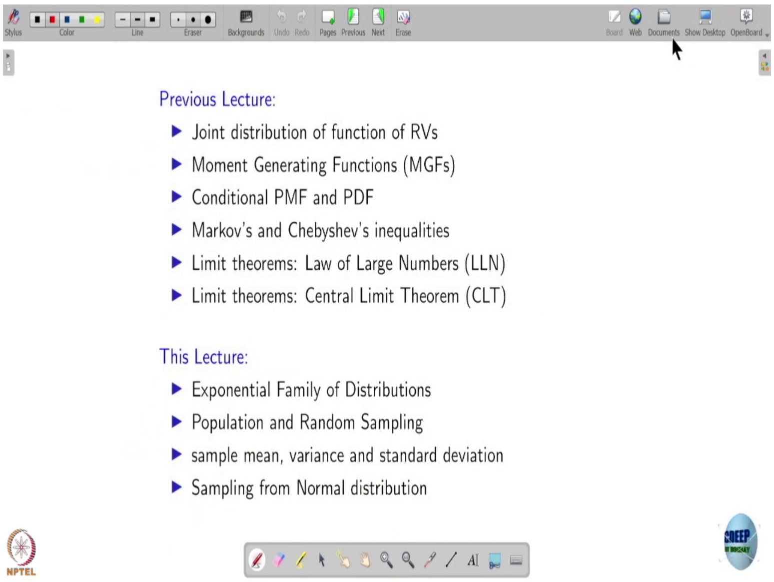
Engineering Statistics

Professor Manjesh Hanawal
Department of Industrial Engineering and Operation
Indian Institute of Technology, Bombay
Week 4
Lecture 19
Gamma and Chi-square distributions

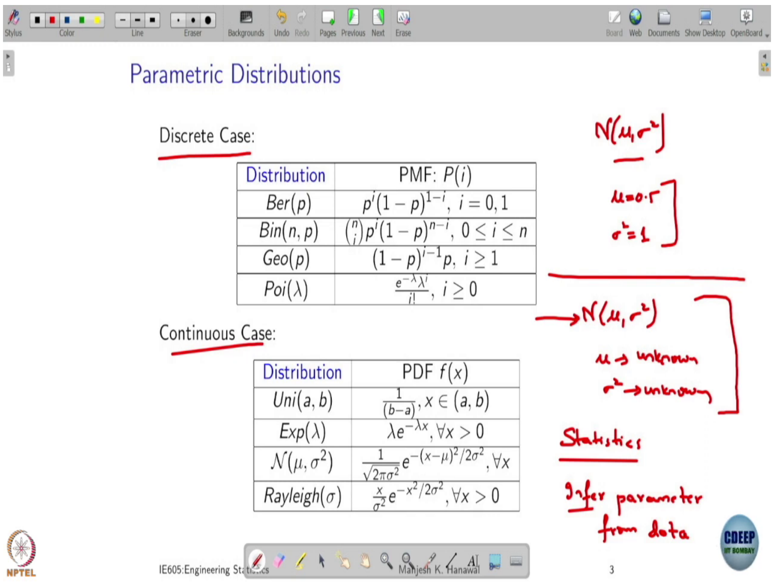
Any questions so far what I have discussed?

(Refer Slide Time: 00:20)



So, we have been discussing, I think in the last class we talked about limit theorems, law of large numbers and central limit theorem.

(Refer Slide Time: 00:30)



Now, as we move from probability to statistics, now, what we will be doing is, we will be trying to look mainly at the data and trying to see that, from which underlying distribution are, what are the parameter associated with the distribution that is likely generating this data. Whereas the probability was about, we already took some distribution, and try to understand its properties.

And in probability like we already start with some distributions, all the parameters, everything is known. And according to that we talk about data being generated. Now, in the statistics, we do not start with distribution directly, we may start with some model of the distribution, but we do not know its complete characterization. We do not know what are its parameter.

But we will start, we will have access to the data. And from the data, we will try to see that what is the underlying distribution or what are the parameters of that underlying distributions. For that, let us have a quick recall on all the distributions we talked so far. We will expand this in this class today. But as a recap, in the discrete case, we talked about Bernoulli, Binomial, Geometric and Poisson. And notice that all of them are coming with certain parameters.

And also, in the continuous case, we talked about Uniform, Exponential, Gaussian, Rayleigh. And I think we made some remark about other distributions like a Laplace distribution and couple of other distributions. Now, if I tell that, data is normal with mean and sigma squared, I may say that data is coming according to the law, which is Gaussian with parameters sigma and sigma, μ and σ^2 .

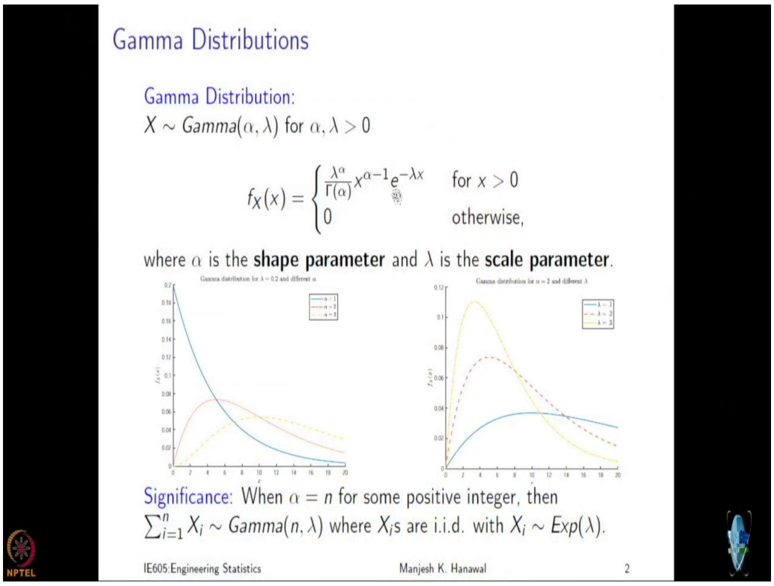
I will just say that, data is coming from some Gaussian distribution with some parameter mu sigma square. If I specify what is the value of μ , let us say if I tell μ is 0.5 or some value like σ^2 is 1, then I have completely specified the distribution and here you do not, somebody do not need to give you data you can yourself generate the data as using this parameter.

So, this is like if I tell this parameter this is completely specifying your probability. This is like a of probability. On the other hand, I may say that, data is coming from some distribution which I am going to assuming to be μ^2 , but μ is unknown and maybe sometimes even σ^2 is also unknown.

Here, what I am just assuming data is going to be, I am instead of assuming some arbitrary distribution I am going to restricting to Gaussian distribution but here I just do not know what is that parameter μ and σ^2 are. Now, in statistics we will try to infer this parameter μ and σ^2 from data.

So, some time infer, we call sometimes we called estimate and all like we will make that more formal. So, to infer these parameters we need data. And that is why, even though we start with some underlying probability model when I talk about statistics I will always say that we will start with data and from that data we have to find out that parameter of the distributions.

(Refer Slide Time: 05:10)



Now, as I said in, we will want to little bit expand our family of distributions we know and see what are the parameters they are associated with. The first distributions we are going to look into is gamma distribution. Here, I cannot write, so I cannot write in this I will just use the mouse. Gamma distribution is parameterized by 2 parameters alpha and lambda, both alpha and lambda are positive.

And last time we discussed the definition of gamma function. So, we all know the classic gamma function, but that gamma function can be further parameterized with some parameter like if you just if you ignore this denominator and let lambda equals to 1, so that is x to the power alpha minus 1 e to the power minus x. When you integrate it between 0 to a certain number that will define the gamma function at that number.

But now, you can bring in this another parameter lambda. So, then if you add alpha to the power lambda x alpha minus 1 and e to the power minus lambda x that is like a generalization of your gamma function. And now, if you see that if you normalize this integral by this actual integrand you will see that this actually when you integrate it over the whole range from 0 to infinity, this normalizes to 1, this integrates to 1, that is what we discussed last time.

So, this is a valid probability density function. So, this is positive valued and the area under this curve is 1. Now, as you see, I have plotted it in 2 scenarios. In the first scenario, I have hold this lambda to be constant to 0.2 and then I am wearing alpha. As you see that as I go from alpha equals to 1 to 2 to 3, when alpha equals to 1 this is like a monotonically decreasing, when I made alpha equals to 2 it started first increasing and then decreasing.

And as I am increasing alpha it looks like it is started only showing increasing behavior. So, that is why when you have, when you, if you look into different value of alpha it is shape is changing, from monotonically decreasing to all the way monotonically increasing. So, the shape is changing, that is what we call this alpha shape parameter. And here I have plotted for a fixed alpha different value of lambda.

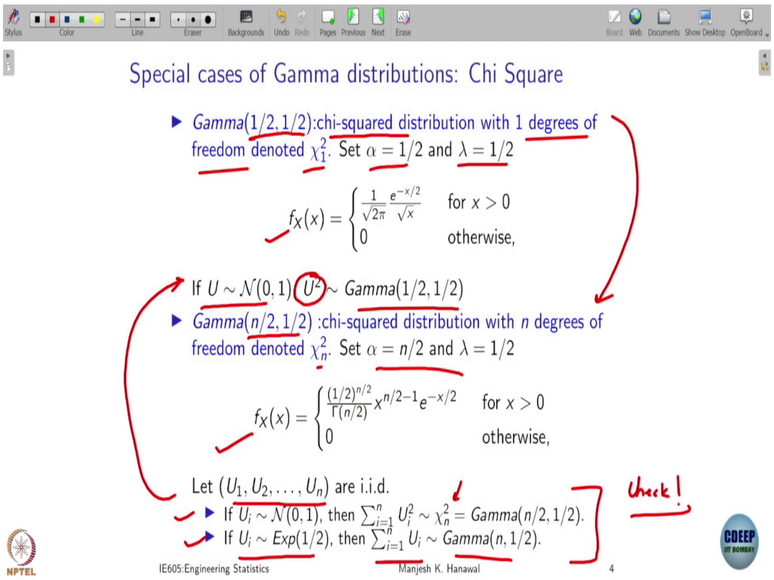
When it is lambda equals to 0.1 it was like monotonically increasing, but it also showing some small tendency of decreasing. When lambda equals to 0.2 this is also increasing and decreasing. Lambda equals to 0.3 increasing and decreasing. So, as we are increasing lambda it is taking larger and larger values in some particular peak regions.

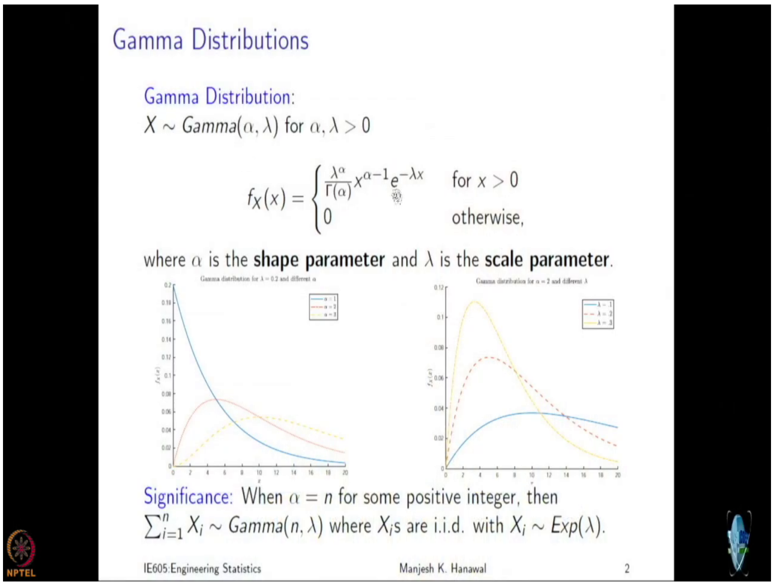
So, basically what is happening lambda is in kind of scaling and that is also kind of evident from this, maybe not so evident, but anyway as you see like as we are increasing lambda it is kind of scaling like y values are getting more scaled. Not across entire range, but in some particular regions where the peak is happening. Because of this sometimes this is also called scale parameter.

Now, even though we have discussed this gamma function gamma distribution in terms of gamma function, it has relation to some previous distribution that we know. What is that? Suppose, you take some integer n and take n number of Gaussian distributions that are, sorry n number of exponential distributions coming from parameter lambda with lambda and they are all i.i.d.

If you add them the resulting distribution is exactly going to follow this PDF, that is it is going to be gamma with parameter n and lambda. So, this gamma distribution can be interpreted as gamma distribution when n is integer it can be interpreted as sum of n i.i.d. exponential random variables with parameter lambda.

(Refer Slide Time: 10:39)





So, basically the gamma distributions we can now take any value of n and lambda because these are n and lambda parameters. It is not necessary we have said, we have never said that so far in the gamma distribution n has to be integer. Only as a special case we showed that when n is integer then gamma is related to the sum of i.i.d. exponential random variables. So, now, let us look into what happens if I restrict myself to some specific values of alpha and lambda.

So, here I have taken alpha equals to half and lambda equals to half. So, here the first parameter is not integer I can interpret it as a sum of half exponential random variable with parameter half I can interpret. But this distribution gamma half and half this has a special name and special significance, this is called Chi-squared distribution with 1 degrees of freedom and it is denoted as this symbol Chi with subscript 1, 1 is denoted here 1 degrees of freedom and Chi squared the 2 is, the superscript 2 is denoting Chi squared.

In the gamma distribution if you are going to set alpha equals to 1 and lambda equals to half this is the distribution you will end up, the simplified one for the specific value of alpha and. Now, you may be wondering, fine you have taken alpha equals to half then where is this 1 degrees of freedom, why that term is coming. And just to see that why could that why this has been given the name of 1 degrees of freedom.

So, suppose let us take uniform distribution, sorry, let us say U is some random variable which is normally distributed. Now, you can show that if you take the square of that that has this gamma distribution that is square of a normal distribution is chi square distribution with 1 degrees of freedom.

You see that now U is I have just taken U to be 1 random variable it can take any value as per that probability density, and if I just square it, I am getting gamma distribution and here U is one component which can take any value it likes as per that. And if I am just squaring that I am getting this gamma distribution. Next, there is another, if you now, this instead of gamma half and half, if you take gamma n by 2 and half.

Now, this is also given a special name, this is just like extension of this chi-squared distribution with n degrees of freedom and here n is like an integer and that is denoted as chi square with a subscript of n. So, then, what is chi square distribution with n degrees of freedom? This is nothing but a gamma distribution with parameter n by 2 and half. And if I plug in the value n by 2 and half in this PDF we will, you will get back this value.

Now again, this chi-squared distribution with n degrees of freedom. This is again related to other distributions. So, let us say if you have n random variables, n i.i.d. random variables, I am now considering 2 cases, case 1, where all of the random variables have normal distribution. If you square them and add that will have chi square distribution with n degrees of freedom.

So, when n equals to 1, when this n equals to 1, we had already this case that we have gotten. So, it is not just that only Gaussian is related to this gamma distributions. You can even verify that if all these Ui's are exponentially distributed with parameter half and if you are going to add them and all of them are i.i.d. then also you will get gamma distribution with parameter n comma half. Notice that here it is not n by 2 it is n here.

And now even though I have stated them here you should check this, check this, this is indeed correct. Make sure that if you are going to add all this, if you generate n i.i.d. random variables which are normal, and if you are going to add them, you will indeed get this chi-squared distributions. Now, let me see I think at some point, we should actually start simulating this and say.