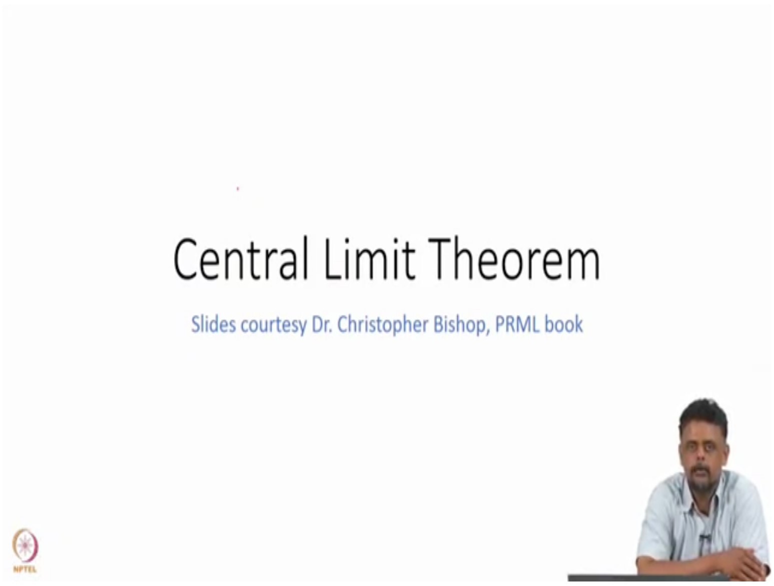
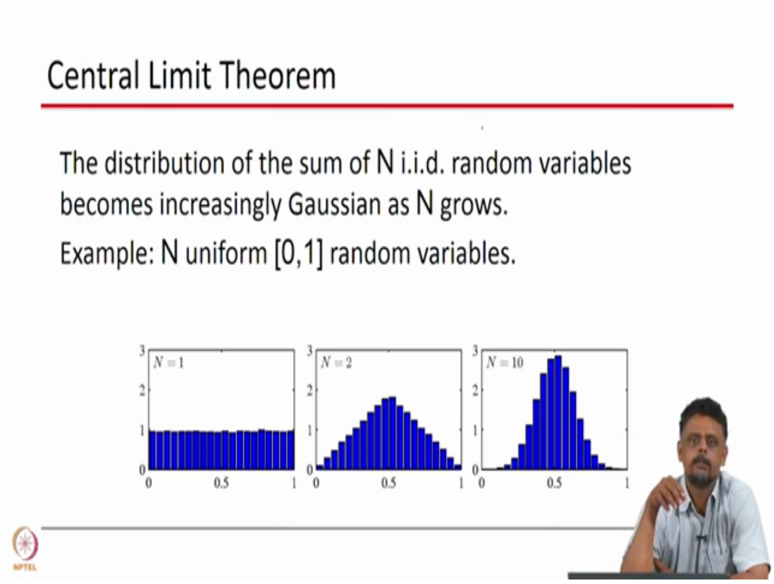
Machine Learning for Engineering and Science Applications
Professor Dr. Ganapathy Krishnamurthi
Department of Engineering Design
Indian Institute of Technology, Madras
Central Limit Theorem

(Refer Slide Time: 00:13)



Hello and welcome back, so in this video we will look at the central limit theorem there is only one slide and courtesy by Dr. Christopher Bishop based on his PRML textbook.

(Refer Slide Time: 00:24)



So what is central theorem limit theorem state, now that we have looked at we have familiar with Gaussian distribution and the (berno) Bernoulli distribution, what it says is that the distribution of the sum of N independent and identically distributed random numbers random variables becomes increasingly Gaussian ok, so as N becomes very large ok so if you look at these plots so the first plot (cones) corresponds to N equal to 1 remember that we are considering the sum of N independent identically distributed random variables.

So in this case N equal to 1 the first there is no summation here, so we are just drawn from a numbers from a random distribution from a uniform random distribution between 0 and 1 ok now if we consider sum of two numbers each of which is drawn from is drawn from a uniform random distribution and we plot the histogram of those drawers then we see that again it is approaching a Gaussian as we go to N equal to 10 it looks more and more like a Gaussian, ok.

So again to reiterate we are only considering sum of N numbers ok the summation of N I will say numbers but summation of N random variables which are drawn from the same distribution, so that is the condition yeah independently drawn from the same risk of independent and identically distributed ok, so why is this important? Because if you know if you if you look in many of our problems if you look machine learning problems they are used probabilistic models typically we will end up using the Gaussian distribution as the model, ok.

So then how do we justify that so typically the justification comes from here where if you look if you consider your data point for instance let us say you have a bunch of data points and you say ok we decide that it they are Gaussian distributed so then what is the justification so one justification you give depending on the problem is that each of these data points can be considered as a sum of N numbers drawn from a similar distribution and it is a very large number then we can say that oh then each of them we can say that this now then (curres) is drawn from a Gaussian distribution, so that is the idea behind our using the central limit theorem.

So one example is you are all familiar with you know cell phone cameras where you take pictures right ok, so but then what is that cell phone, cell phone camera do it is just that it is collecting the light photons that are reflected off the object that you are photographing ok and it is integrating the combo it is integrating them in a sense it is counting the number of photons literally ok that is what you are the camera did the camera detector in your cell phone that is what it does.

So even though if you if you look at counting statistics that counting statistics are not they are usually called Poisson distribution ok, we have not done Poisson but (())(03:09) Poisson distribution however when you consider a large number of light photons and you are into indoor and your detector is actually you know integrating the counts ok each time a photon falls are there these the intensity increases you can think of it that way.

So then when you get the output picture each of the pixels corresponds to a detector in the camera a detector element in the camera each of the pixels in an image but then that detector is the these even though the individual statistics are poison because we are integrating over a large number of those photons we can see that each pixel can be modelled the intensity can be modelled as a Gaussian distribution ok, this is the idea behind you behind importance behind central limit theorem, is that the limit of as N being very large and we are considering a sum of i.i.d random variables gone from the same distribution however i.i.d means that then we can say that the result can be interpreted as being from a Gaussian distribution ok, thank you.