Essentials of Data Science with R Software – 1
Professor. Shalabh
Department of Mathematics & Statistics
Indian Institute of Technology Kanpur
Lecture No. 60
Method of Moments

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, up to now, we have understood that what are the good properties of an estimator and when those properties are there, we can believe that this value is going to be dependable and reliable value.

So, now, the question is, how to get that value that you want to compute or in very simple words, how to get an estimator of a parameter? So now, we are going to begin with this topic that how are you going to estimate the parameters on the basis of a random sample. That means, you have got the observation, now, what do you want to do? Do you want to take a sample mean or you want to compute the median or mode, harmonic mean, geometric mean etc what, in case if you want to find out the central tendency of the population.

So, now, the estimate of the parameter depends on the population, and population is characterized by the probability density function. Now, we are going to talk about two methods in this course, one is method of moments, and another is maximum likelihood estimation. I am not saying that these are the only two approaches to estimate the parameters, there are many, many approaches, but these two are the fundamental approaches and they give us a good value.

And in the beginning of this course, when you are trying to create the foundations of this data sciences, then I think that these two are sufficient for you at the moment, but definitely I am not saying that they are the only one, you need to learn more. And now, with the advent of the computer, more computational estimation techniques are there. So, let us try to begin this lecture and try to understand it. But before that, let me ask you a very simple question.

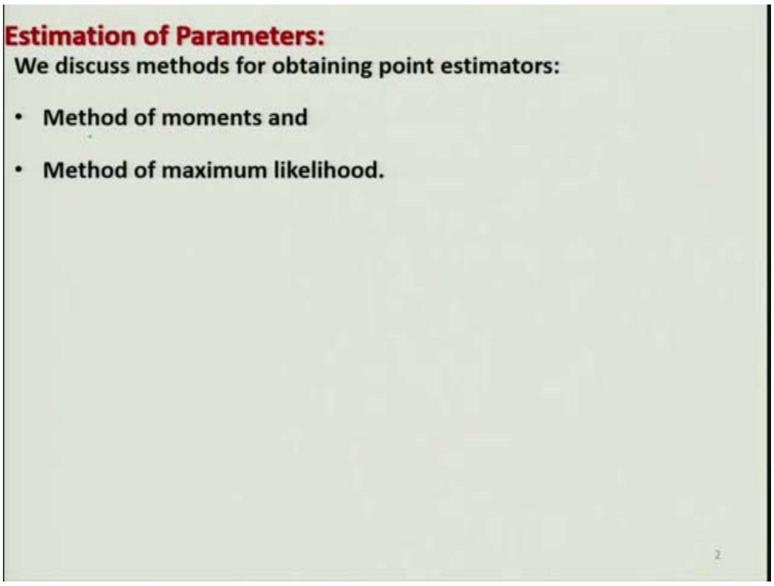
Have you ever gone to a shop to buy some wheat or rice or some food grains? I am sure, yes. Have you seen there is a big heap of the grains there are 100 kg bags for the wheat, rice etc. and you want to see what is there inside the bag or what is the quality of the grain inside the bag, what do you really do? You simply try to take a small sample in your hand and then you try to see how is the quality, and if that small quantity of grain has got some holes that means they have been infected by some insects, then you believe that okay the grain inside the entire bag is also infected by the insects.

And in case, if you find that in that small sample that is the of say 100 grains, if there are an impurity if suppose out of 100 grains, there are 10 pieces of impurity, then possibly do not you think that you zoom without thinking that okay this bag is having almost 10 percent impurities and on the other hand, if your all the greens are good, you assume that, the entire bag of the green is good. Think about it, what are you trying to do? You are looking at the sample and then you are just magnifying it to the entire population.

And whatever are the characteristics inside the sample you are simply assuming that they are the same, which are available in the population also. So, how this is happening? And that is the characteristic, which I am going to use in the method of moments. Now, as soon as I say method of moments, I expect that you remember what we had learned in the topic of moments. Means I had told you that these are the things, which are not going to be used as such but they will help us a lot.

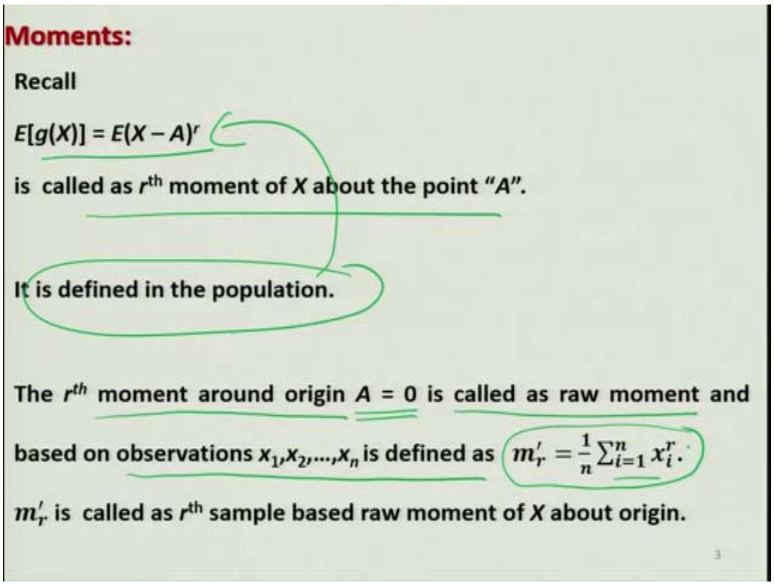
So, now, in case if you have not revised that part, I would request you that you first try to have a quick look on the moments, although, I will give you here a brief information about those moments I will try to repeat it but surely if you understand it from your earlier lectures in detail, that will be helpful. So let us begin with the Method of Moments.

(Refer Slide Time: 05:05)



So, now, as I said, we are going to talk about the method of moments and method of maximum likelihood, method of moments we are going to talk in this lecture and method of maximum likelihood in the next lecture.

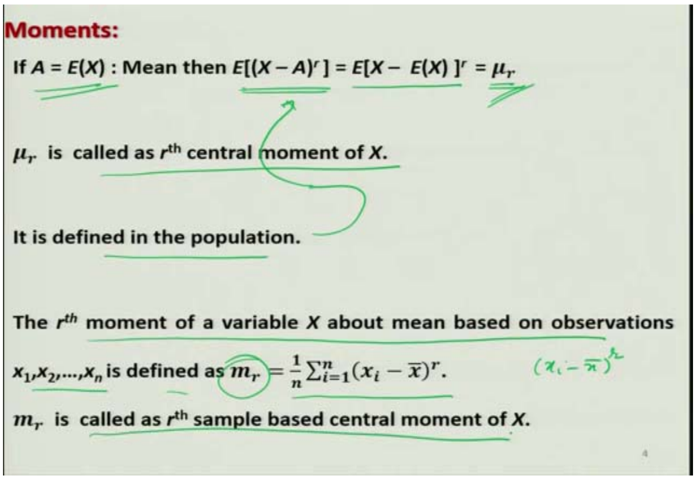
(Refer Slide Time: 05:15)



You can recall that we had considered the expected value of a function of a random variable as $E(X - A)^r$ and this was called as the r^{th} moment of X about the point A, and it is defined in the population. This is what you have to keep in mind. You can see here this is the expected value. And in case if you try to take this capital A point to be 0, then we say that the r^{th} moment around the origin equal to 0 is called as raw moment.

And in case if you try to compute this population value on the basis of the given sample of data as $X_1, X_2, \ldots, X_n$, then we try to define the rth raw sample moment as $m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$. So, this is your here r^{th} sample-based raw moment of X about origin.

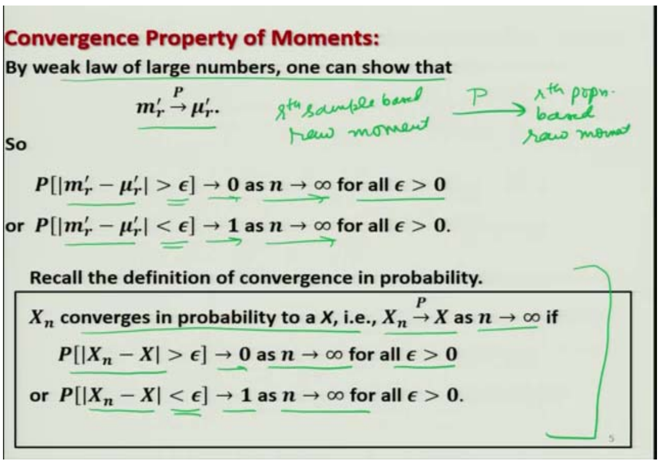
(Refer Slide Time: 06:18)



And in case if you try to take here A is equal to expected value of X that is the mean then $E[(X - A)^r]$ this becomes an $E[X - E(X)]^r$, which is equal to here μ_r and μ_r is called as the r^{th} central moment of X and this is again, defined for the entire population that you can see here, μ is the mean for the entire population for example, and similarly, the μ_r is the mean rth central moment for the entire population.

Now, in case if you want to compute it on the basis of given sample of data, again the rth moment of a variable X about mean based on the observation $X_1, X_2, \ldots, X_n$ is defined as say $m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r$. So, this is only the arithmetic mean of the deviations $X_i - \bar{X}$ with our r and this m_r is called the r^{th} sample-based central moment of X.

(Refer Slide Time: 7:22)



Now, I try to give you here one property, which I am not going to prove but I am just stating it. Do you remember the definition of the convergence in probability that we had used earlier also in that case of consistency of the estimator. So, we try to say that X_n converges in probability to X that is X_n is going to X in probability as n goes to infinity, if $[|X_n - X| > \epsilon] \to 0$ as $n \to \infty$ for all $\epsilon > 0$ $P[|X_n - X| < \epsilon] \to 1$ as $n \to \infty$ for all $\epsilon > 0$.

So, now, based on that, I can inform you that using the weak law of large numbers we can show that $m'_r \xrightarrow{P} \mu'_r$ that means, the rth sample based raw moment tends to the rth population-based rth moment raw moment. And this convergence happens following the rules of convergence in probability. That means, $P[|m'_r - \mu'_r| > \epsilon] \to 0$ as $n \to \infty$ for all $\epsilon > 0$.

And the probability that the absolute deviation between m'_r and μ'_r are smaller than epsilon is going to 1 as n goes to infinity for all epsilon greater than 0. So, this means, if you try to increase the sample size, your rth sample-based raw moment will converge to the rth population-based raw moment.

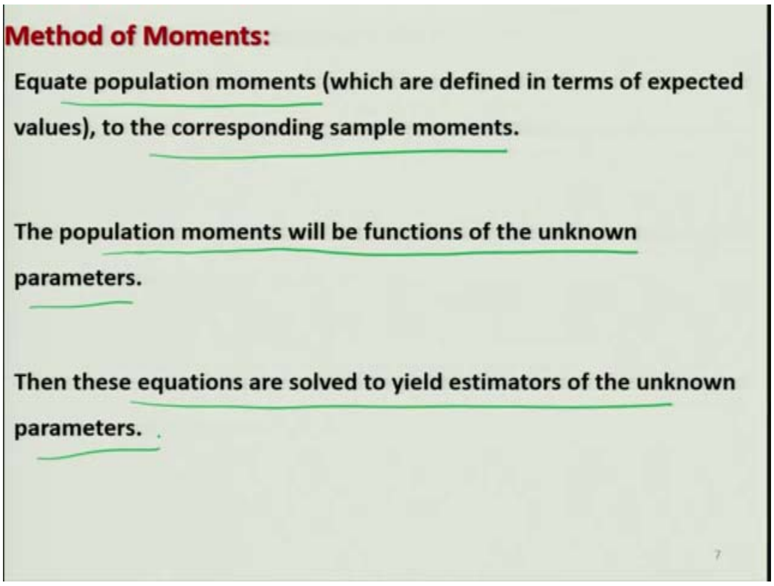
(Refer Slide Time: 09:28)

Now, you have seen this type of picture well, these are the places where people try to store the grains and there are so many bags of in possibly every bag has a weight of 100 kg, but whenever we want to buy the wheat, we will simply take out a one bag and we try to look into the quality of the week by drawing a sample like with this type of instrument or we try to simply use our hand and we try to look into the quality of the wheat.

Now, and then what we try to do whatever is the quality that you are trying to estimate here you try to equate this quality that you have obtained on the basis of a sample to the entire population. So, whatever is the quality in this amount of wheat that you try to translate it to this entire bags. And similarly, if you whatever is the quality that you are obtaining through this small instrument that you try to translate it to this entire population of the bags.

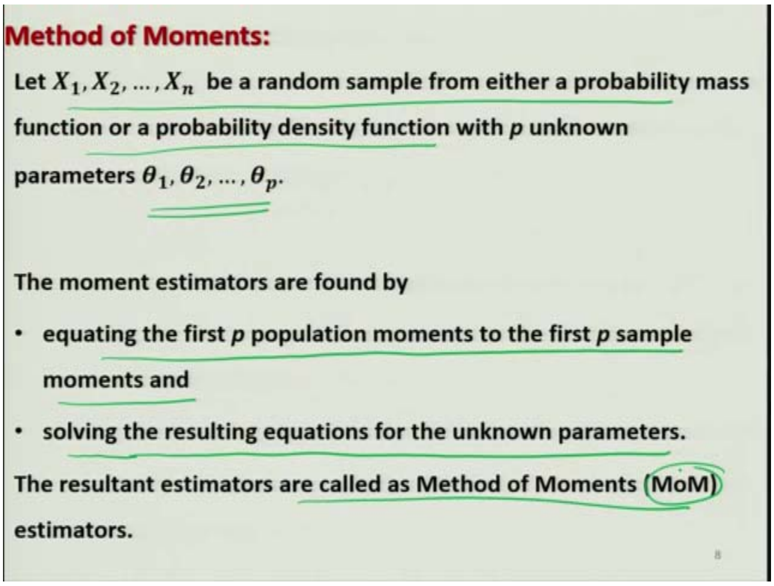
So, that is the concept here. You are trying to assume that whatever are the characteristics, whatever are the properties which are contained in the population, they are also in the sample. And whatever other properties that we have learned they can be prescribed through the moments.

(Refer Slide Time: 10:48)



So, what we try to do here that we try to equate the population moments to the corresponding sample moments, that's all. The population moments they are defined in terms of expected value and they are going to be a function of some unknown parameters. And then we are going to have such equation we try to solve those equations and they will try to yield the estimators of those unknown parameters.

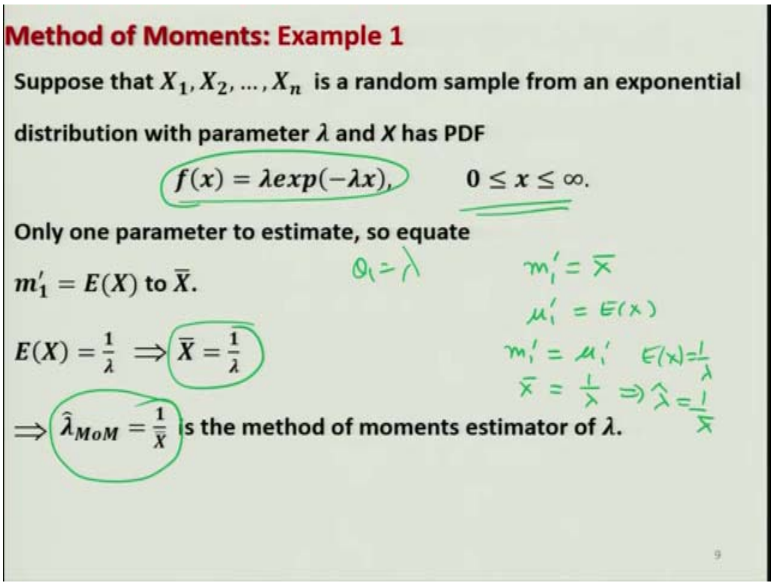
(Refer Slide Time: 11:19)



And they are called as Method of Moments estimator. So, formally I can say that they let X_1, $X_2, \ldots, X_n$ be a random sample from probability density function or former probability mass function whatever it is, and suppose there are p unknown parameters $\theta_1, \theta_2, \ldots, \theta_p$. So, in order to find out the method of moments estimators what we try to do, we create the first p population moment to the first p sample moments and then we try to solve those equations.

And the solving of those equations results into the values of those parameters, which are the estimators, and those estimators are called as Method of Moments Estimator or say MOM Method of Moments Estimators.

(Refer Slide Time: 12:08)



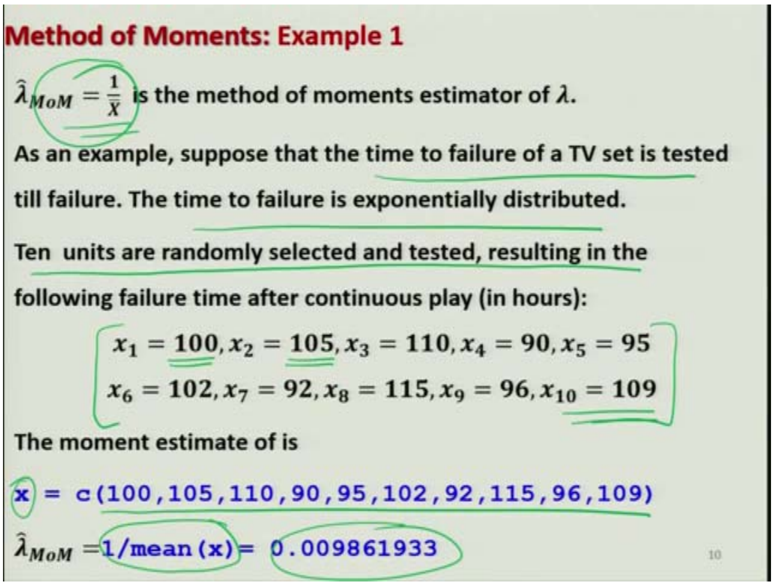
So, let me try to take here some example to show you that how things work. So, let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential distribution with parameter λ, and whose PDF is given by this $f(x) = \lambda exp(-\lambda x), \quad 0 \leq x \leq \infty$.

Now, you can see here there is only one parameter here. What is that? lambda. So, what you try to do now, if you try to look into this definition $\theta_1, \theta_2, \ldots, \theta_p$. Now, here in this case p equal to here 1 that means, there is only one parameter θ_1 is equal to λ. So, this means, we simply have to compute the first raw moment based on sample and population.

So, the first raw moment based on sample is indicated by m_1' and this is equal to here $\bar{X}$ and the first raw moment based on the population is here μ_1' which is actually here expected value of X. So, now, I simply have to substitute μ_1' is equal to m_1' like this. And once you try to do it here you can see here that sample mean capital $\bar{X}$ becomes equal to say here $1/\lambda$ because we know that in the case of exponential distribution the expected value of X is given by $1/\lambda$.

So, now, you can see here this is what I have written here. Now, once you try to solve it here, this will give here I say λ is equal to $1/\bar{X}$ and hence λ head becomes equal to $1/\bar{X}$ and this is here the method of moments estimator of lambda.

(Refer Slide Time: 14:08)



Now, in case if I ask you that can you really find it on the basis of R? Do you think that finding out the value of $1/\bar{X}$ is difficult? So, now, let us try to take an example and try to show you that what do we really mean by this estimator and how do we interpret it, how do we calculate it on the basis of R software.

So, now, we have understood that in this case, $\hat{\lambda} = 1/\bar{X}$ is the method of moments estimator for lambda. So, now, suppose we have some data set on the time to failure of our TV set. So, certain number of TV sets are tested and it has been recorded that after what time they are going to be failed, and that time to failure is exponentially distributed, because this is a lifetime.

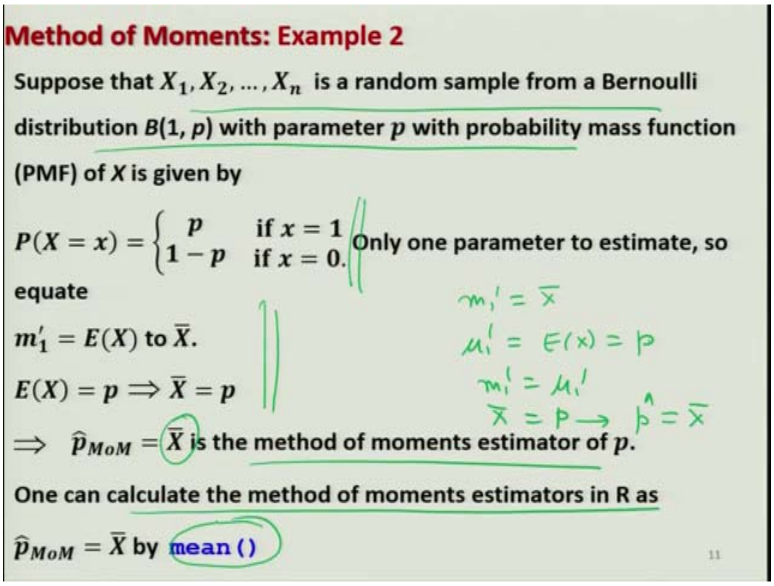
Now, you see, in reality, if you are trying to deal with such situation now, you know the basic background that if you are getting a data on the lifetime of a product then possibly exponential distribution can be used here. So, this test is conducted on 10 units which are randomly selected and tested and these following values are obtained on the number of playing hours.

So, the first TV set when 400 hours, second TV set went for 105 hours and so, on and the tenth TV set lasted 109 hours. Now, you want to know what is the average life after which the TV set may fail to work. So, that means, now, you have to take a call what you really want to know? Mean or variance or skewness or kurtosis this question can be answered by considering the mean of the population. And now, you are trying to employ the method of moments to find out the mean of the population and you have obtained that the method of moments estimate is $1/\bar{X}$.

So, now, you simply have to compute the value of $1/\bar{X}$ on the basis of this given set of data. So, now, if you ask me that how to compute this value in the R software or any software, I will say simply try to enter the data like here this and simply try to use the command 1/mean of X where X is the data vector containing all such observations and you can see here this will come out to be here like this. So, you can see here now, estimating the parameter is not difficult at all.

And once you have obtained the value of the λ you can compute different types of probabilities. Now, the complete distribution here is known to you. If you try to see here now, this entire PDF is known to you because λ is known to you and that is why λ was called as a parameter. Once you know the probability density function, you can draw any type of information what you want.

(Refer Slide Time: 17:19)

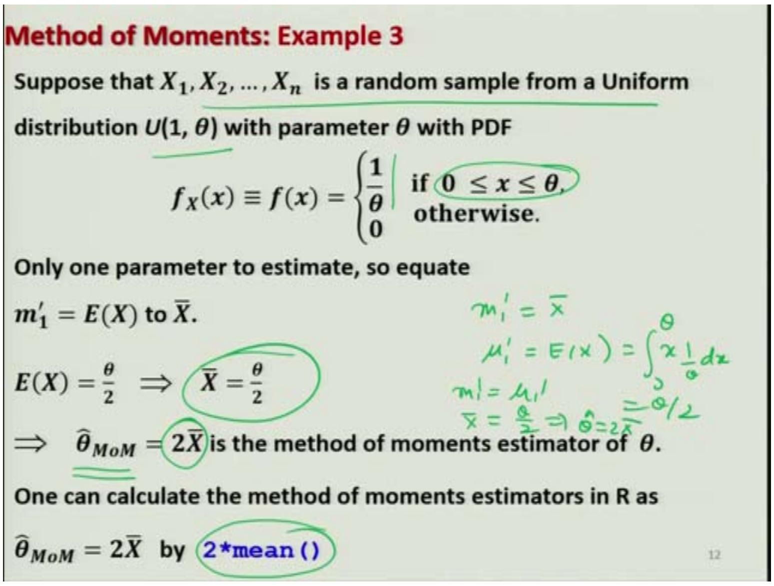


So, that was the advantage. Now, let me try to take one more example very simple example, suppose, let $X_1, X_2, \dots, X_n$ is a random sample from a Bernoulli distribution with parameter P whose probability mass function is given here like this. That we know, and in this case also there is only one parameter P that we want to estimate. So, once again I have to do the same thing, the sample-based first raw moment is here m_1' which is equal to here $\bar{X}$, and the μ_1' which is the first raw population moment this is expected value of a here X now, expected value of X that we already have found is equal to p.

So, you try to equate both here say $m_1' = \mu_1'$ and we get here $\bar{X}$ equal to here P. This is what I have done here. And now, once you are getting this thing, then I can write down here that $\hat{p} = \bar{X}$ that means, the method of moments estimator for the population proportion in the case of Bernoulli distribution can be estimated by the sample mean, and this is the method of moments estimator of p. So, now, if you ask me how to compute the mean, do you think that could I inform you?

Yes, I should, because one can calculate the method of moments estimator in R as the command mean that is all So, you can see here the life is now become so, simple, so, easy once you know all these things.

(Refer Slide Time: 18:51)

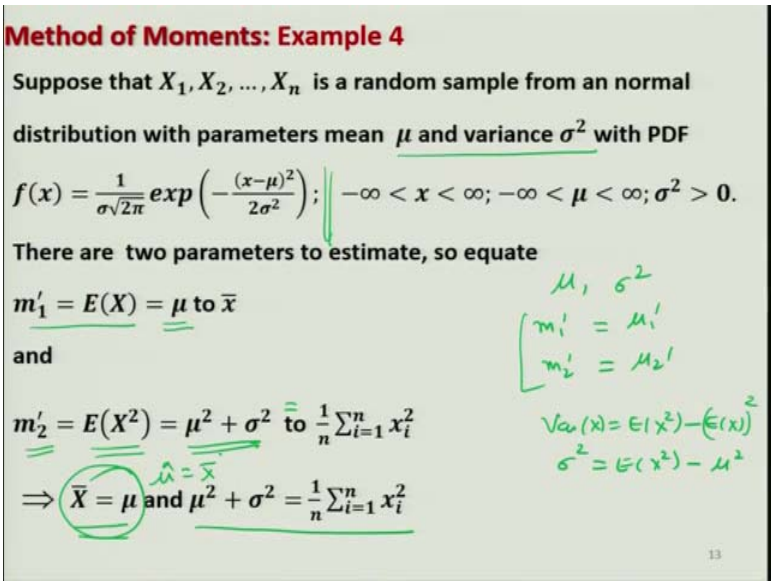


Now, I try to give you one more example. Suppose, X_1, X_2,…, X_n they are a random sample from a uniform distribution, $U(1, \theta)$where this $f(x) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x \le \theta, \\ 0 & \text{otherwise.} \end{cases}$. So, in this case, we have to do the same thing the first raw moment m_1' is obtained here as say $\bar{X}$, the first raw population-based moment that is given by expected value of here X.

So, and then I have to simply equate it. So and in this case if you try to find out the value of expected value of X, this is going to be simply 0 to θ X 1/θ into dx and if you simply try to obtain it, this will come out to be θ by 2. So, that is what I try to equate here that $m_1' = \mu_1'$ and then this will become here $\bar{X} = \theta/2$ which I have written here.

And now, once you do this thing, because we'll imply that $\hat{\theta} = 2\bar{X}$. And so, the $\hat{\theta}$ is method of moments estimator, which is equal to $2\bar{X}$ for the parameter θ. So now, if you want to compute this estimator, you have to simply write down here 2 star mean in the R software, this is pretty simple.

(Refer Slide Time: 20:17)

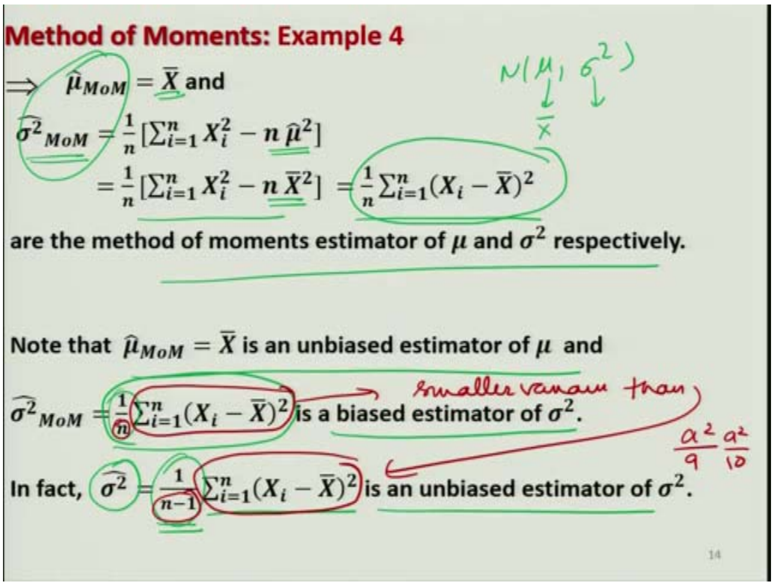


Now, similarly, if I try to take here, the last example that $X_1, X_2, \ldots, X_n$, they are coming from a normal distribution with parameter μ and σ^2 and their PDF is given here like this. Now, in this case, you can see there are two parameters. One is here is mean, and another here is variance. So what we have to do here? We have to find out the first two sample-based raw moments and first two population-based raw moments. So we need to find out here m_1' and m_2', and then we need to find out here, μ_1' and μ_2', and then we have to equate them.

Now here, we have here two equations and two unknowns, and we simply solve them and get the value of μ and σ^2. So, that is what I am going to do here that m_1' = E(X) that is equal to μ and m_2' = $E(X^2)$ that you can obtain very easily that variance of X is equal to $E(X^2)$ $-(E(X))^2$.

So, variance here is σ^2 and minus μ^2. So, you can very easily compute that $E(X^2) = \mu^2 + \sigma^2$. So, we try to equate $\mu^2 + \sigma^2$ 2 $\frac{1}{n}\sum_{i=1}^{n} x_i^2$. So, we try to now equate here $\bar{X}$ equal to μ and $\mu^2 + \sigma^2$ is equal to $\frac{1}{n}\sum_{i=1}^{n} x_i^2$. So, now, if you try to solve this thing, this is giving you directly here that $\hat{\mu}$ is equal to $\bar{X}$.

(Refer Slide Time: 21:13)



And if you try to substitute these values over here, you can obtain the estimate of σ^2 as here like this. So, here the method of moments estimator for μ is the sample mean. And now for the method of moments estimator for the variance, you can write down here the same thing $\frac{1}{n}\sum_{i=1}^{n} x_i^2$ minus say, instead of here μ, you try to replace μ by $\hat{\mu}$ so this will become here minus n $\hat{\mu}$ square so this becomes your n times $\bar{X}^2$.

And if you try to solve this quantity, this is simply here $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. So now you can see here, now you have got here the population $N(\mu, \sigma^2)$, you do not know the value of μ, you do not know the value of σ^2, I am simply now giving you this very simple solution, just use the method of moments and that is trying to indicate you or that is trying to inform you that the value of μ can be obtained just by finding or the arithmetic mean of the sample observation and the value of σ^2 can be obtained by the sample-based variance, which is given here like this.

So, this $\hat{\mu}$ and $\hat{\sigma}^2$ they are the method of moments estimator for μ and σ^2 respectively. And one thing I can just show you and you can verify it here, that this $\bar{X}$ is an unbiased estimator of μ whereas this $\hat{\sigma}^2$, which is obtained here, as a method of moments estimator, this is a biased estimator of σ^2.

In fact, if you try to make a small modification, then this $\hat{\sigma}^2$ can be converted into an unbiased estimator, what is that a modification just try to take in the denominator n minus 1.

So, if you try to take another estimator of σ^2 as $\hat{\sigma}^2$, which is $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, then this will become an unbiased estimator of σ^2.

Well, this is not a method of moments estimator, but still I thought I should inform you because many times we have used these two forms, and always I was telling you that, that these two estimator will have different properties. So, you can see here that in these two form with the divisor 1/n and the with the divisor 1/(n-1), the estimator with the divisor (n – 1) that is going to be an unbiased estimator whereas, the estimator with the divisor n this is a biased estimator. And now, you understand what is the meaning of biased and unbiased estimator.

Also in case if you try to think about the variability here, the quantity in the numerator, this is the same. $\sum_{i=1}^{n}(X_i - \overline{X})^2$. The difference is coming only in the denominator that in the method of moments estimator this is divided by n and in the case of $\hat{\sigma}^2$, this is divided by (n – 1). So, obviously, this estimator will have a smaller variance than this. Why? In case if you try to take some quantity here a square and try to divide it by here 9 or say a square divided by 10 what do you think which is a smaller quantity? That is a very simple thing to understand.

So, now, we come to an end to this lecture. And now, you can see here possibly, this was one of the most interesting lecture of the entire course. Now here you are able to implement all the things what you have learned, but my first question to you all is that. I am sure that up to now you must all be getting extremely bored because I was trying to initiate different types of concept. Sometimes there is some probability, sometimes there is probability function, sometimes there is a continuous random variable, sometimes it is a discrete random variable, etcetera, etcetera.

But now, you can see, after this lecture, I am asking you a question and I will leave the answer up to you. If you had not understood those basic fundamentals, can you really understand anything in this lecture? This is the lecture which you wanted from the beginning that how are you going to estimate the unknown parameters?

Now, in case if I tell you, you have got an exponential population, you try to estimate the parameter λ by 1/mean, and if you have a normal population try to estimate the population mean by simply by sample mean? How do you know which one is correct? And I would like that you take this call, not me. But if you do not know these fundamentals, how will you know that whether the mean has to be estimated by $\overline{X}$ or $1/\overline{X}$.

So, that is precisely what I told you in the beginning of the lecture when I posed the question before you what you really want to be? A compounder or a doctor. I am sure, that now you are going to becoming the doctor very soon. So, you try to have a look into the books try to take some examples try to practice it. And I will see you in the next lecture with method of maximum likelihood. Till then, goodbye.