data Science with R Software – I in which we are trying to understand the basic concept of probability theory and statistical inference. So, you can recall that in the last lecture, we had understood the concept of unbiased estimators. And towards the end of the lecture, I had shown you that there are four possible estimators which are trying to estimate the population mean of a normal population, and all those four estimators were unbiased estimator for the population mean μ.

So, the question was out of them, which one we have to choose? And in order to find the answer of such a question, we need to have some more criteria. That is the same thing that suppose in a class there are two students who have got this 100 percent marks and then we have to decide that who is going to be at the first position and who is going to be the second position so some time we make a rule that if two persons have got this same marks, then we will try to compare the marks in their particular subject.

For example, say mathematics. And even if the mathematics marks are same, then we go for physics or chemistry and so on. So, we try to impose some conditions. That is the same thing that in a soccer game also in the football game also if both the teams are striking the same number of goals at the end of the game, we try to give them the penalty strokes, and the team which strokes more goal that team is said to be the winner.

Similarly, here in the statistical inference also we need to impose some more criteria gradually, so, that we get a good estimator, but now, in this lecture, we are going to talk about the efficiency of the estimator. That means, we are going to impose one more criterion of efficiency on these estimators and we will try to choose the estimator which is more efficient, but now, the bigger question comes how are you going to define the efficiency?

Tell me one thing, suppose in an office there are two workers who are trying to do the same job, say typing. How do you say that which of the two typist is a better typist? You simply try to say the type is which is making the smaller number of typing mistakes that person is said to be better than the other typist. And similarly, in general, we always say that a person is efficient that the, if person can do the job in the less time with less number of mistakes.

So, similar concept is there in statistics also, but the question is now, how to measure this efficiency. So, we know the concept of variability. We always want our statistical decision in which the variability should be as small as possible. For example, as earlier I had discussed an example where you have to find out the time taken from your home to your college. Now, you have two options. You get me the value say 20 minutes and you say their variability will be just by 1 to 2 minutes.

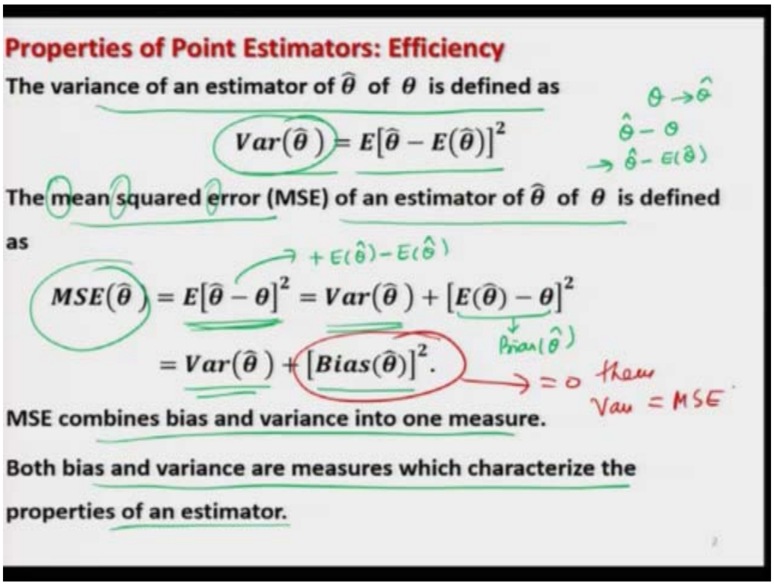
That means, you can take the time between 18 minutes to 22 minutes. But if you say the average value is 20 minutes, but the variability will be say between say 15 to 20 minutes, that means, what? You are trying to say well, the average value is 20, but, the actual value will be something like 15 to 20 minutes less than the 20 minutes and 15 to 20 minutes more than that 20 minutes.

Or similarly, if you try to say that your friend is coming at a particular point, so, sometimes you say that the friend will be getting late by 5 minutes. So, that means if the waiting time is going to be just say, not more than 5 minutes, but if the person says okay he may get or delayed say between say this 30 minutes to say 40 minutes, then that means the whatever the time the person has given to reach that can be less than 30 minutes from that time or after that 30 minute of that time.

What do you think? Which of the statement is better? So, definitely where you are getting less, variability the person says that I will be there within + - 5 minutes or within + - 20 minutes, the statement with within + - 5 minutes is said to be a more efficient estimator or a more efficient value. So, similarly in the case of statistical estimation, we try to compute the variance of these statistics and these statistics are the random.

And since, these estimators are going to be the functions of the random variable, so, that is why they are statistics and so, we can find out there variants also. So, now, that is the topic which we are going to discuss in this lecture. And once we try to impose the criteria of this smaller variability after that, some more topics will come that I will try to discuss in this lecture. So, let us begin our lecture.

(Refer Slide Time: 06:00)



So, now, we talk about efficiency of estimators. So, now, you can recall our discussion, when we discussed the development of the concept of variability. We had estimated a parameter θ by $\hat{\theta}$. Now, there are two options. I try to measure the division of $\hat{\theta}$ from θ like this $\hat{\theta}$ - θ or if I try to measure the division of $\hat{\theta}$ from θ in terms of its mean value. So, I tried to consider here $\hat{\theta}$ - E($\hat{\theta}$).

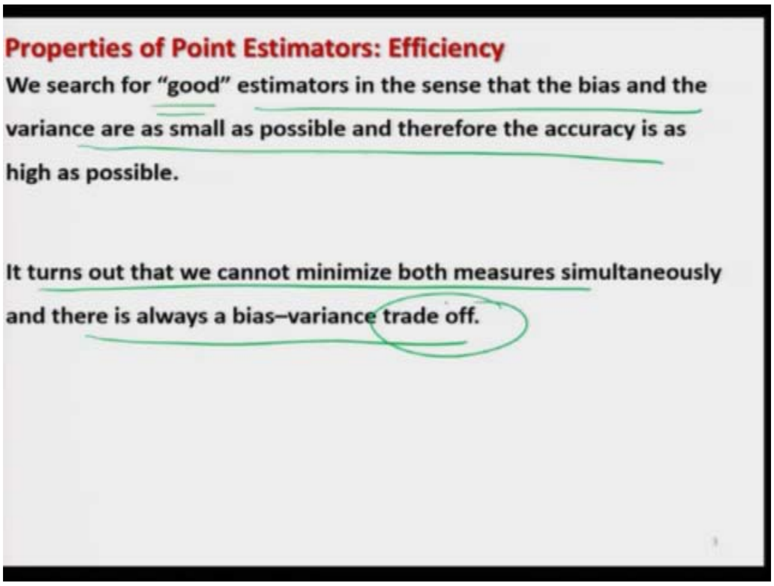
So, now, using these two concepts, we have two measures of variability one is here, variance and another here is mean squared error. So, we define the variance of an estimator $\hat{\theta}$ as the variance of θ equal to $E(\hat{\theta} - E(\hat{\theta}))^2$, and mean squared error of an estimator $\hat{\theta}$ is defined as $E(\hat{\theta} - \theta)^2$.

So, now means taking expectation with respect to the mean value E($\hat{\theta}$) or θ that is not a very difficult job for you that you can do very easily, but this is how we define the two concepts of variability when we are talking of the efficiency of the estimator that the estimator has been obtained as $\hat{\theta}$, now we want to measure its variability. So, remember one thing that you can estimate the variance of $\hat{\theta}$ or you can also estimate MSE of $\hat{\theta}$ depending on the situation what we really want to do, but these are the two measures about which we are going to talk about.

So, definitely when you are talking of the mean squared error or that is commonly called as MSE. So, this M is for mean S is for squared and E is for error. So, this MSE of $\hat{\theta}$ is defined like this $E(\hat{\theta} - \theta)^2$. And if you try to just add and subtract here the $E(\hat{\theta})$ inside the bracket and if you try to expand it, you can write down here very clearly as the $Var(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2$. And if you try to identify what is this thing, this is nothing but your bias of $\hat{\theta}$.

So, the MSE or $\hat{\theta}$ can be expressed as the $Var(\hat{\theta}) + [Bias(\hat{\theta})]^2$. So, this MSE combines bias and variance into one measure. And you know that this bias and variance both are the properties of an estimator that are trying to characterize the estimator.

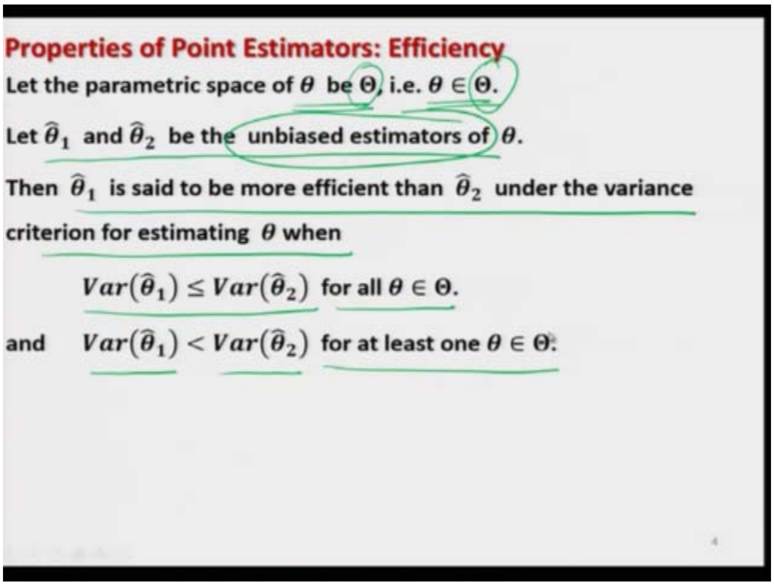
(Refer Slide Time: 08:43)



And now, we are looking for those estimators or we are going to call any estimator as a good estimator whose bias should be as smallest possible and the variability or the variance should be as small as possible. So, we search for say quote unquote, good estimators in the sense that bias and the variance are as small as possible. And once you do it and the accuracy of the estimator becomes as high as possible.

But when we try to do such exercise mathematically, then it turns out that we cannot minimize both the measures simultaneously, that is the bias as well as variance, both of them cannot be minimized at the same time. And so, there is always our bias various trade-offs, that means, you have to strike a balance between the amount of bias in the estimator and the variance of the estimator.

(Refer Slide Time: 09:41)

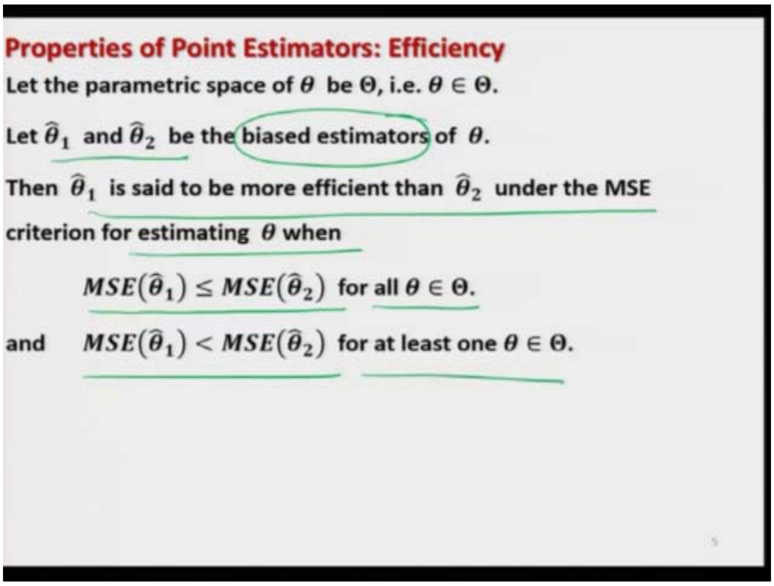


Now, the question comes that, how are you going to define that which of the estimator is going to be more efficient. So, suppose there are two parameters or a parameter θ, and this θ has the parametric space, which is here Θ like this one. So, $\theta \in \Theta$. So, be careful about my wordings. When I say theta, θ means small θ. And when I have to speak out for this Θ then I will always say Θ.

So, now, suppose we want to estimate this parameter θ and suppose $\hat{\theta}_1$ and $\hat{\theta}_2$ are the two estimators of θ and support both are unbiased. So, then in this case $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ under the criteria of variance for estimating the parameter θ when $Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$ for all $\theta \in \Theta$ and $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$ for at least one $\theta \in \Theta$.

So, in general you are trying to say that any estimator whose variance is going to be less that will be called as more efficient than the other estimator whose variance is higher than this.

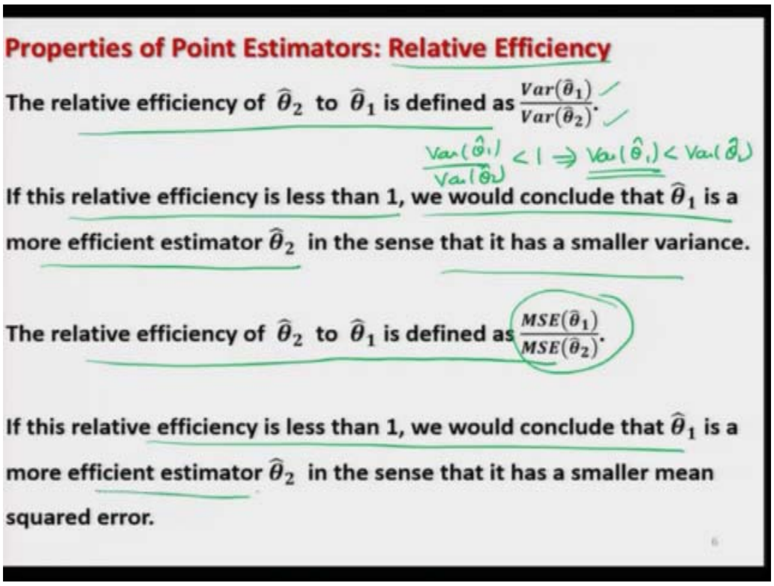
(Refer Slide Time: 11:09)



Now, the same definition can also be given in terms of the mean squared error of $\hat{\theta}_1$ and $\hat{\theta}_2$. But in this case, we assume that let $\hat{\theta}_1$ and $\hat{\theta}_2$ be the biased estimators of θ.

So, now in this case $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ under the criteria of mean squared error or we call it say MSE criteria for estimating the parameter θ when $MSE(\hat{\theta}_1) \leq MSE(\hat{\theta}_2)$ for all $\theta \in \Theta$ and $MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$ for at least one $\theta \in \Theta$.

(Refer Slide Time: 11:53)

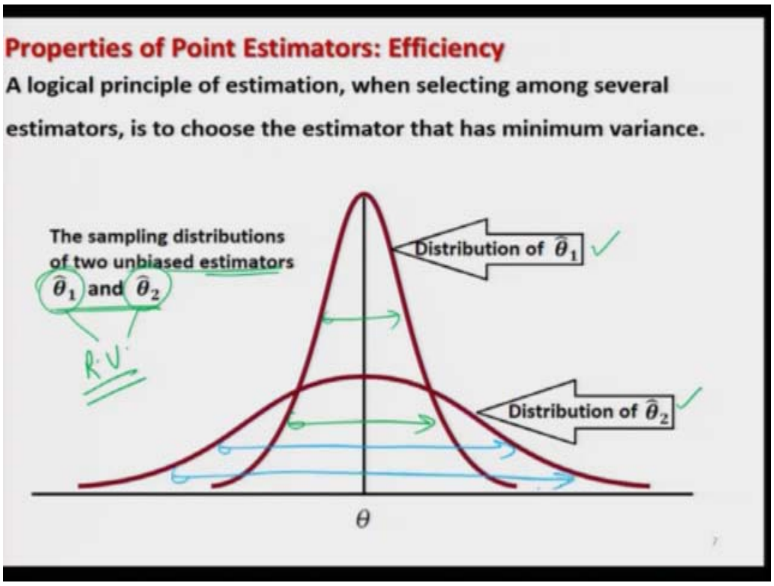


So, that is what we try to do. And another way to measure the efficiency of estimator is the relative efficiency. That mean, you try to compare the variances or MSEs of $\hat{\theta}_1$ and $\hat{\theta}_2$. So, the relative efficiency of $\hat{\theta}_2$ to $\hat{\theta}_1$ is defined as $\frac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)}$. And obviously, if this relative efficiency is less than one that means I am trying to say here, variance of $\hat{\theta}_1$ is less than variance of $\hat{\theta}_2$. That means, if this ratio is less than 1 that mean, it is implying that variance of $\hat{\theta}_1$ is less than variants of $\hat{\theta}_2$. So obviously, this $\hat{\theta}_1$ becomes more efficient.

So, now, we can say that if this relative efficiency is less than 1, you would conclude that $\hat{\theta}_1$ is a more efficient estimator than $\hat{\theta}_2$, in the sense that it has a smaller variance. And similarly, if you want to define the relative efficiency in terms of MSE, then the relative efficiency of $\hat{\theta}_2$ to $\hat{\theta}_1$ is defined as exactly in the same way. You simply have to replace it variance by MSE.

So, that will become $\frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}$. And if this relative efficiency is less than 1, you would conclude that $\hat{\theta}_1$ is a more efficient estimator, than $\hat{\theta}_2$ in the sense that it has a smaller mean squared error.

(Refer Slide Time: 13:23)

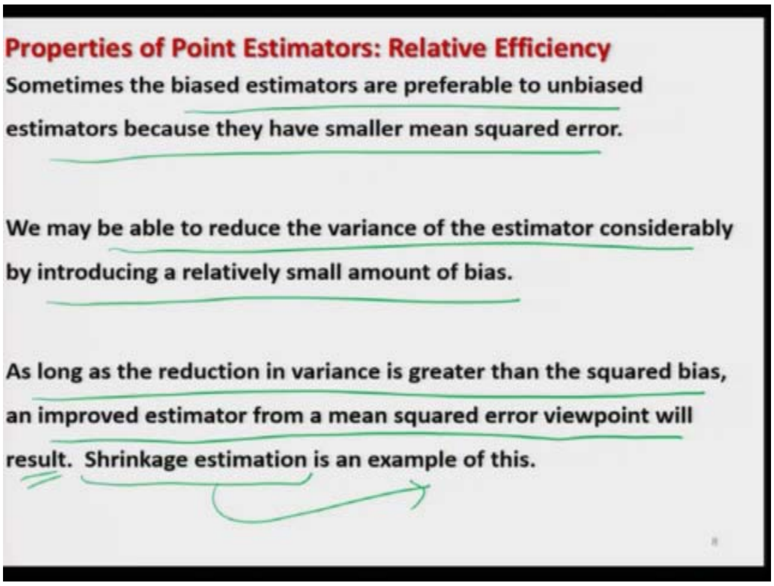


So, now, in case if you try to see graphically what will really happen? That suppose we have got suppose, two estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$. So, now both the $\hat{\theta}_1$ and $\hat{\theta}_2$ suppose they are unbiased estimator, obviously, they are random variable, because they are the function of random variables so they are a statistic.

So, now, once they have got a random variable, then these statistics will also have a probability distribution, and you have done such. For example, when you said that let X_1, $X_2, \ldots, X_n$ they are coming from normal μ σ^2. And suppose they all are IID identically in and independently distributed, then the distribution of the sample mean is also normal with mean μ and variance σ^2/n. So, similarly, you can think in general that all estimators will also have a sampling distribution.

So, suppose, I try to plot sampling distribution of the two statistics $\hat{\theta}_1$ and $\hat{\theta}_2$ which are being used to estimate the same parameter θ. Suppose their curves looks like this. So, you can see here that in the case of $\hat{\theta}_1$, this variability, which is indicated by the curve that is much, much smaller than the variability, which is exhibited by the sampling distribution of $\hat{\theta}_2$. So, this is how you can see that this criterion can be looked upon.

(Refer Slide Time: 14:58)

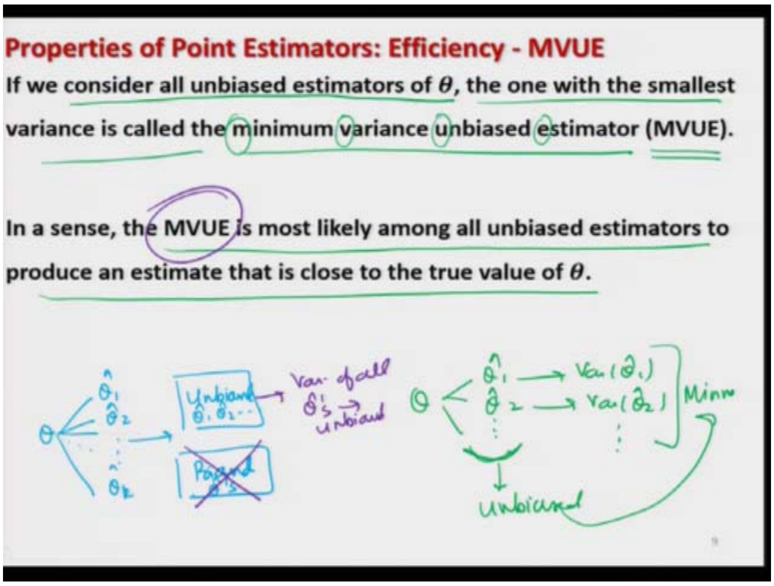


In fact, when we are trying to talk about this variance and MSE, then you have seen that there is a specific condition that variance is defined when the estimator is unbiased and MSE is usually defined when the estimator is biased. And obviously, if you try to look into the expression of this MSE here, you can see very clearly here that in case if this bias becomes 0, if this is equal to 0, then variance is equal to MSE, that means MSE and various both become the same.

So, that is why we have these two definitions. But what I want to tell you here that in practice, many times we are dealing with some situations where the bias estimators are preferable to unbiased estimators, because they have a smaller mean squared error. And suppose we want to have an estimator which has got our lower variability and the criteria of biasness is not that important in a given situation. So, in those situations, this type of concept is useful, and we may be able to reduce the variance of the estimator considerably just by introducing a relatively small amount of bias.

And as long as the reduction in the various is greater than the squared bias, an improved estimator from a mean squared error point of view will result, right. For example, you might have heard the name shrinkage estimation. The shrinkage estimation as an example of this type of estimation that we try to strike a balance between the bias and the variability and we get an estimator which is biased, but that has a smaller variability that than the variability of an unbiased estimator, but anyway, we are not going to talk about this concept here, but surely, I would like you to know that that there exists such estimator.

(Refer Slide Time: 17:11)



And in case if we are considering only the unbiased estimator of data. Suppose there is a parameter θ and there are suppose more than one estimators, which can estimate it like this one, then what we will try to do that, we will try to find out here the variance of each of this estimator. Variance of $\hat{\theta}_1$ variance of $\hat{\theta}_2$ and so on. And then I will try to see that which of these various value is the minimum.

So, what is happening now, we have here all the estimators which are unbiased. And now out of them, we are trying to choose the estimator out of this for which the variability is the lowest, the minimum. This type of estimator, do not you think can we call as minimum variance unbiased estimator? So, this is also an outcome that once you are trying to find out the estimator, you have to impose a condition that any estimator whose variance will be less that will be more preferable.

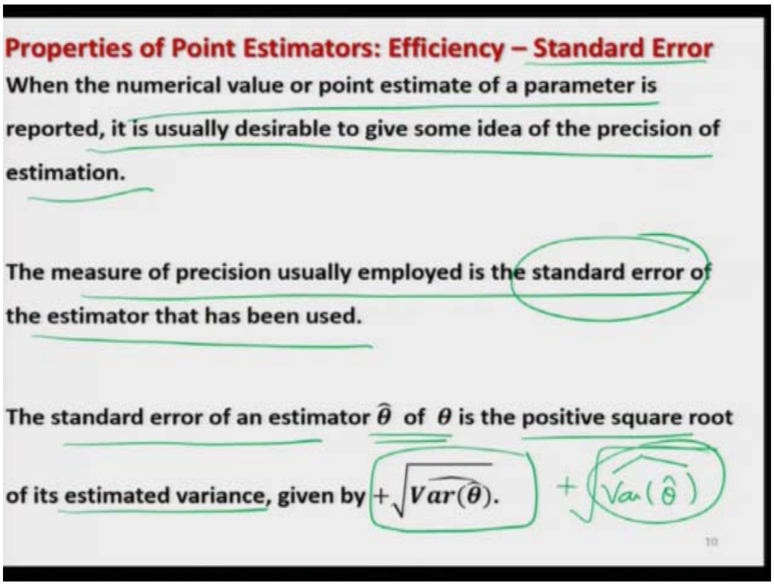
So, if we consider all unbiased estimators of θ, the one with the smallest variance is called the minimum variance unbiased estimator and that is briefly denoted as MVUE M means minimum V means variance U means unbiased and E means estimator. So, in essence, the MVUE is the most likely among all unbiased estimator to produce an estimate that is close to the true value of θ. So, now, what you have done?

Suppose, if I try to explain you here, then suppose we have an unknown parameter θ and then we had here suppose here K possible estimate $\hat{\theta}_1$, $\hat{\theta}_2$ $\hat{\theta}_2$, up to your say $\hat{\theta}_K$

. Now, I tried to divide them into two groups. One are here unbiased estimators and then another is a group of biased estimators. So, definitely there will be suppose more than two estimators in both cases. So, there will be $\hat{\theta}_1$, $\hat{\theta}_2$ and something like this, and here also some $\hat{\theta}$ s are there.

Now, what we try to do? That we ignore the all $\hat{\theta}$ s which are bias and we then we try to find out the variances of all $\hat{\theta}$ s, which are unbiased. And then out of them, I tried to choose the estimator which has got the smallest variance, and this is called as MVUE. And this is the way we are trying to say that when the variance is low, so we expect that the estimated value is going to be more closer to the true value.

(Refer Slide Time: 20:04)



So, whenever in practice we are working and whenever we are trying to estimate a parameter, we always have to report the estimate of this variance that what is the variability which is expected from that given estimator. So, you know that variance is a squared quantity. And for example, if I have a random variable here height, then suppose height is measured in meters. So, variance will be measured in meters square. So, in practice it becomes many times convenient that instead of reporting the values in terms of squared quantities, we want to report them in the same quantities.

So, one option is that if you want to convert this meter square into the same unit meter, one can take the square root. And when we try to find out the estimate of the variance that means, we are trying to find out the value of the variance of the estimator on the basis of given sample of data, and then we try to take a positive square root that is called as a standard error.

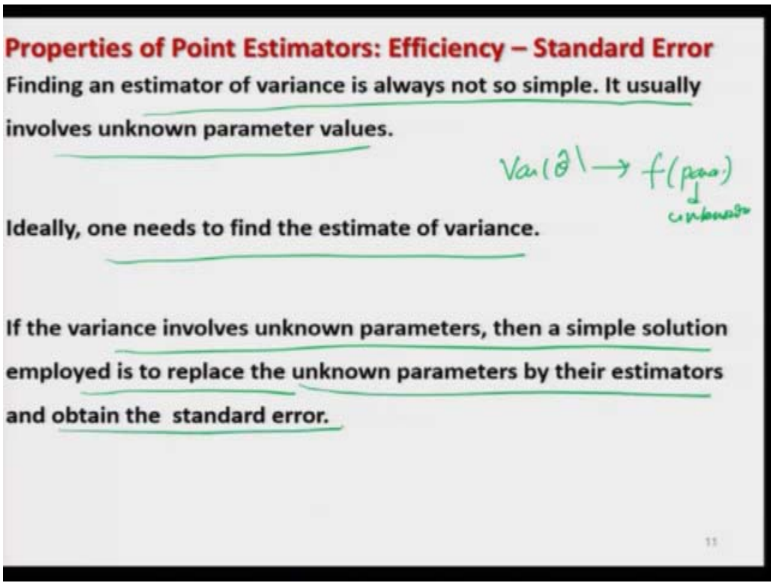
So, when the numerical values or point estimate of a parameter is reported, it is usually desirable to give some idea about the precision of the estimation or the variability of the estimator. So, the measure of precision is usually employed is the standard error of the estimator that has been used.

And then remember one thing we are using here the term is standard error, we are not using the term here a standard deviation that we already have discussed in one of the earlier lecture that what is the difference between standard deviation and standard error. The standard deviation is for the population value and this quantity and this expression will involve some population parameter, but the standard error is entirely based on the sample values.

So, if I am trying to estimate a parameter θ by $\hat{\theta}$, then the standard error of $\hat{\theta}$ is the positive square root of its estimated variance which is given by like this is square root of estimator of variance of θ and its positive square root. And you see, you are trying to find out here the value of variance of $\hat{\theta}$ and you are trying to estimate it.

So, we know that when we are trying to estimate it, I have to put here a hat. So, that is why I am trying to put here up big hat so that, you can know that this entire quantity is estimated, and then I am trying to take a square root that will have two possibilities + sign and - sign, so, I am trying to take care of the positive sign.

(Refer Slide Time: 22:45)



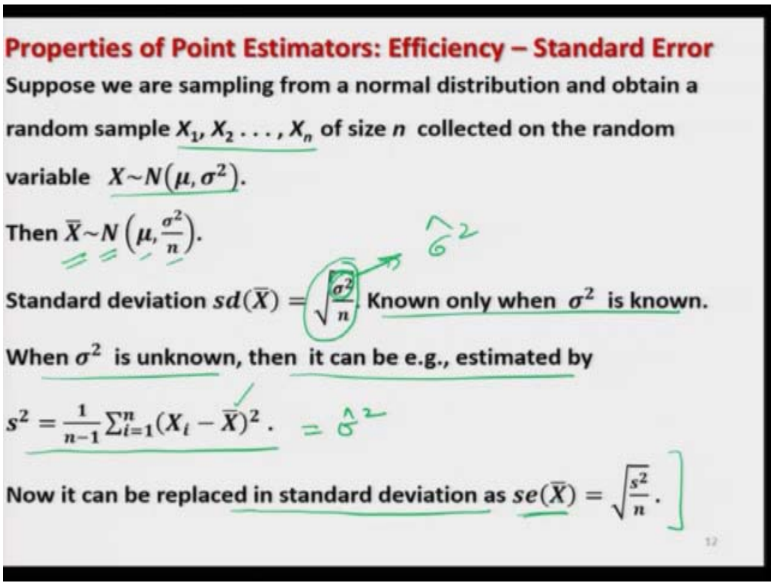
So, this is a very nice, expected, desirable value that any experimenter would like to have in real life, but surely finding an estimator of the variance is always not so simple and straightforward, because that may involve some unknown parameter value also. Because whenever you are trying to find out the variance of $\hat{\theta}$, that is going to be a complicated function and that may involve some parameters which are unknown. So you just cannot estimate it.

But in real life, whenever we are trying to work, we always want to have it because that is the same situation that somebody goes to a doctor, and the and if the doctor says, no, we don't have the medicine for this disease then it is not a very nice thing to say, and nobody will be happy. But if the doctor says we do not know about the right medicine for this disease, but there are some possible medicines and we are trying to use them.

So, it is possible that that if the ailment can be cured in say 5 days of time in case if the medicine would have been developed, but now the element is going to be say cured in say, 3 more days. So, instead of 5 days, it is assured that it will get cured in 8 days. So now, if I ask you what is a better option? That the doctor do not give the medicine, that there is no medicine for this ailment or the doctor tries to give you some good medicine or the doctor tries to give the patient some good medicine that is expected to cure it and say in a couple of days more.

So, that is not going to be as efficient as the true medicine, but definitely that is not going to be a bad solution also. So ideally, one always needs to find the estimate of the variance, but if the variance involves unknown parameter, then a simple solution with that is employed in real life is to replace the unknown parameter by their estimators and obtain the standard error. So this is something like a feasible version. So you may not get here the exact value, but you will get a nice or good approximate value.

(Refer Slide Time: 25:03)



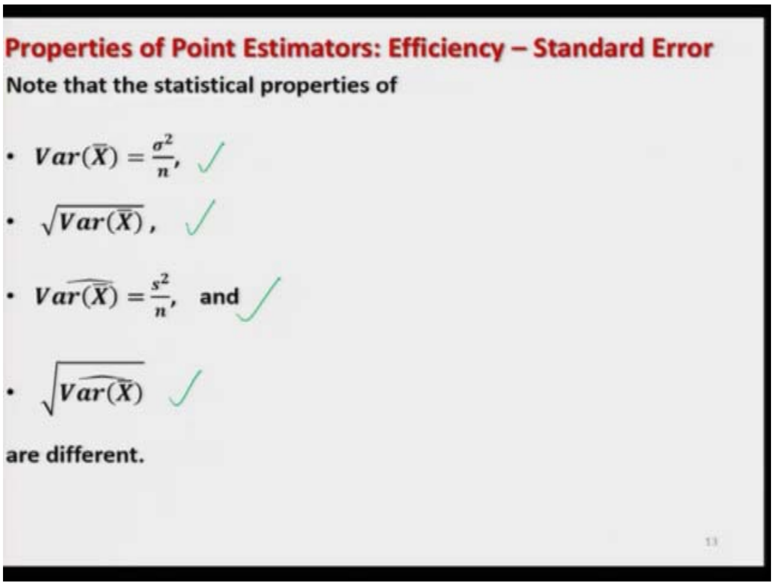
So, now, let me try to take a simple example to explain you this feature that suppose that we have a random sample $X_1, X_2, \ldots, X_n$ of size n from $N(\mu, \sigma^2)$ then we know that the sample mean of $X_1, X_2, \ldots, X_n$ will follow our normal distribution with mean μ and variance σ^2/n, and that we already have done.

Now, the standard deviation of $\bar{X}$ is defined here as the square root of σ^2/n. So, now, you can see here there is a quantity which is here σ^2 involved. So, now, this standard deviation is going to be known only when σ^2 is known to us and in practice usually this will not be known to us because we will be working only on the basis of a random sample.

So, one option is just that when σ^2 is unknown, then it can be estimated by for example, this quantity $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Yes, the question is how this quantity is coming that we will try to see when we are trying to find out the estimator of say mean and variance in the case of normal population in the forthcoming lectures. But at this moment, you can just take it as this is one possible choice for estimating the value of σ^2. So, this is actually like $\hat{\sigma}^2$.

So, now, what we can do we can replace this σ^2 by this quantity σ^2 by here $\hat{\sigma}^2$ and then we have this quantity here square root of s^2/n this is the standard deviation which is obtained after replacing the unknown σ^2 by $\hat{\sigma}^2$ and this is called as the standard error of $\bar{X}$.

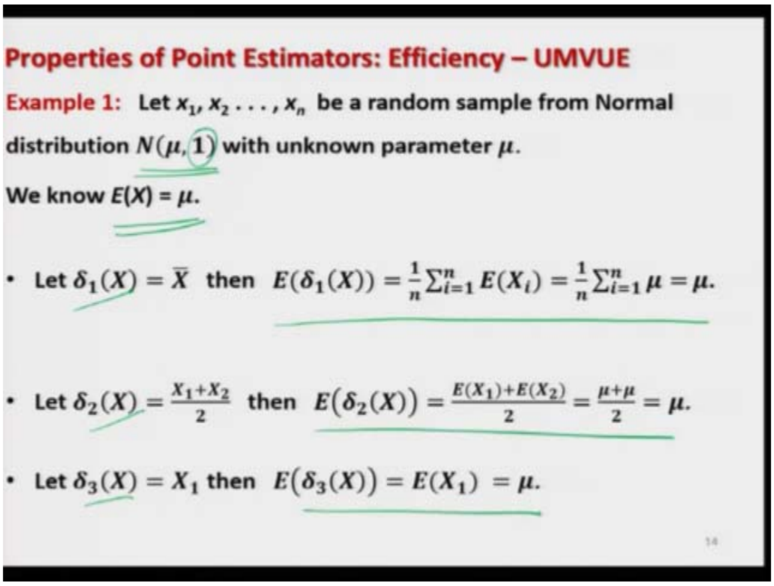
(Refer Slide Time: 26:50)

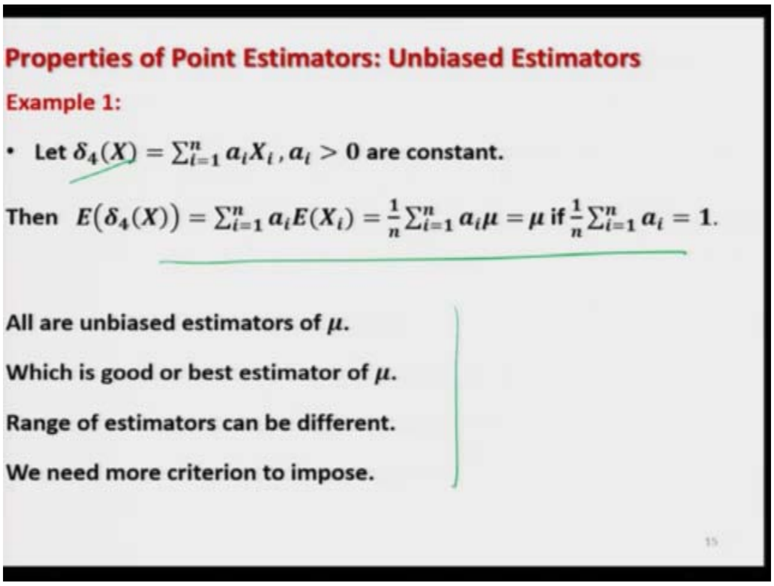


But now, the problem is that the statistical properties of variance of $\bar{X}$ is the standard deviation estimate of variance of $\bar{X}$ and the standard error they are different, they have got different statistical properties. And having one property does not directly employ the properties of the other estimators. So, that is what you have to keep in mind, but still we are looking forward for us solution.

So, definitely the first choice will be if you can find out the exact value of an estimator of the variance MSE or if not, one possible solution which can be easily employed particularly in data sciences is that you try to estimate all the parameters separately and then you try to replace the estimated value in the place of unknown parameters in the expression of the variance. Possibly this will give you a good value.

(Refer Slide Time: 27:42)

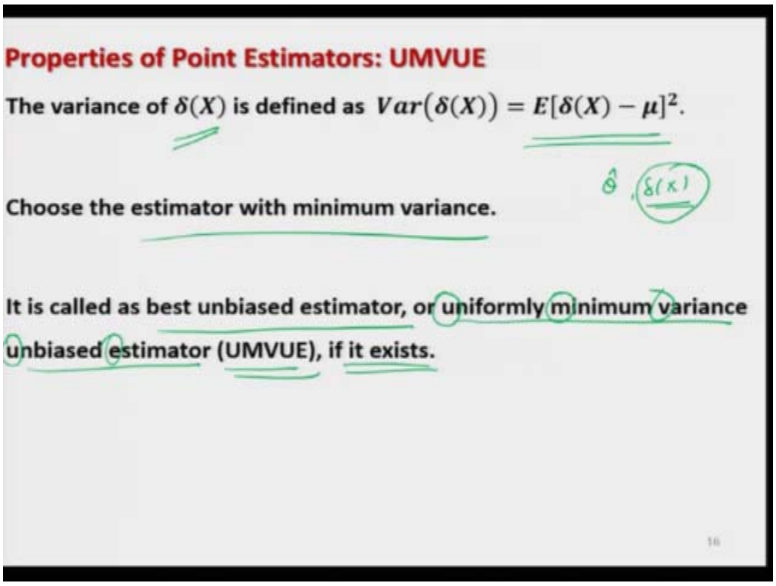




Now, let me try to take here one simple example to explain you these concept. Suppose, we have got a random sample from the normal distribution whose mean is μ and variances suppose known as 1. So, now, we know that expected value of X here is μ and we already have discussed in the earlier lecture that if I try to consider here four possible estimator $\delta_1(X)$, $\delta_2(X)$, $\delta_3(X)$ and here $\delta_4(X)$, then what is going to happen, we had obtained the bias of all of four estimator and we had shown in the earlier lecture that all this estimators are going to be unbiased estimators of the population mean. So, I am not repeating that algebra.

And for now, we have here four estimators of the population mean, and we want to know which one is the best estimator. So, we try to now impose here criteria of various and we try to compute the variance of all these estimator.

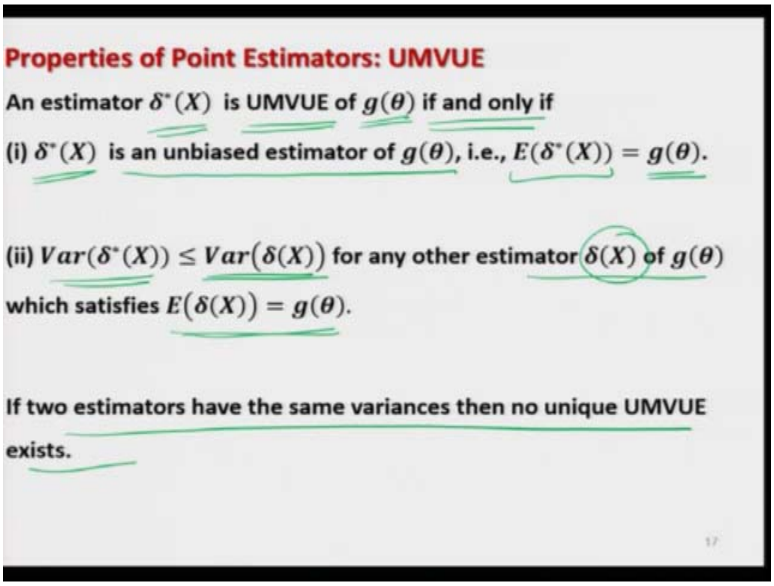
(Refer Slide Time: 28:39)



So, we are going to find out the variance of an estimator say delta X as expected value of delta X - μ whole square, and out of them we are going to choose that $\delta(X)$ for which the variance is minimum, and that is going to be called as best unbiased estimator or that will have uniformly minimum variance and bias estimator, which is briefly called us you UMVUE, U-M-V-U and E they are coming from here. It is not always that UMV always exist.

So, now, let us try to see, what are we going to do here. One thing you may notice here that sometimes I am trying to indicate the estimator of an unknown parameter by $\hat{\theta}$, or sometimes by $\hat{\delta}$. So means I am just using it conveniently so that I can explain you in a better way. For example, when I am using delta X that is clearly indicating that this is a function of X. So, do not get confused.

(Refer Slide Time: 29:46)

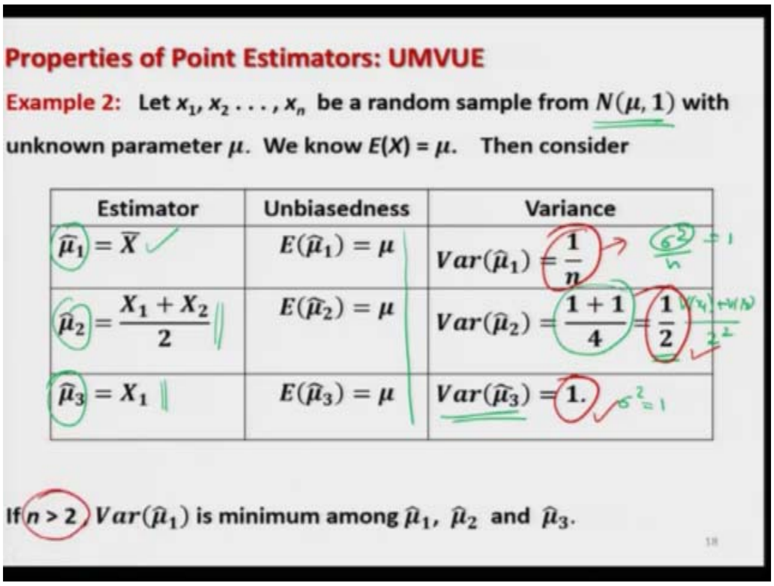


So, an estimator, $\delta^*(X)$ is the uniformly minimum various and estimated of some function of parameter if and only if this estimator say $\delta^*(X)$ is an unbiased emitter of g(θ) that is expected value of $\delta^*(X)$ is equal to g(θ). And variants of this $\delta^*(X)$ is always less than or equal to variants of delta X, where $\delta(X)$ is any other estimator of g(θ). And $\delta(X)$is also an unbiased estimator of θ.

So, what are we trying to do here? That we are trying to simply choose the estimator from a group or a class of estimators in which all the estimators are unbiased estimators of the given parameter and then we are trying to find out their variances, and we are trying to find out that whose value is the least that is which estimator has got the minimum variance and then we are trying to choose that estimator, which has got the minimum variance, and that is going to be called as uniformly minimum various and bias estimator of that respective parameter.

So, this is the definition of UMVUE of g(θ). So, obviously, you are now imposing the word here uniformly minimum. So, that means, if there are two estimators which have got the same variance that means, there cannot be unique uniformly minimum variance and bias estimator.

(Refer Slide Time: 31:23)



So, now, now, in the same example, which I have just considered where we had the 4 parameters just for the sake of understanding I am trying to take here, the 3 estimators and we and now, I am trying to denote them here. So, now, let me try to take here the example which I just consider, and I tried to consider here three possible estimators of the parameter μ.

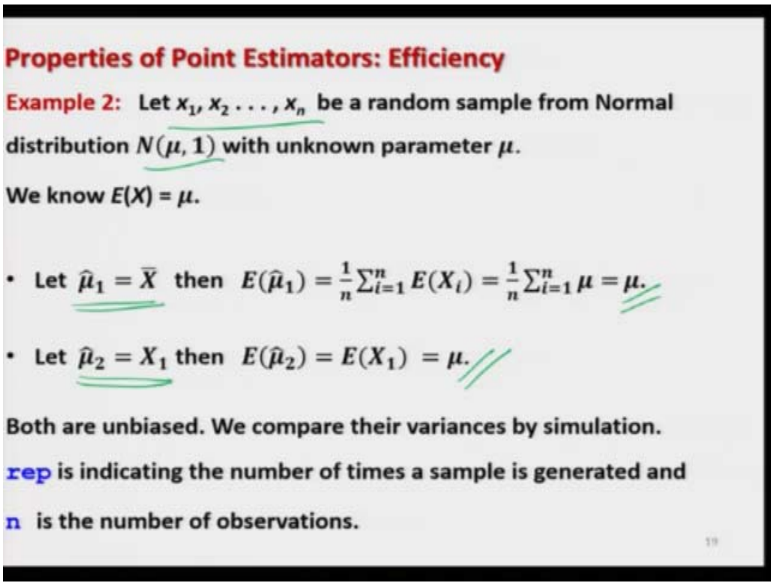
So, we have a sample from normal μ 1 and we want to estimate the parameter μ and for that we are estimating μ by say, $\hat{\mu}_1$, which is the sample mean $\hat{\mu}_2$, which is the mean of the first two observation and $\hat{\mu}_3$, which is the first observation. I am not considering here the $\delta_4(X)$ because otherwise you will get confused because the condition will be on submission ai square also.

So, now, you can see here, we already have found that all the three estimators they are unbiased estimator. Now, we try to find out their variances. So, the variance of $\hat{\mu}_1$ will come out to be 1 upon n because this is σ^2/n and σ^2 here is 1. Similarly, if you try to find out the variance of $\hat{\mu}_2$ this will come out to be here variance of X_1 + variants of here X_2 + covariance will become 0 because they are independent and divided by here 2 square.

So, the variance of X_1 and variance of X_2 they are 1, 1. So, this will become here one + 1 + 1 divided by 4 which is equal to 1 by 2. And similarly, if you try to find out the variance of $\hat{\mu}_3$, this will be simply here σ^2 which is equal to here 1. So, now, you can see here that out of these three values 1/n, 1/2 and here 1. This 1 upon n is going to be smaller than 1/2 and 1 if n is greater than 2. And definitely when you are trying to deal with statistics and data sciences, you expect that the observations will be at least more than 2.

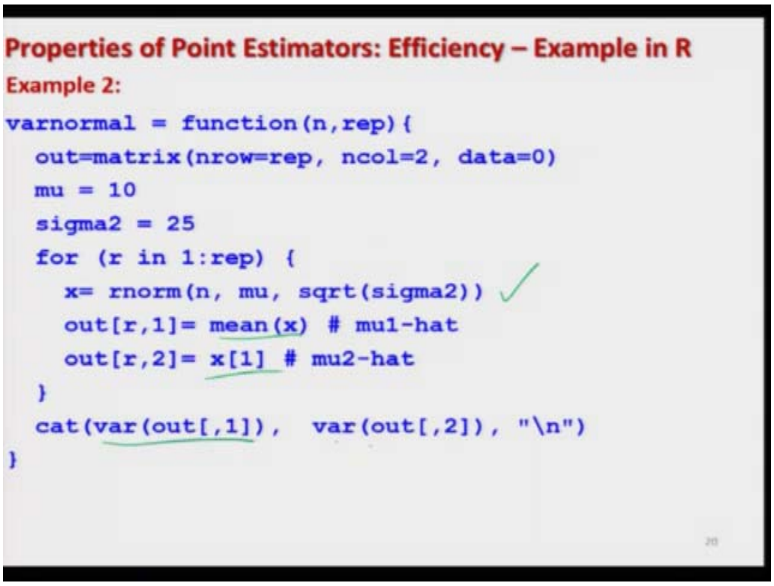
So, this condition is usually going to be satisfied in all the practical applications. And you can see here that for n greater than 2 the variance of $\hat{\mu}_1$ is minimum among the variances of $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\mu}_3$. So, I can see here that this is the best estimator.

(Refer Slide Time: 33:50)



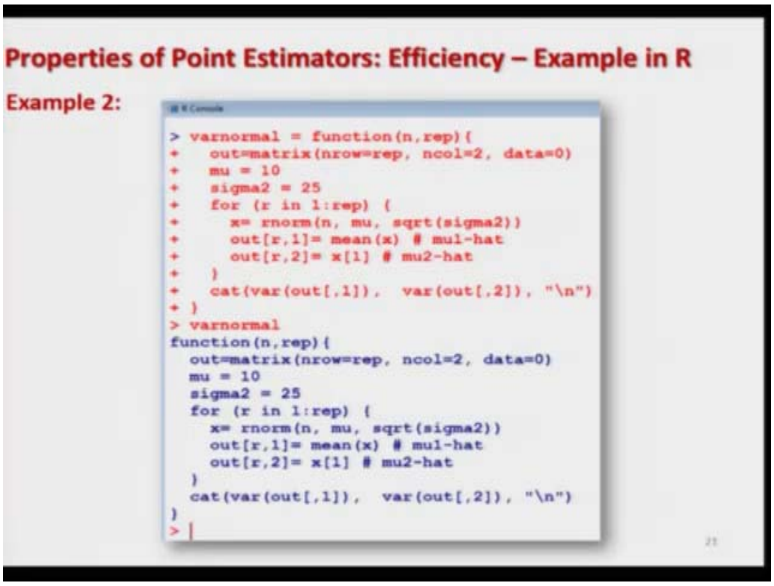
So, now, I try to simulate the same experiment in the R software, just to convince you that how it will look like. So, I try to consider here only two estimators just for the sake of simplicity, and this is here $\hat{\mu}_1$ is equal to $\bar{X}$ and I am considering $\hat{\mu}_2$ equal to X_1. So, both of them are going to be unbiased estimator and I try to generate the random samples $X_1, X_2, \ldots, X_n$ from $N(\mu, 1)$ and I try to generate large number of samples and I try to see what happens.

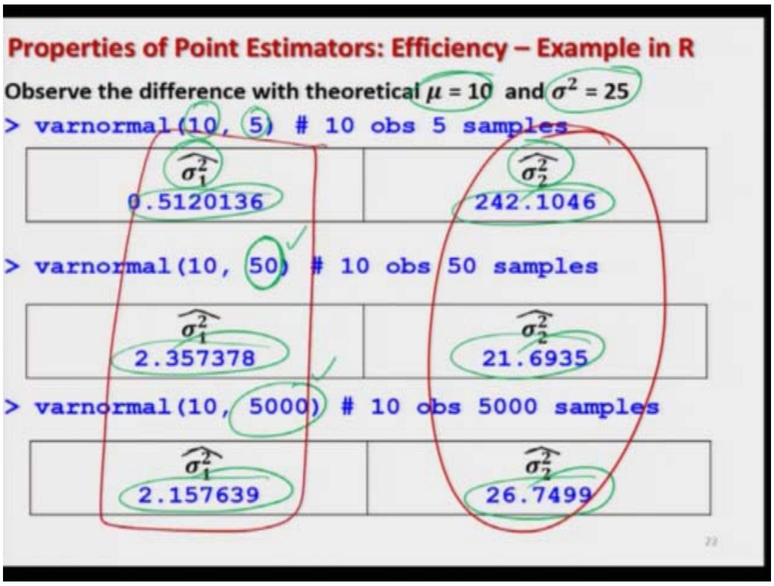
(Refer Slide Time: 34:30)



So, this is the program which you can actually use. So, you can see here it is trying to generate the random numbers from the normal population and then we are trying to find out its say here mean. And then we are trying to find out the from the values of $\hat{\mu}_1$ and $\hat{\mu}_2$, we are trying to find out their variances which are you can see here like this one.

(Refer Slide Time: 34:49)



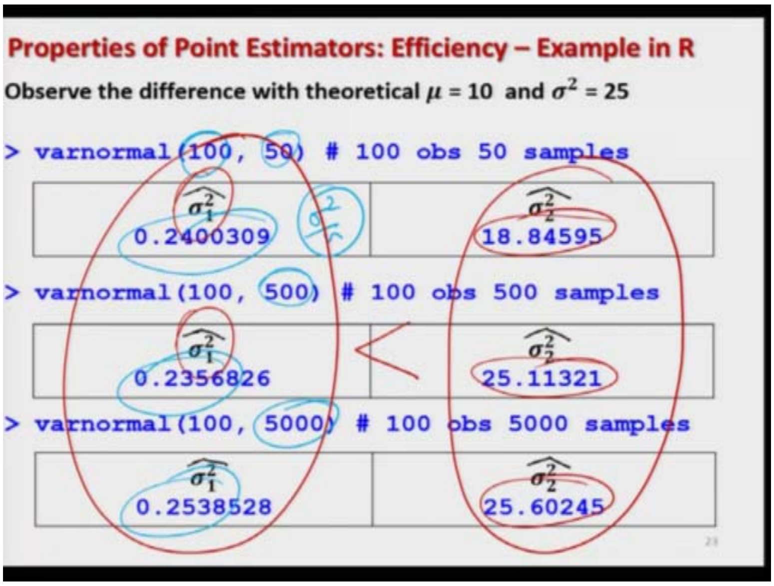


So, now, this is the program here that is the screen shot now, means I already have obtained these values and I would like to discuss it with you. So, you can see here in this case, the true value of μ is 10 and the true value of σ^2 is 25 by which I have generated these observations.

Now, if you try to obtain the value of $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ you can see here when you are trying to just take a sample of size 10 and you are getting only 5 samples here, then the value of $\hat{\sigma}_1^2$ is coming to be close to 0.51, but can it become 2.35 then it becomes 2.15. So, there is a lot of fluctuation even when you are trying to increase the sample numbers that means, you are trying to generate 50 sample or say even 5,000 samples.

And in case if you try to look into the $\hat{\sigma}_2^2$ this value is like 242, 21, 26. So, you can see both the values are fluctuating they are varying, but definitely the variations in the values of $\hat{\sigma}_2^2$ are much, much larger than the values of variance in the case of $\hat{\sigma}_1^2$.

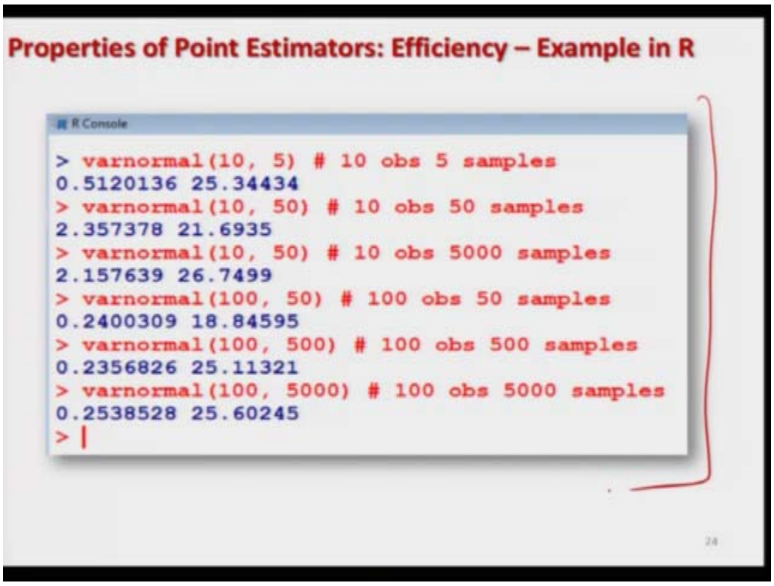
(Refer Slide Time: 36:12)



And same thing you can see here that if I try to increase here the number of observations here 200 and then I try to generate a 50 sample 500 samples and 5000 samples here the values of $\hat{\sigma}_1^2$ are close to 0.24, 0.23, 0.25 because they are going to be determined by σ^2/n. So, these values are much, much closer than the values of $\hat{\sigma}_2^2$ you can see they are close to 18, 25 and 25. So, you can see here that all these variances are greater than the variances that you have obtained using the expression of $\hat{\sigma}_1^2$.

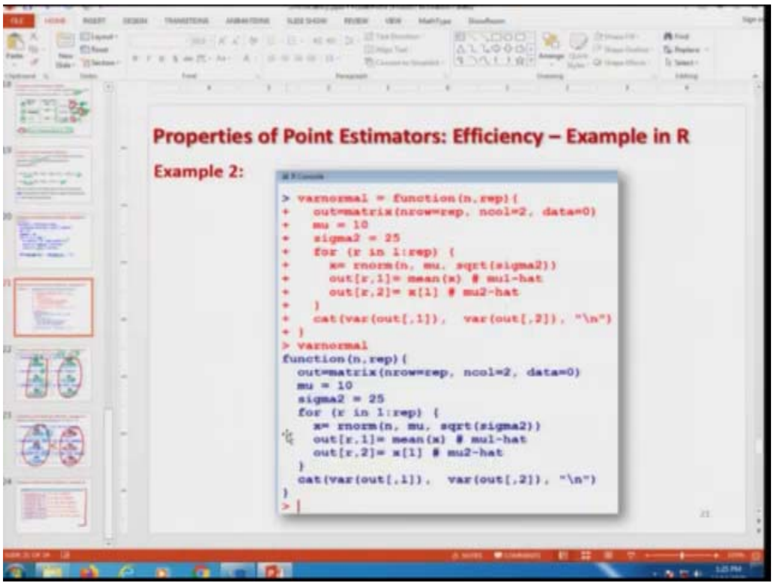
So, this clearly shows you that the variance $\hat{\sigma}_1^2$ is smaller than the values of $\hat{\sigma}_2^2$ in general. And you try to repeat this experiment for a variety of condition whatever you can expect that can happen in the data science, and then try to see possible you will get an idea that how this variance is larger or smaller.

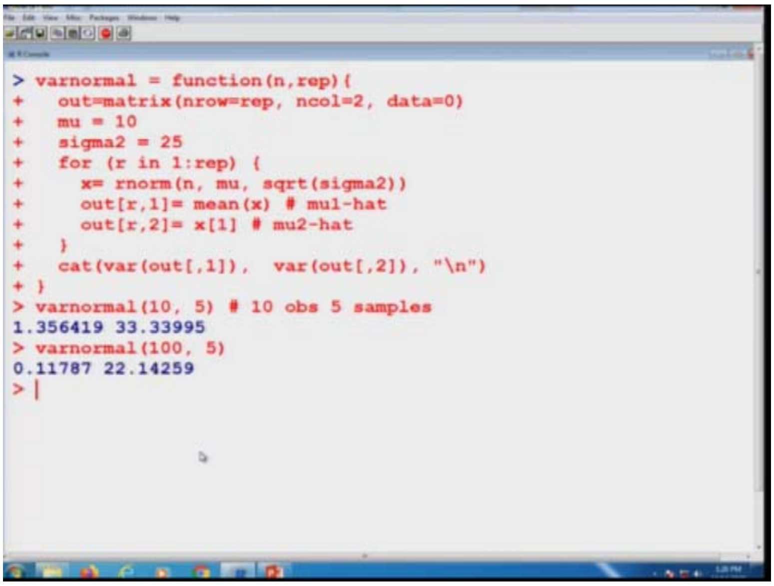
(Refer Slide Time: 37:21)



And this is the screenshot of the same outcome, which I just shown you. Now let me try to show you this thing on the R console and then we are done with this lecture also. So, so, I tried to copy this program on the R console, you can see here, and then I try to execute it.

(Refer Slide Time: 37:42)





You can see here this is like here, this one you can see here, these are the values like this one. And if you try to increase the number of observations over here, you can see here 100 observations this values are like this. So, you can also I mean, I would like that you try to conduct this experiment. And can you try to convince yourself that and try to understand what is happening inside this phenomena?

So, now, it will let me come to an end to this lecture. And I have tried my best to give you the concept of efficiency. And I also have introduced that how this efficiency can be embedded into the estimators, and for that we have defined the minimum variance and bias estimator or say uniformly minimum variance and bias estimator. Yes, the question is still remains that how are you going to find out such estimators. So, that we will try to discuss in the forthcoming lectures.

So, now, you try to have a look try to experiment these things on the R console, try to take an example from the book, try to solve it manually means to using the statistical concept and try to simulate it inside the R software. I promise you, this will give you a wonderful experience and better knowledge than anybody can have it. So, you try to practice it, and I will see you in the next lecture. Till then, goodbye.