Probability Methods in Civil Engineering

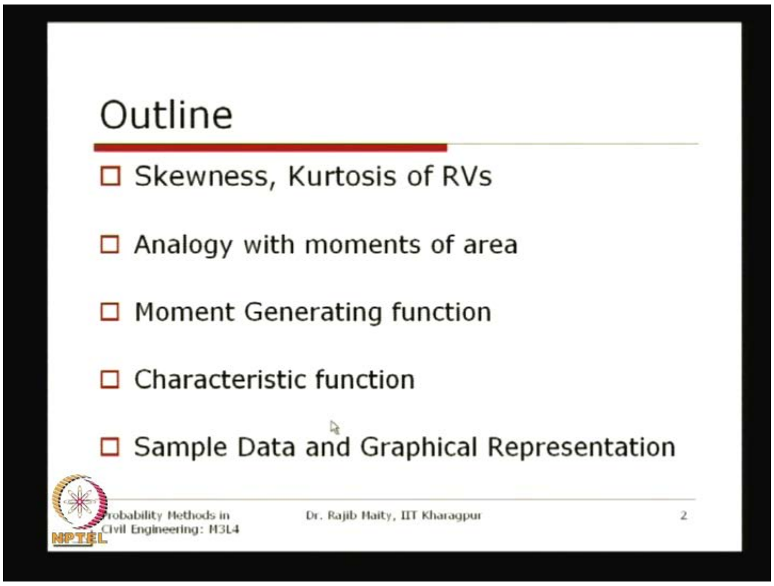
Prof. RajibMaity
Department of Civil Engineering
Indian Institute of Technology, Kharagpur

Lecture No. # 09
Further Descriptors of Random Variables

Hello and welcome to the forth lecture of our, this current module,module 3 this module is on random variable. And in this lecture,we will cover thatfurther descriptors of random variable.Basically, in the last classwhat we did?We started withthe description ofcdfand after that, we started some of the descriptors of the random variable, if the;in absence of thea proper definition of thispdfandcdf,if we know some sample data then from the sample data there can be, we can define some descriptorsof the random variable, and with that we get some idea about how theparticular random variablepdfdistribution the behavior of thepdfwe know.

So, in the last class, we saw that it is central tendencyand that central tendency as we saw that it is in terms ofit is mean mode or median.And then we saw thatcoefficient of dispersion that is measure ofdispersionin terms of variance and standard deviation.So, today's lecture that we will start with some more further descriptors of this random variable.

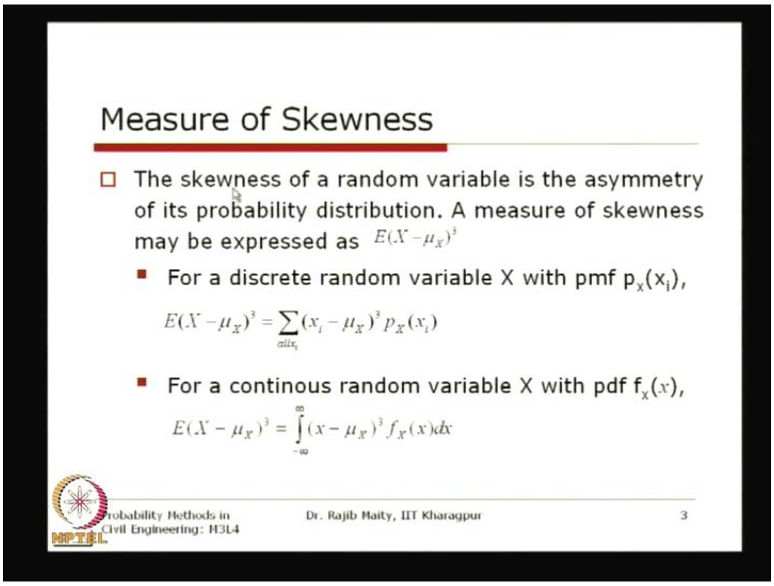
(Refer Slide Time: 01:40)



And we will start with skewness and kurtosis. This skewness and kurtosis are the highermomentsas we have just described being the last that analogy,that we have given that analogy with these moments.This is the higher moments with the respect to the mean and standarddeviation sowe will see that one, firstthat skewness and kurtosis.Then we will see some analogy with the moment of area that we indicated in this last class. And with this, we will start that moment generating function and characteristic function, these two are very useful function and if we can define these functions,then we will see that the description ofall such moments that is basically, this mean, standard deviation or mean,variance, Skewness,Kurtosis this is up to fourth. And above also that highermoments are also possible.

So, we will see that how,if we know the moment generating function and characteristic function, and then how all this moments in any ordermoment of any other, how it can be defined with a help of this twofive functions. Then we will start, because these are for thisfor somedistribution for if some sample data is up level, then how we can that that particular sample data, we can represent graphically,sothat we can have some idea about thisdescriptors with respect to it isgraphical representation.

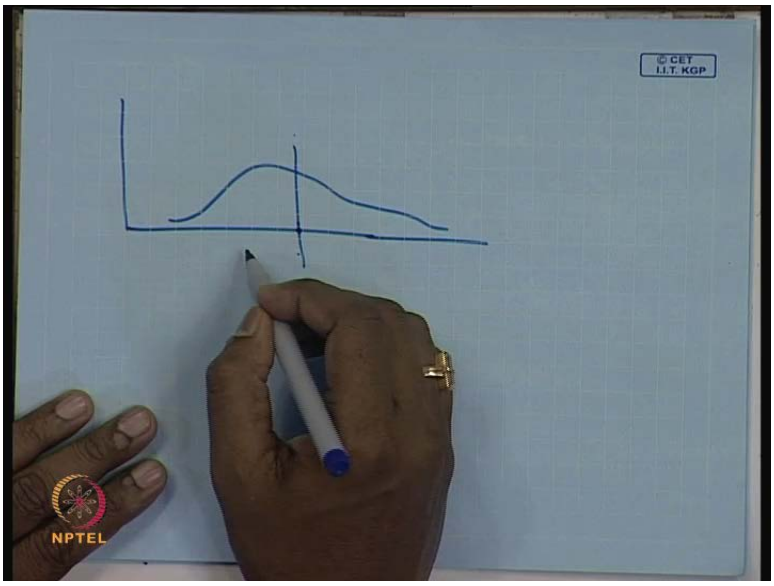
(Refer Slide Time: 03:09)



So, this is we will try to covertoday’s lecture and we will start with the skewness. As we told that this skewness of a random variable is the asymmetry of it is probability distribution. So, this is a measure of the measure of skewness may be express in terms of this expectation of X minusmu x. If you recall, that in the last class what wediscuss that?That the first moment when we are taking,when we are lookingfor thecentral tendency, we take the momentwith respect to theorigin. And so,it gives that the distance from the origin to that so it is nothing but that it is mean in the location of it is mean that also minute in terms of some graphical representation.

And now, all the higher ordermoments,if we take, takeit with respect to the mean if we take. And in the last class, we concluded the if the first moment with respect to the mean is becoming 0. So, this skewness that what we were are talking about, this is also a moment with respect to this mean and that meanit is the thirdmomentand second moment we discussed in this last class, last lecture that it was a variance. So, this is a measure of symmetry about this meanso, how this particulardistribution, if we see it here,then we will come to know that.

(Refer Slide Time: 04:38)

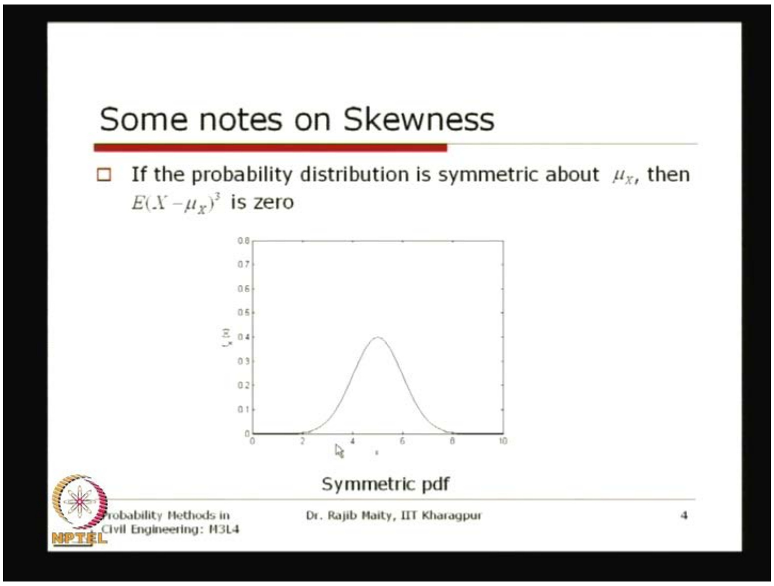


Now,from the origin I know how it is distributed and all. So, we know that where it is the location of this mean.Now, what we are trying to do from with respect to thissymmetry that with respect to this mean weatherit is symmetric or not. So, if it is not symmetric then how it isdistributed about the mean that is the goal. So, that is why when we are taking the third moment X minus this mu sothis sign remains.Sign remains means,it will show thatif it is more this side, then it will be negative, then if it is more this side it will be positive, means more means that more disperse.If it is the dispersion ismore this side than it will be,this will be negative and if it isthis sidethen it will bepositive.

So, what we will do? We will just take the expectation of this one, as usual for incase of this discrete random variable and incase of thus,in case of continuous random variable, for this discreet the expectation we know that.Now, this is becoming awe see, this is becoming a function of thus random variable. So, X is our random variable and this X minusmu x power cube is a function of our random variable X. So, we know how to take the moment or we know how to take thisfunctions as the expectation of a function, that we described in the last lecture with respect to the g x. So, this function will come, willbe multipliedby the thatprobability density were the; for the discreteit is the probability value for that particular outcome. And we will sum up all this things to get the,getwhat is the moment with respectwhat is the third moment with respect to the mean.

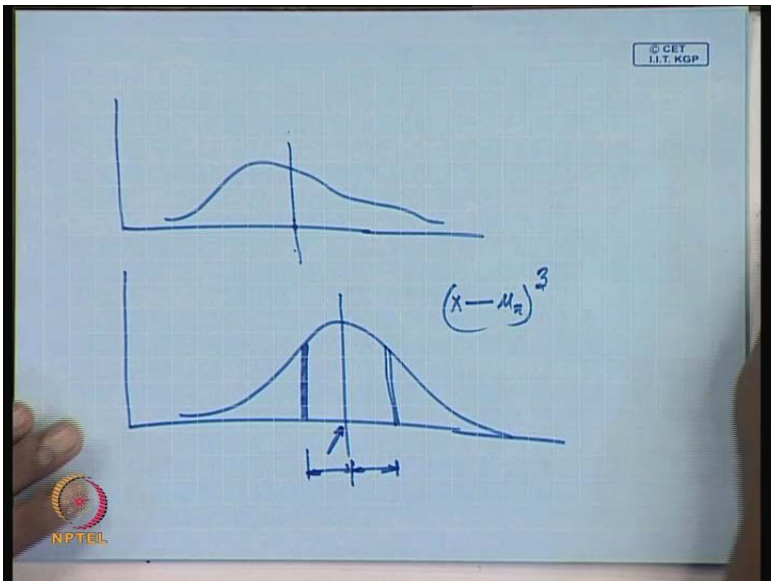
And similarly for the continuous random variable X forit ispdf, if it ispdfis defined by thisf X x.Then this expectation of this function, again we will be express in terms of asothis summation now isconverts to this integration.And this integration over this entire supportof this random variable X, we have taken it from minus infinity plus infinity. So, this gives that cube multiplied by the density and if we do this integration, we will get the measure of skewness. So, this is; this will get the third moment,third moment of that variable with respect to mean.

(Refer Slide Time: 07:05)



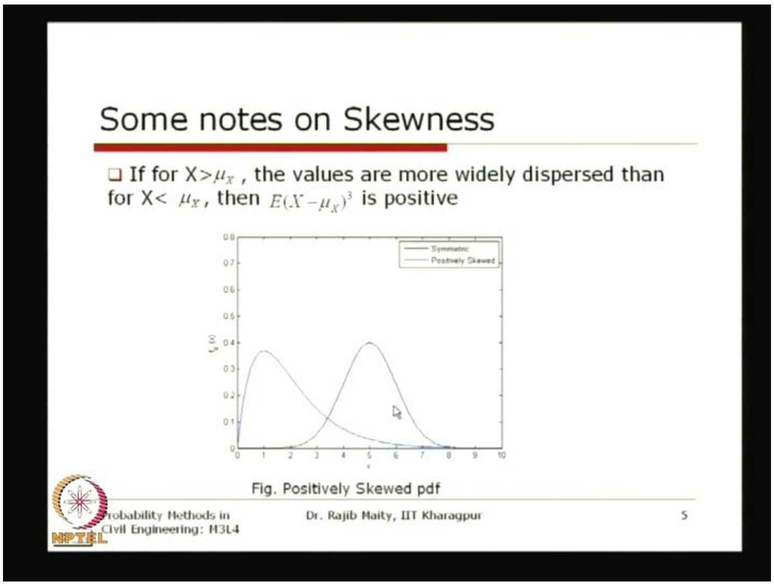
Now, if we; now, see thatif the probability distribution is symmetric about this mean then it is, thismoment will be equal to 0, this will be 0, because whatever now, if you see here.

(Refer Slide Time: 07:25)



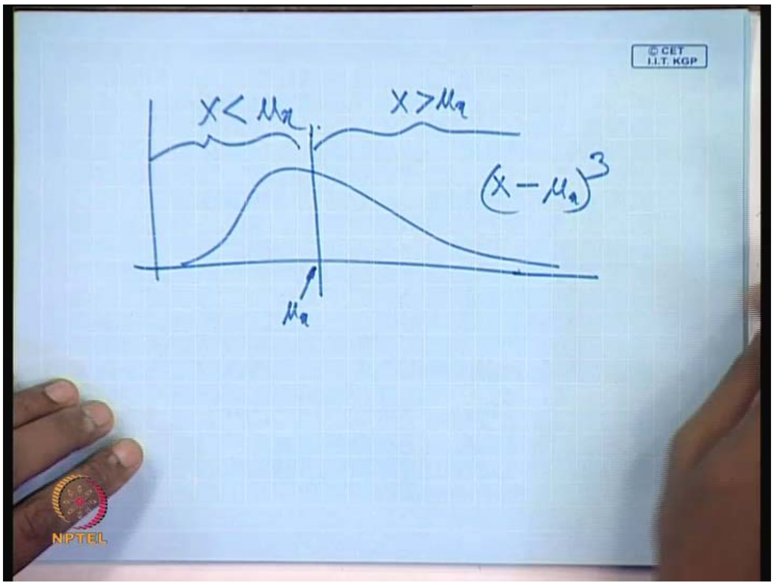
Now, if I draw one particular distribution which is symmetric now, this is your location of mean.Now, what we are doing actually?We are taking a small, a smallarea and as this is that distance as,what we are taking, X minus mu x and this power is cube. So that whether this is on the negative side withrespect to this mean or this is, sothis is on the positive side with respect to this mean.Now, being there, being it symmetric with respect to this one,being this power 3 thistwo are cancels out. So, if thisone is symmetric, then this moment will be will be exactly equals to 0.So, this is what is explained here, if theprobabilitydistribution is symmetric about mu x, then this expectation of X minus mu xpower cubethis is third moment is equals to 0, which is one example is also shown in this, in this diagram.

(Refer Slide Time: 08:23)



Now,we will compare with this symmetric, this black one again we have just written which is we know that this is a symmetric or which thisthusthird moment with respect to mean is 0.Now, if we see this blue one, this blue one is known as the positively skewed.Now,we will see how thisone is positively skewed, how it is distributed. If for X greater than mux, the values aremore widely disperse than for the X less than equals to mu x then this will obviously, become positive. So, if this is becoming this is becoming positive, then thisskewness,the skewness of this bluepdfis apositivelyskewed,we say it ispositivelyskewed. Now, this one is greater than for this zone.

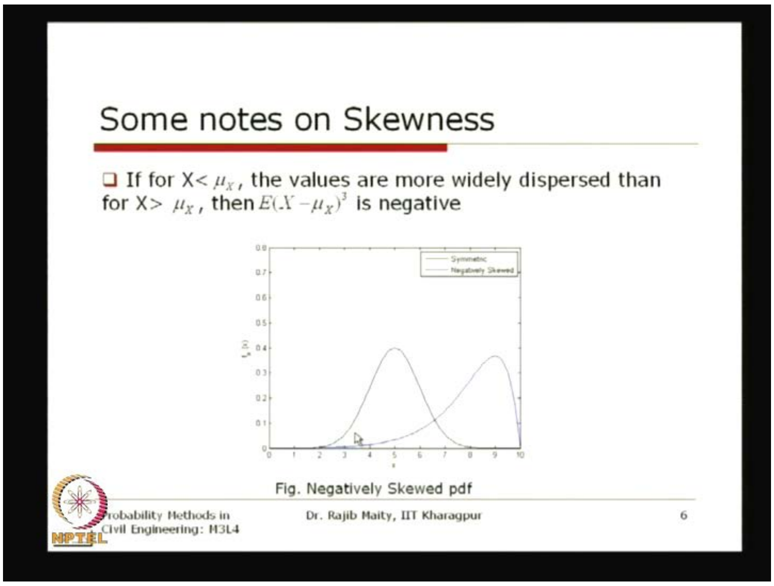
(Refer Slide Time: 09:34)



Now, if I just saw, if I just draw another one here,what is meant by this oneis.Now, as this side is disperse if we take this mean, may be the mean will be somewhere at this location, this mean will come here. Now, if we see that data range from the; from towards the negative side it was left side of this meanand the right side of this mean that is this is your mu x.So, thissideif I say, this is your X is greater than mu x now, and this side is your X less than mu x. Now,you can see that this side only widely spread, this is widely spread than compared to thisone, this is lesser, this is more dense compared to thus right handside. Then when we are taking that X minus mu x power cube this is givingyoumoreweightage that is why the totalsummation will be positive in casewhen it is skewedlike this.

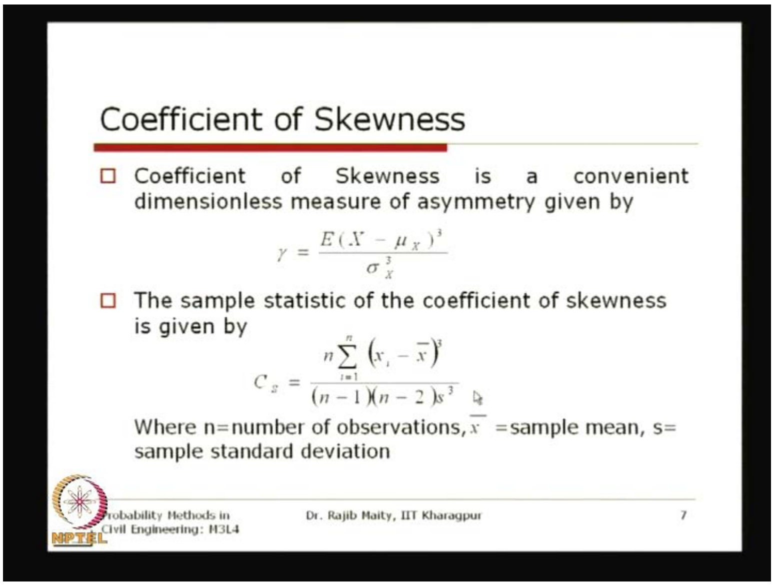
So,that example is shownhere. So, somewhere the main location will be in thishere where the right side, the positive side with respect to the mean will be more disperse, than compared to this aleft hand side. So, that resulting in thatsothis side it is most this quantitywill become positive. So, this bluepdfwhat is shown here, that is positively skewed.

(Refer Slide Time: 11:08)



Just opposite if we see,we will get that whatis called the negatively skewed. So, in this case it is just opposite that is X less than equals to mu x in this zone, in this region the values are more widely disperse, than for that X greater than mu x. So,somewhere in this location, in this point thisthat mean will come and this side it will be less disperse compare to this the side.So, obviously, due to thispower cube this quantity will become negative. So, what is shown here by this blue line thispdfis known as negatively skewed p d f.

(Refer Slide Time: 11:49)



Now, for up to this secondmoment what we have seen that,it can beviewed asa power of this random variable, when we are taking this moment with respect to the mean. And we are taking is as a square then; obviously, that particular quantity is also having the same unit that the random variable is having.Now, forever convenience, when we are increasing the order of this moment, when we are going for this third moment of fourth moment, then keeping the same unit may not bethat convenient. So, make it convenient what we do, is thatvalue, that moment is generally normalized with respect to it is some power of the standard deviation. So, the power of the standard deviation is selected in such a way, that the quantity becomesdimensionless. So, this is exactly is done here to get that and which is known, is known this coefficient of skewness.

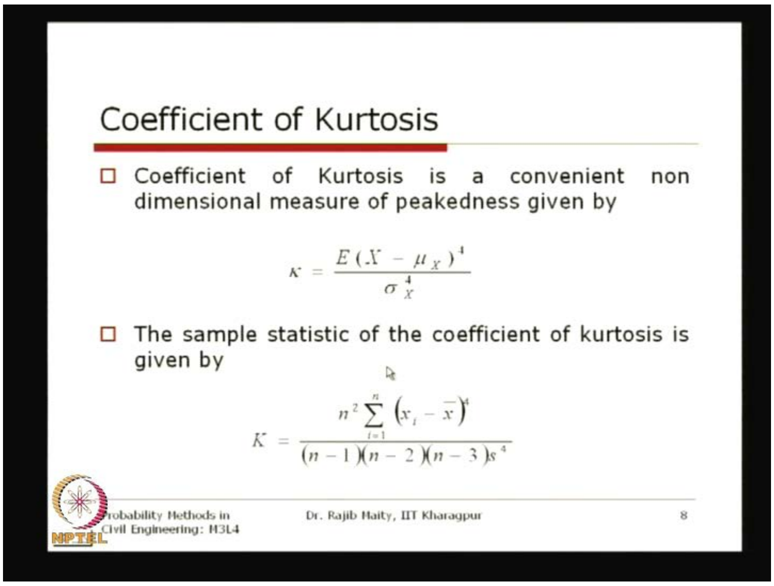
So, the coefficient of skewness here is the convenient dimensionlessmeasure of asymmetrygiven byso this is your expectation of this,this function which is the thirdorder moment.Now, this is obviously, a this quantity is having an unit of this cube of the unit of this random variable. So, the standard deviation also, we are taking the same power to get the coefficient of skewness. So, the coefficient of skewness as we are dividing it by this power 3, this is becoming the total quantity, this gamma becomingdimensionless.

So, this now,the; if this becomes 0, in case of the perfectly symmetricpdf, perfectly symmetric Probability Density Function, then this when it is becoming 0, this function is becoming, this value becoming 0. So, the gamma equals to 0 indicate the; this is asymmetricpdfand if it is positive then it is positivelyskewed, if it is negative then it is negatively skewed. So, thisdenominatoris only to make this quantity dimensionless.

Now, this is from thenow, if we have somesample data, if we want to calculate this one,calculate this measure that is this coefficient of skewness from the sample data. There are some samples statistic based on which we can calculate this coefficient of skewness. Now, this coefficient of Skewness,the sample estimate of thiscoefficient of skewness is expressedas this, where we take the individual observation and deducted from the mean. So, it is giving basically the distance from; distance of each and every observation from the mean. And we are taking x power cube summing it upfor the all observation, this n is the number of observation as shown it here, n equals to number of observation.

So, we are summing it up multiplying it by n, will come to this one and divided bythis is the sample estimate, this s is the sample estimate of this standard deviation that we discuss in the last lecture. So, this is the sample estimate of the standard deviation, we are taking it is power cube. So, this unit and this unit get cancelled andthese are the normalizing constantsso, this n divided by n minus 1 multiplied by n by 2 just make this one is consistent and unbiased. So, this is this total quantity what is given is that, this is you are the coefficient of Skewness,depending on this data available from x 1 to x n.

(Refer Slide Time: 15:39)

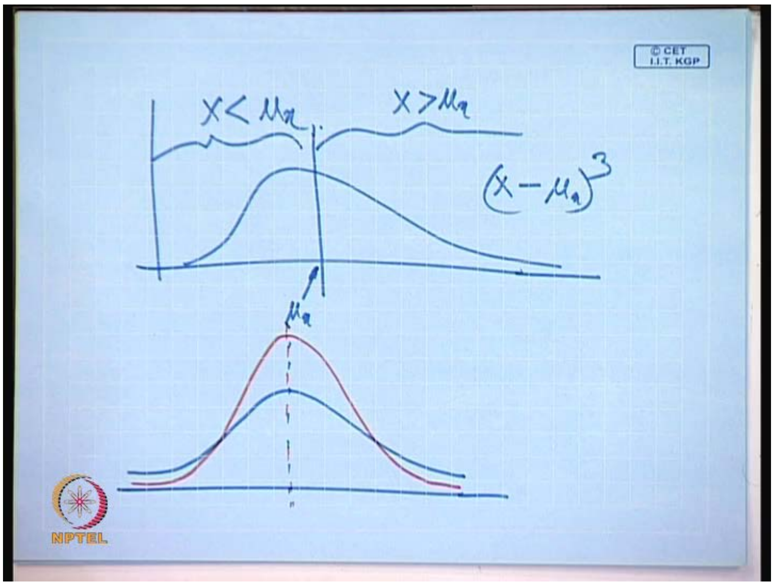


Now, we will go to this forth moment,forth moment with respect to the mean and which is known as thekurtosis.Now, sothe measure of kurtosis is the fourth moment with respect to mean. So, we are taking the expectation of the function X minus mu x power 4.Soand again to make this one dimensionless, what we are using is? We are using that this sigma x power 4.Now, as this again similar to this coefficient of skewness that is we are making, we are taking the power of thisstandard deviation 4 to make this total unit,total this one as the unitless.

Now, what is important here? What does this coefficient of kurtosis implies? Forexample, thuswe have seen that mean, what does it mean implies, because this is a central tendency, thatthen the coefficient of variation it is the this is indicating that it is dispersion with respect to mean. Then skewnees,skewnees is showing howweather it is most spread towards the right hand side, right hand side of mean or the left hand side of the mean sothat is the measure of symmetry.

Now, what is this kurtosis if we see, then we will see thatthis is also power 4soas it is power 4. And obviously, for the whateveron the left hand side and the right hand side that will be negated andthat will make 0.Now, these 4 is actually is a measure of peakedness.Now, this measure of peakednessmeans, how the peak, howdoes that particular peak of thispdflook like.

(Refer Slide Time: 17:29)



Now, if we see here, if we see that difference between say this is one,so this is also a symmetric distribution. And if I draw another one, this is also another symmetric distribution and they are mean are also saysame. So, the first moment is same, second moment is the variance also will be same if we see just the total thing.So, we will see that the variance of these two distributions with respect to the mean will be same.Now, thecoefficient of theskweness,skweness also will be same in the both the cases it is symmetric so that is why it is both the cases will be 0.Now,the difference of this peak of this, how it is peak at this point, that will be reflected with respect to it is fourth moment.And that is why this fourthmoment, when it is normalized by the fourth power of this standard deviationwhich is known this coefficient of kurtosis, this is known as this measure of peakedness.

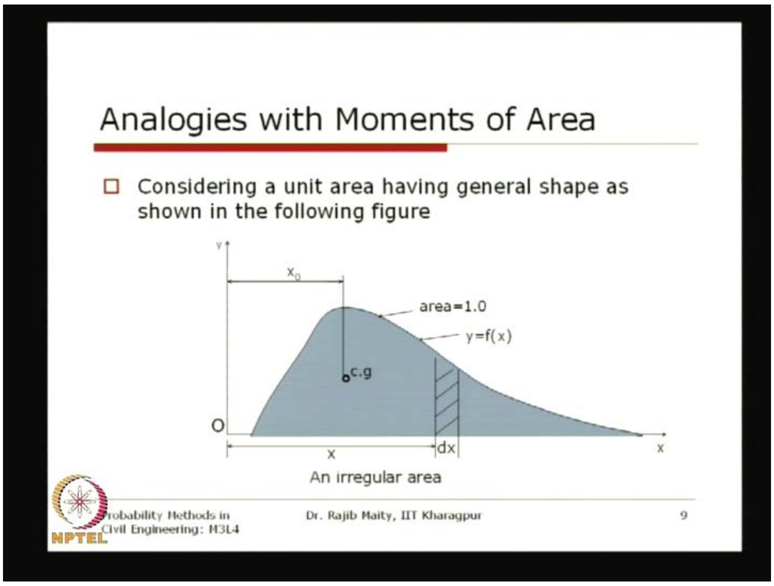
So, this measureof peakedness with this one having a sample statistics of thiskurtosis again the sample estimates.Sample estimate of this coefficient of kurtosis is expressed in terms of this that is each and every individual it is measure. The distance from this mean power 4 summing up for this all the observation and divided it by the 4 power of this standard deviationand this normalizingconstantto make this estimateunbiased and consistent.Now, this1 will be equal to3, in case of this most,mostly use distribution known as this Gaussiandistribution or normal distribution. So, for this normal distribution, if wecalculate this quantity sothis measure of peakedness will become equal to 3.

Now, if I say, that this measure of peakedness is less than 3, then we will say that this peak is at a new is little bit lesser than this standardnormal distribution and for thathere. And if it is more, if it is peak is more, if it is more than 3 then it is more peak, than with respect to in compared to thatnormal distribution.So, this one sometimes in some of the text, this sample estimate is also expressed in terms of this minus 3, just to make that this k, if the k is equals to 0, then it is for the same to this normal distribution. And if it is negative then it is lower than normal distribution, if it is positive higher the normal distribution, but this is also correct. So, we can, we have to take that this is equals to 3in case of this normal distribution.

So, with this what is a thereafterwards you can now, understand that we have discuss about this first moment, second moment, third moment, forth moment. And obviously, the first moment with respect to the origin we discuss,then we have seen that first moment with the respect to the mean is 0. Second moment with respect to the mean and we understoodwhat is this implication, thenthird moment with respect to mean,forth moment with respect to mean. And in this way we cango on increasing, we can go for the fifth,sixth and seventh moment and for each cases you will get sum of this descriptors and up to this fourth oneit isalmost describing almost everything.

But still we will say, if we can get, that theif we get the measure of this all this kinds ofcoefficient that is all this moments with respect to mean. If we say, thenwe can equivalently, we can say that this the all the properties of this p; of thepdfis known to us. But so, in that one we will just see in a minute, that how this one is deleted, howwe can say that in a from a singlefunction, how can get all thismeasures, all this moments that we will see.But,before that these moments this analogy of thisnow, why this is calledmoment that is it is analogywith respect to the area of thepdfwill be discuss now.

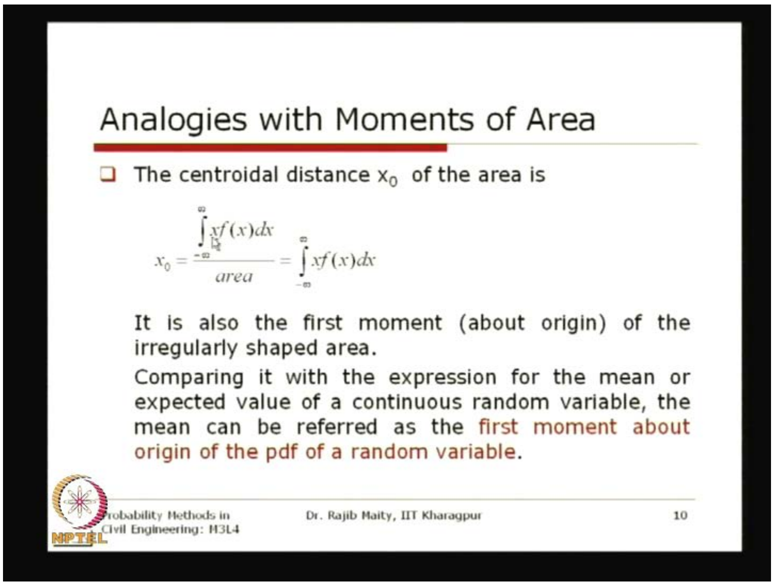
(Refer Slide Time: 21:53)



So, now see if we say that this is one of the standardpdfthat we have shown here. Now, it ismean will be some point where it is cg is.Now, considering a unit area having the general shape as shown in this following figure.Now, this is if I take case a very small area very small length along this X axis which is d x and if we calculate, what is the total moment and this is your origin. So, if we calculate, what is the moment, that is due to this particular area then this distance will be multiplied by this total area to get this one.Now, if we integrate this one, this fullarea, then we will getwhat is the total moment of this area with respect to the origin.

Now, if we divide that quantity with respect to this total area, then we will getonedistance that distance from this origin that is passing through it is cg. So, this is, this show and we know that from this properties of this standard,properties of thispdfthat this area total area is equals to 1.Now, if we justsee this discussion here, either the centroidal distance x naught of this area, that is x naught is equals to. So, this one what we are doing is that? That particular area that is f x multiplied by d x. So, what is this? So, this one, this height here which is that f x at the location x that multiplied by this smallin fact, small length d x thus is giving this area.

(Refer Slide Time: 23:27)

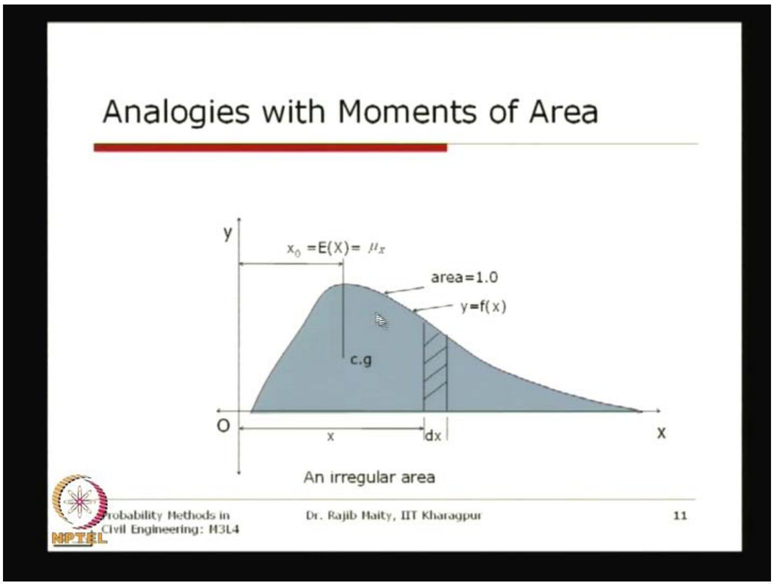


That one multiplied by this x is giving with this moment. And if we integrate from minus infinity to plus infinity,we are getting the total moment and that divided byarea; obviously, is giving the distance of this centroid of this particular area. And we knowthat this total area is 1 sothis1 is equals to x, the centroidaldistance from the origin is equals to integration minus infinity plus infinity x multiplied by f x d x. Now, what is this again? This is nothing but the expectation of the x.The way we define that expectation of the x is, that x multiplied by f x d x. So, this distance is thecentroidal distance of this area from the origin is nothing but it is mean the distance from the origin to this one is nothing but it is mean.

So, it is; so, as this one is also the moment withabout origin of this irregularly shaped area. Just comparing this with the expression for the mean or the expected value of a continuous random variable, the mean can be referred as the first moment about the origin of thepdfwith respect to the random variable. So, this is how we justjoin the analogy between thisanalogy between thistotal area taking it is with respect to this moment and multiplied by it is distance from the origin. So, this first moment about the origin is nothing but is it is mean which is equals to this function is nothing but equals to it isexpectation.

Now, if we want to take the second moment with respect to the origin then this area multiplied by this distances square will give you divided by total area, will give you the I amsorry. So, this area multiplied by this distance square willgive you that second moment.

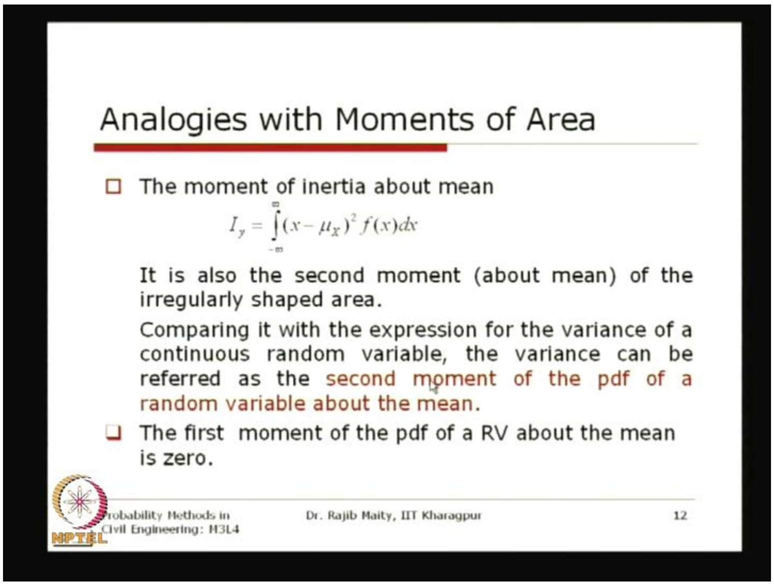
(Refer Slide Time: 25:27)



Now, this second moment the; what we are; what we have seen is that with respect tothis with the instead of taking with respect to the origin for our convenience, we take the second moment onwards we take that with respect to this location of this mean. So, this x naught in case of this first moment, this x naughtwhat we have shown that as a distance from the origin to this centroid of this area isnothing but from this expression which is reflecting that this is nothing butyourthe expectation of this x which is nothing but this mean.

So, this point is the distance from this origin to this mean now, the second moment.Now, what I told I repeat, that from this second moment onwards, we calculate it with respect to this mean and we take this distance from this mean to this inferential small areaand multiplied by it is distance and calculate it is moments.

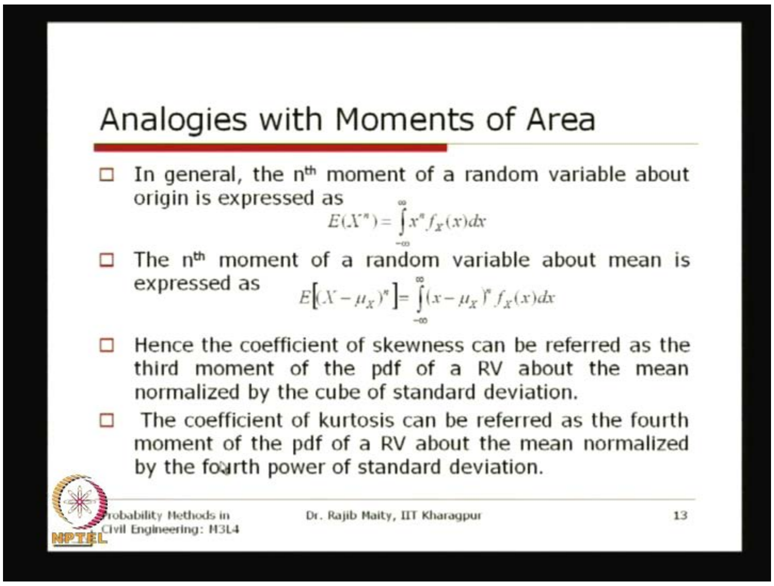
(Refer Slide Time: 26:29)



So, the moment of inertia about the mean, that I was telling that the second there is moment of inertia, second moment that is x minusmu x. So, this x minus mu x since, coming from that as, because we are taking it with respect to the mean. So, the distance now, this function which we are taking first only x because that was from the origin that wasactually that was x minus 0.So, as we are taking it with respect to the mean, we are taking that x minus mu x andsecond moments, we are taking it is square multipliedby this small length d x and it is value at that point x. And this quantity, we are integrating it for this full support area to get this one.

So,this is also what we see is that, this is also now.Now,this one from thisearlier expression of this moments, this is the second moment about the mean of thisirregularly shaped area. So, comparing it with the expression for the variance of a continuous random variable, the variance can be referred as the second moment of thepdfof a random variable about the mean, what wediscusssofar. And also in the last class, we at the towards the end, we discuss the first moment with respect to, first moment of thepdfof a random variable about the mean is zero.And this is discussedwhy it will bezero, because whatever the right hand side of this mean and left hand side of this mean and power; obviously, will be zero,will be negative each, other will cancel each other that is why that moment will always becomezero,withrespective of the safe of the p d f.

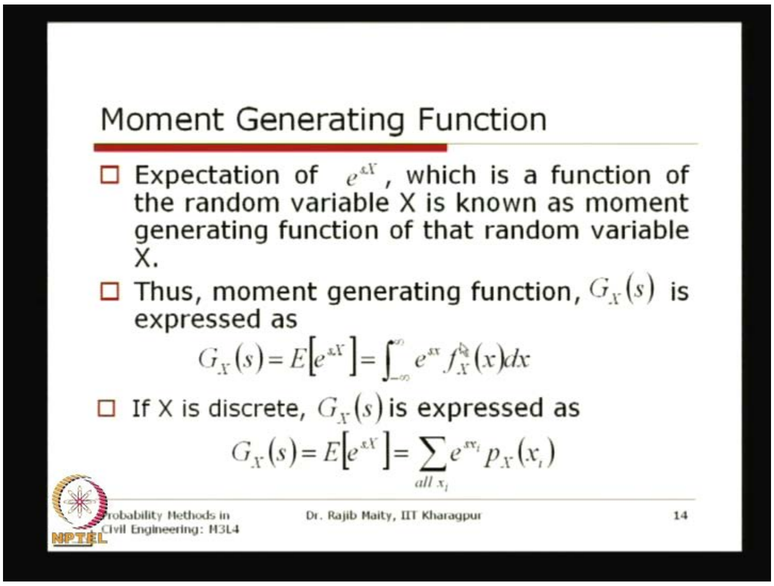
(Refer Slide Time: 28:18)



So, in general if I see, that nth moment of a random variableabout the origin is expressed as expectation of X power n. If I just now, we are taking as the with respect to the origin that is why x power n is minus infinity to plus infinity x power n f x of x d x. And the similarly, that nth moment of the random variable about mean is expressed as that, expectation X minus mu x power n is equals to minus infinity toplus infinity from this x minus. So, this expression that we have seen that, this is for the nth moment with respect to the mean.Hence the coefficient of skewness can be referred as the third moment of thepdfof a random variable about the mean normalized by it is cube of the standard deviation is exactly what we discuss,when we are discussing it isestimates.

So, this is the cube of the standarddeviation means, this we are taking the third moment to cancel this one to make it dimensionless, we are normalizing it by with respect to the cube ofthe standard deviation.Similarly, for this coefficient of kurtosis this can be referred as the fourth moment of thepdfof the random variable about the mean normalized by its fourth power of the standard deviation.So, this analogiesif you keep in mind, then you will know that why we call thisquantity as the moment of thepdfwith respect to origin or with respectto mean as the case may be.

(Refer Slide Time: 30:02)



Now, as justfew minutes back I was discussing that these four things, that we have discussed that is central tendency, variance, variation dispersion, thenit is measure of symmetry, measure of peakadness. All this things are the representation of one moment with respect to one moment, with respect to that;with respect to mean. And in this way we can go on increasing to take the moments from 5^{th},6^{th}, 7^{th},and in this way. So, every time I have to define that function I have tonormalize to this one.

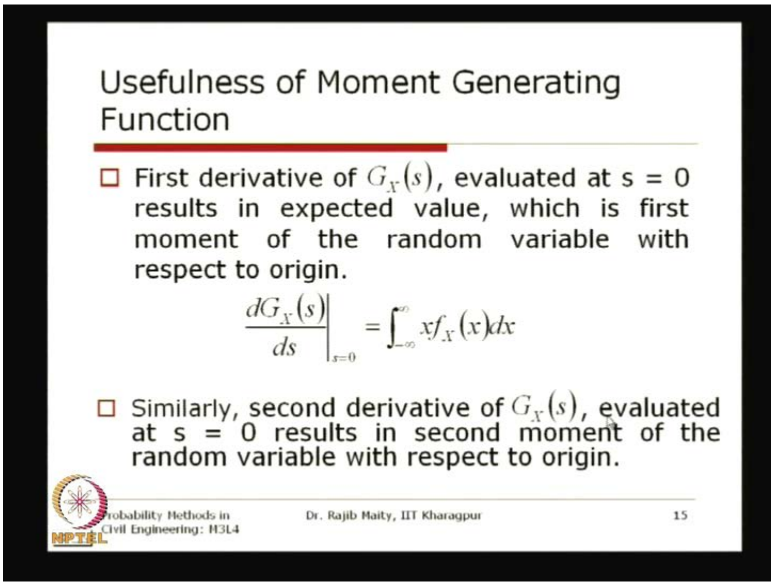
Now, this the next thing, that we are going to discuss is the moment generating function and we will see with the help of a single function, how all the moments can be known. And this is very important in the sense, if all the moments are known, then it is equivalent to know all the propertiesof thispdfthat is why, this moment generating function is very important and this is what we are going to discuss next.

So, this moment generating function, the expectation of epower s x this x is the random variable here. So, this epowers is again a function of the random variable. So, expectation of epower x s x, which is a function of the random variable X is known as the moment generating function of the random variable X. So, I have to take the expectation of this value. So, how this expectationlooks like.

So, this moment generating function which is denoted by G X s is expressed as G X s equals to expectation of epower s X equals to minus infinity to plus infinity epower s x this function I am taking and that multiplied by thisprobability density function and integrating with respect tod xminus infinity to infinity. So, this is; this function, this particular function is your that moment generating function.So,if x is discrete, then this is for this continuous that is why we took this integration.Now, if x is x is discrete then again show G X is expressed as the similar way to how we express that expectation of a function of random variable of a function of a discrete random variable, this is the summation of all x i epower s x i multiplied by the value of this probability allthose respective points.

So, this is your. So, this is, this expression for thediscrete random variable and this is for this continuous random variable is nothing but this moment generating function.Now,we willsee why these things are here, why this how these things are useful and why these things are important.

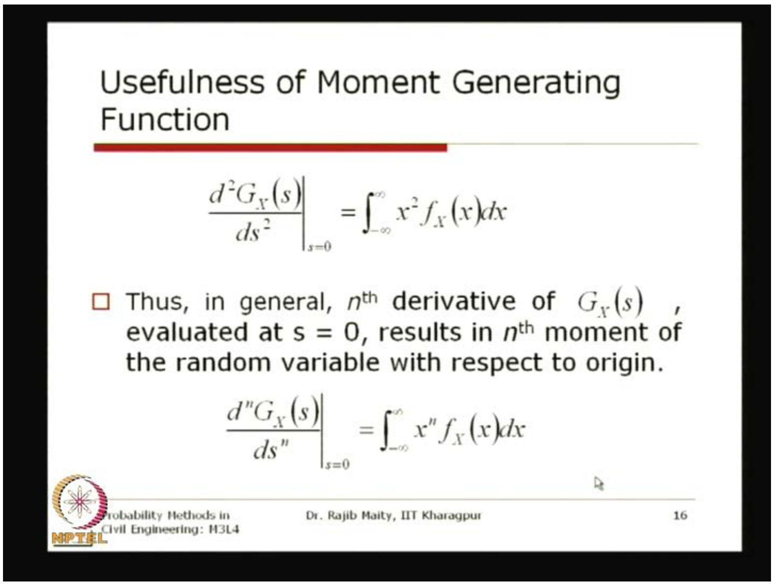
(Refer Slide Time: 32:59)



And these usefulness of this moment generating function. That first derivative of that G X s, evaluated as s equals to 0 results in the expected value, which is the first moment of the random variable with respect to the origin.Now, derivative of this G s G X s and sothe this is the first derivative I am taking andafter this derivative I am putting that x equals to 0.So, by this simplecalculus,we will get that from this will basically turn to that minus infinity to plus infinity 1 x we will get due to the derivative and epower s x when we are putting s equals to 0 that becoming 1 sothat is 1 multiplied by f s x d x.Now, this function is nothing but the first moment of that of the p d f. So, when we are taking this derivative, when we are taking the first derivative, the relative function is yourisyour first moment with respect to the origin.

Similarly, the second derivative of G X s, evaluated at s equals to 0 results in the second moment of the random variable with respect to the origin.

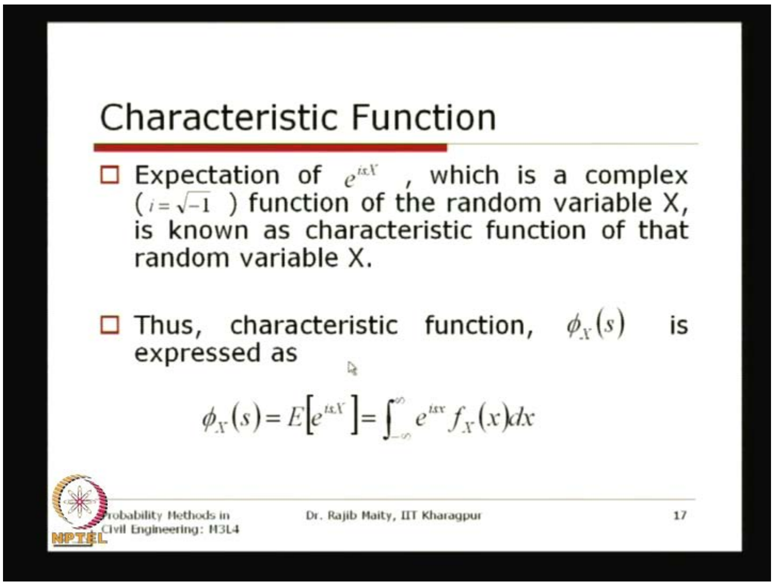
(Refer Slide Time: 34:15)



Let ussee. So, we are taking this derivative double; the double derivative an evaluated with respect to x. So, this x now, is becoming x multiplied by x that is x squire multiplied epower s x s equals to 0. So, epower s x is equals to 1 multiplied by this f x x. So, this is nothing but again that second moment with the respect to, this is expression for the second moment with respect to origin. So, thus if we just continue, thefirst moment, second moment and in general, if we want to express that the nth moment, nth derivative of G s x, evaluated at x equals to0x equals to 0 results in the nthmoment of the random variable with respect to origin.

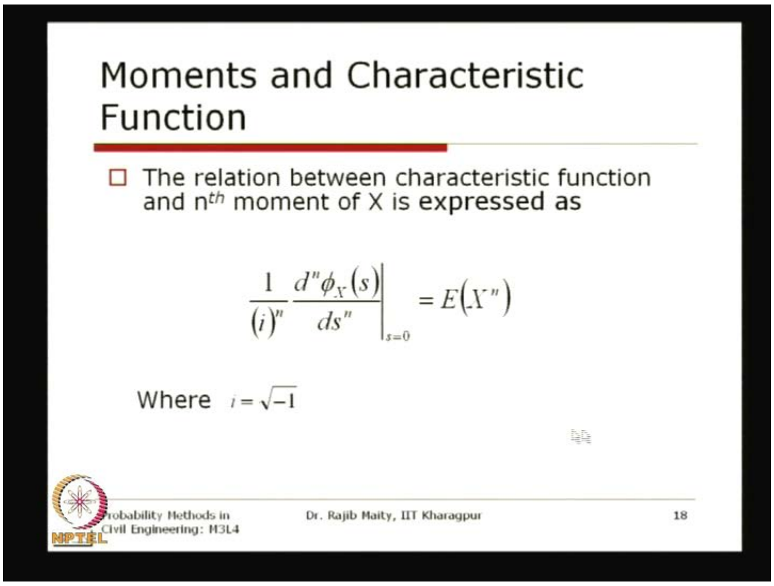
So, this that nth derivative here againsothis as, we are taking this nth derivative, this x power is becoming n and which is nothing but the expression for the nth moment of therandom variable basically, thepdfof the random variable here.So,this is the, this is how we can use this one. So, if we know this G X.Now, we can take the derivative of whatever the derivative that we need. So, whateverorder of this derivative, that we need that thenorder. So, this n can be in general. So, nth order derivative, we can take andweput that x equals to 0. So, we will get directly that the nth moment nth moment. So, this is the usefulness of this moment generating function.

(Refer Slide Time: 35:53)



Now, there is another onecalled the characteristic function, this expectationof this epower i s x. Now, instead of the putting that s x, we are puttingone complex variable that is i is equals to square root of minus 1. So, ifwe take the expectation of this one this expectation is known as the characteristic function of that random variable x. So, and it is denoted as this characteristic function is, denotes denoted as phi x s. So, we know if it is a expectation of this one. So, this can be expressed as that expectation of this epower i s x, which is the integration of this minus infinity to plus infinity epoweri s x f x xd x. Now, how we can relate this particular function with respect to the moments that.

(Refer Slide Time: 36:51)

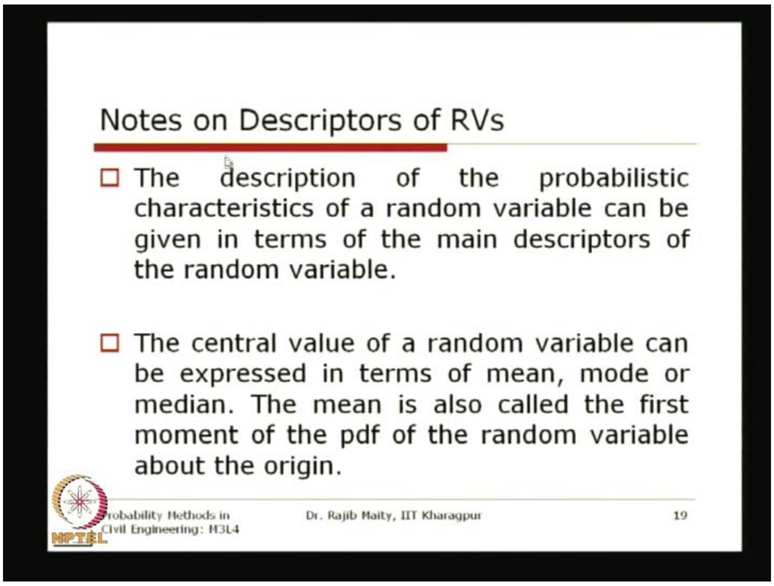


So, we can take that again for this characteristic function also, we can take that nth derivative of thatone and we canmultiply it by 1 byipower n. And thatfunction is evaluated at x equals to 0, which is nothing but we will give you that expectation of X power nthat is the minus infinity to plus infinity X power n multiplied by f x x d x. Where iequals to 1,this characteristic function,use of this characteristic function is mostly is seenin the electrical engineering not much in the civil engineering, but this moment generating function isimportant in this field.

Now, as we have seen that, we have somedescriptors of thisrandom variable and that random variable, we have discussed. And from this;now, if we have some sample data that is available to us, now that sample data we have to first of all represent graphically,just to see how it is shape look like, how it is disperse sothathow it is displayed. Sothat, we can get some idea about that; about it is distributionpackand about it is distribution. So, nowthese graphical representation of thissample data,sample data mean for any random experiment that outcome, that data that we are having in the contest of this civil engineering. We can take that screenflow data, rainfall data are the strength of theconcrete number of accidentson the particular stretch of high way. So, these are some sample data that we can see.

And now, if we want to see that how this random variable behaves. So, we have to plot that data first and we have to see that how this distribution look like. And there are few techniques how graphically we can represent that particular dataset. So that is what we are going to discuss nextto different ways different popular ways, how we can represent the sample data in terms of the graphical waves.

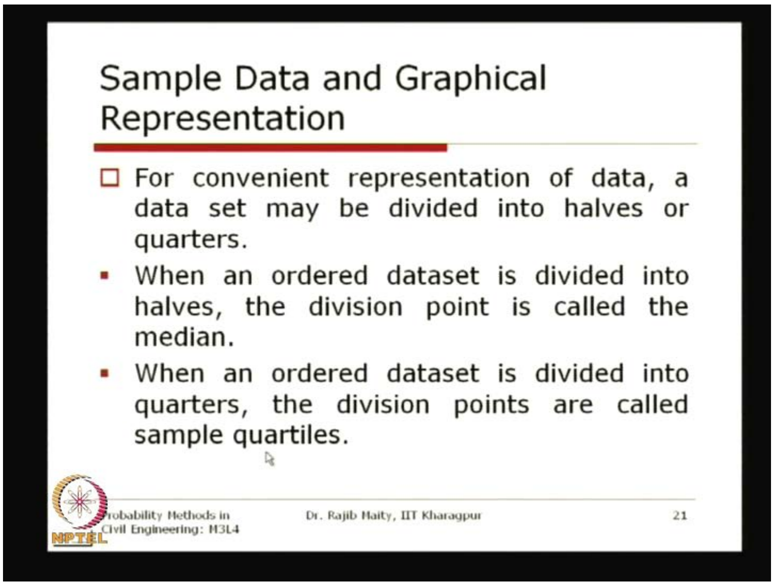
(Refer Slide Time: 39:08)



Before that,we will just to conclude that the descriptive of the random variable,we will put some note on this. That the description of the probabilistic characteristics of a random variable can be given in terms of this main descriptors of this random variable. The central value of a random variable can be expressed in terms of mean, mode or median that we have seen in the last class mostly that mean is also called the first moment of thispdfof the random variable about the origin.That we have seen, when we are discussing thatrepresentation ofthisarea that is analogy with this moment of the area p d f.

Then the dispersion, asymmetry and peakedness of the pd f of a random variabledescribed in terms of their secondmoment, third moment and fourth moment of thepdfof the random variable about it is mean this is important. So that when we are taking this, when we are want toknow that is dispersion, asymmetry and peakedness, we are taking that moment with respect to it is mean.And last, we saw thatthis moment generating function and this characteristic function are helpful, because these help to obtain all the moments of a random variable in an alternative way.

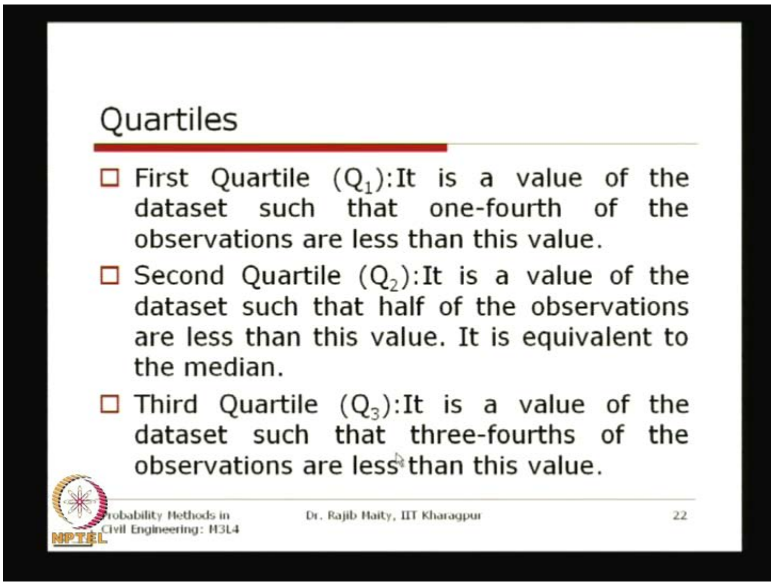
(Refer Slide Time: 40:36)



Now,what just now, we arediscussingthat we will know see how what are thedifferenttechniquesthat this sample data and this sample data how we can represent in termsin graphically,you can represent that sample data.

So, forconvenient representation of data, a dataset may be dividedinto halves or quarters.When an ordered dataset is divided into halves, the division point is called the median. So, ordered dataset here means, thatwe are just arranging the dataset in an increasing order and we will see that where it is exactly dividing it between two and that division point, that particular point is known as themedian. As you have seen that it is that50percent, when the 50percent probability is covered theyhave called it is median. Soandwhen an ordered dataset is divided into quarters,thatis four differentquarters, then the division points are called the sample quartiles.

(Refer Slide Time: 41:50)



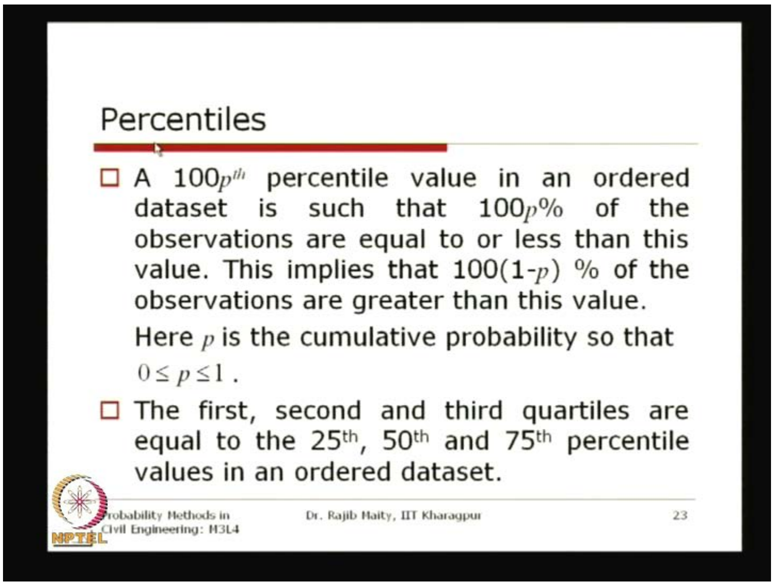
So, these are some terms that will know in detail that is first is that our quartile. So, thisfirst quartilesis the Q 1 represent as Q 1.It is the value of the dataset such thatone-fourth of the observations areless than this particularvalue. So, this is the first fourth. So, when we are taking this a dataset and we are just making it in an increasing order. So, it is that particular point when we say thatone-fourth of the observation as less than that particular value that is first quartile.

Similarly, the second quartile means, when we are talking about this 5th version of the data that is half of the data that is. It is the; it is a value of that dataset such that the half of the observations are less than this particular value. So, this is known as this,this is known as this quartile, which is also equivalent to the median, just what we discuss because it is the half halfof the data less than that particular value. And third quartile is similar to the earlier, but this is thatthree-fourth that is the 75 percent of the data when it is less than. So, it is a value of the dataset such thatthree-fourth of the observations are less than this particular value.

Now, why this things is important, this quartile as importantsothat if we know that the defined quartile, if know that total range of this dataset. And after that, if we know that what are it is quartile, then this will give you at; this will give you some idea that how the distribution of this of the dataset over a particular range. Whether ifyou see that there are lot of data is there within the first quartile or within thebelow the first quartile or in between the first and second quartile or whatever it is. So, this quartile when we are giving this quartile value along with the total range of this dataset this helps us to understand how dispersed the dataset is over it is entire range.

Otherwise the simply; as simply if we see that, simply if we see that what is this median value, that gives only that where the50percentdata is there. So, instead of that if we get a distributed fashion like this,then it will be helpfulto know it is dispersion and it will be moregraphically it is done, in terms of the different boxplotthat will come in a minute before that we will see that in terms of the percentile.

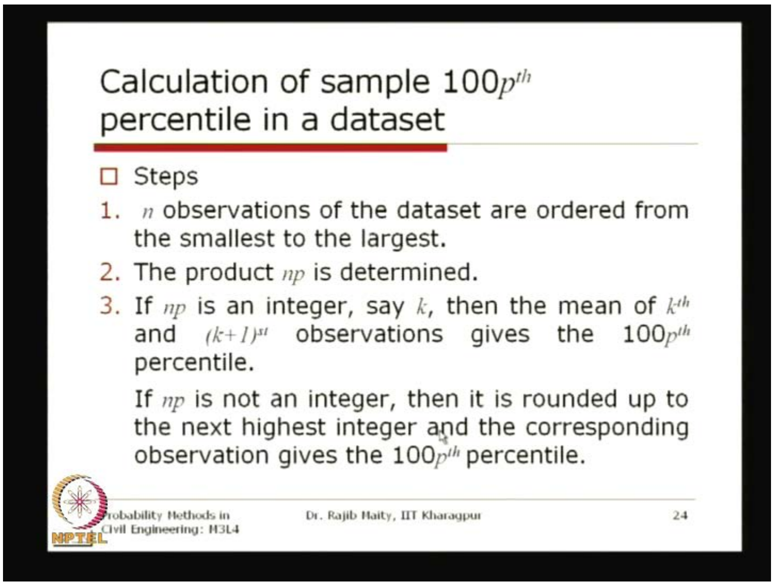
(Refer Slide Time: 44:29)



So, if when we talk about the percentile of a particular dataset that is available to us,that a 100 multiplied by pth percentile. So, before I go this one this p is your cumulativeprobability. So, here p;so, if we see that p is the cumulative probabilitysothat it variesfrom the 0 to 1.Now, for any percentilevalue if we sofor the p equals to 0.1 means, this quantity will give you the 10th percentile or p equals to0.6 will give you the 60th percentile value. So, ingeneral the 100pth percentile value in an ordered dataset, ordered dataset means,you have justordered dataset means it is arranged in an ascending order ordered dataset. In such that the 100 multiplied by p percent of the observations are equal to or less than this particular value.

So, basically from the quartile to the percentile sopercentile we are making it more general and define for each and everyvalue that we need. So, by making this p is any number; is any continuous number from starting from the 0 to 1 put any value that particular percentile you will,we will be obtained.Now, if we just want to see that quartilewith respect to the percentile, then it is very easy to compare that thisfirst, second and the third quartileare equal to the 25^{th},50^{th} and 75^{th}, percentile values in an ordered dataset respectively. So, this first quartile is nothing but the 25th percentile, second quartile is nothing but the 50^{th} percentile and third quartile is nothing but the 75^{th} percentile.

(Refer Slide Time: 46:31)

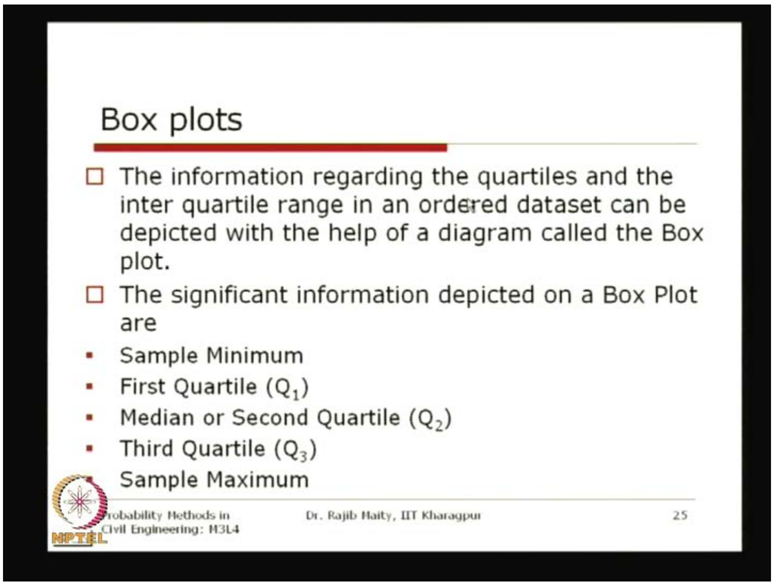


Now, this kind of representation if we want to present in a graphical way, then we will see thatthis can be done through a through a boxplot.Now, to prepare that boxplot we have to know certainwe have to know the percentiles, particularly for thoseat quartiles that is $25^{th}50^{th}$ and 75^{th} and sometimes the 5^{th} and 95^{th} as well. So, to get a particular percentile that is 100pthpercentile, how can we calculate the percentile.

So, these are the steps involved in that. So, suppose that there are total datasetnumber of dataset there is available is n. So, n observations of the dataset are ordered from the smallest to the largest that is in an ascending order,the available datasetof size n is arranged first. The product of n p is determinedsoif p is known a particular desiredpercentile is known. So, that particular p value is known to you. So, I know thus total number of dataset available n. So, n p this product that multiplication of n and p is known.Now, if this n p is an integer. So,n p can becomeone integer or it can be anditmay not be an integer. So, if it is an integer say that integer is k, then the mean of the kth and k plus first observation. So, from kthan immediate next observationthis two observations are taken and their mean is calculatedand that gives that 100pth percentile.

If n p is not an integer, then it is rounded up to the next highest integer and the corresponding observation gives that 100 pthpercentile. This can also be done with respect to that,that you can also linearly change this particular with respect to if it is not an integer, then thistwo values that is kthand in between this k plus firstobservation. In between this two, this also can be linearly; can also be linearly interpolated. Or for simplicity sake it can be taken to the nearest, a nearestinteger value and that particular value we will give the 100pth percentile.

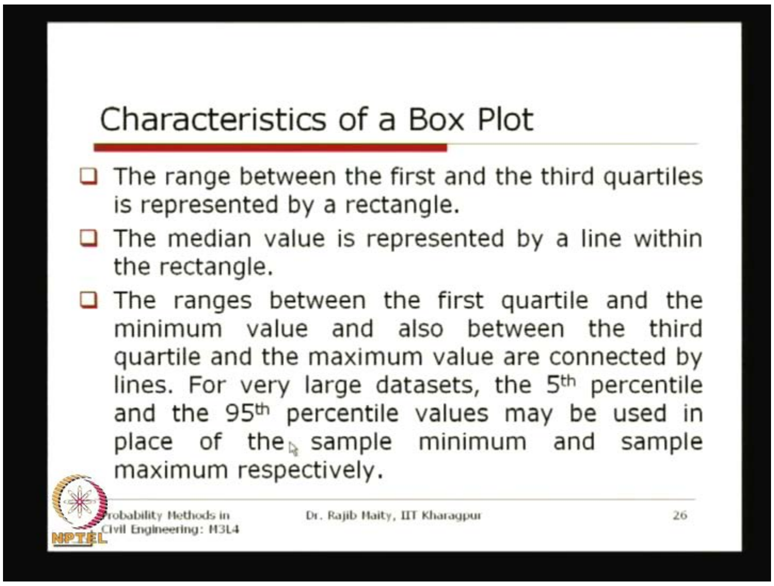
(Refer Slide Time: 48:53)



So, now in this way whatever the desired percentile is requiredthat is known to us and we can represent that information in terms of the box plot. So, the informationregarding the quartile and the inter quartile range in an ordered datasetcan be depicted withrespect;with the help of a diagram called the box plot.Now, the significant information depicted on a box plot are the sample minimum, first quartile, median or second quartile, third quartile and sample maximum.

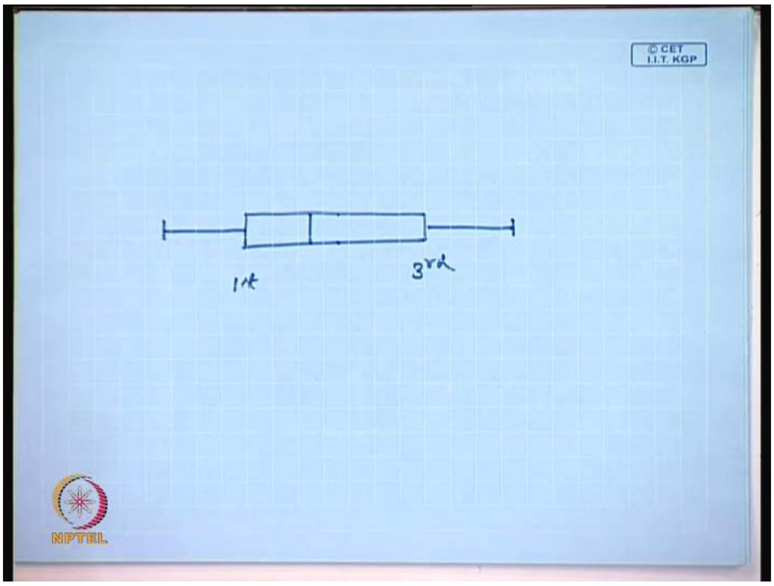
So, thisfive points are essential before we prepareonebox plots. So, we know that how to get that from a dataset, just know we have seen how to obtain this first quartile, how to obtain the second quartile and third quartile as well as a sample minimum value and as well as this sample maximum values.

(Refer Slide Time: 49:54)



Now, once we get this information, this is represented in such a way. That range between the first and the third quartile,first means that 25th percentile and third is the 75th percentile, the range between thesetwo quartiles is represented by a rectangle.

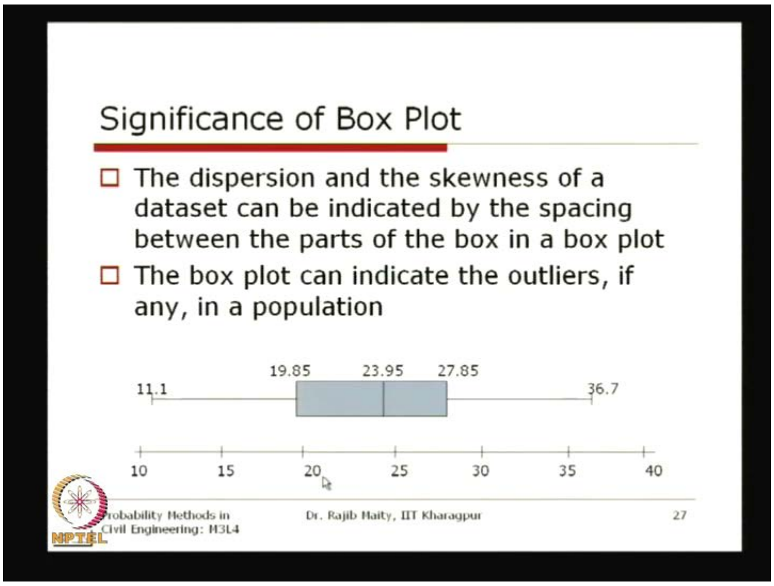
(Refer Slide Time: 50:23)



So,andthe median,so if I just one afteranother if I just draw it here,so thisone isyour, the 1st quartile and this one say is your is your 3rdquartile and this is representedin termsof a rectangle.Now, the second step is that the median value is represented by a line within this rectangle.Now, this median value, with this one is that it can be either, it is not necessary that we will be exactly middle of this rectangle; it can be any point either this side or towards that side it will be.Then,the range between thefirst quartile and the minimum value and also between the third quartile and the maximum value are connected by lines. Sothat the minimum suppose, there this is the minimum value and this the maximum value, then thistwo are connected by the lines.

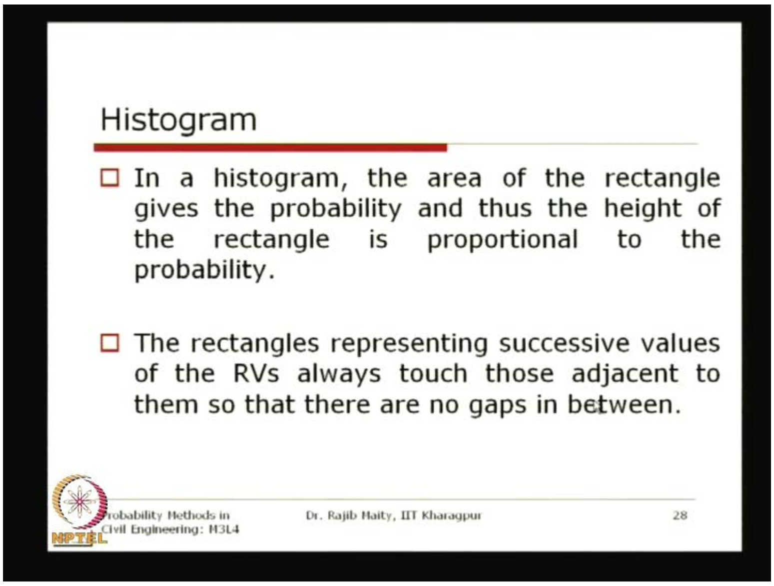
So for,sothistwo are connected by these lines.Now, again for the very large dataset that 5^{th} percentile and 95^{th} percentile values may be used in the place of the sample minimum and the sample maximumrespectively.

(Refer Slide Time: 51:39)



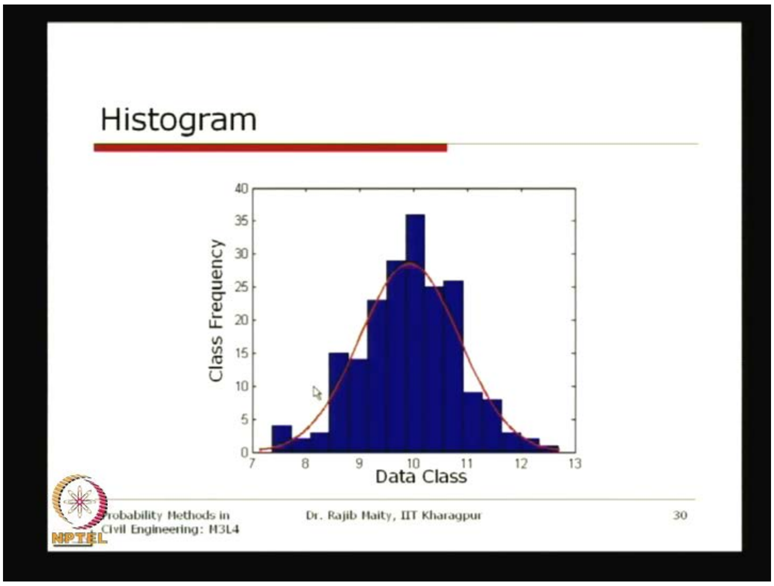
So, here is onerepresentation is shown, the dispersion and the skewness of the dataset as, we just discussis thatcan be indicated by the spacing between theparts of the box in a box plot.The box plot can indicate the outliers, if any, in the population. So, here is one example is given, so this is the real axisand where this box plot is shown. So, this is the 11.1 is the sample minimum or the 5^{th}percentile,19.85 is yourfirst quartile, 23.9 is median or second quartile,27.85 is the third quartile and 36.7 is theis 95th quartile or the sample maximum value.

(Refer Slide Time: 52:25)



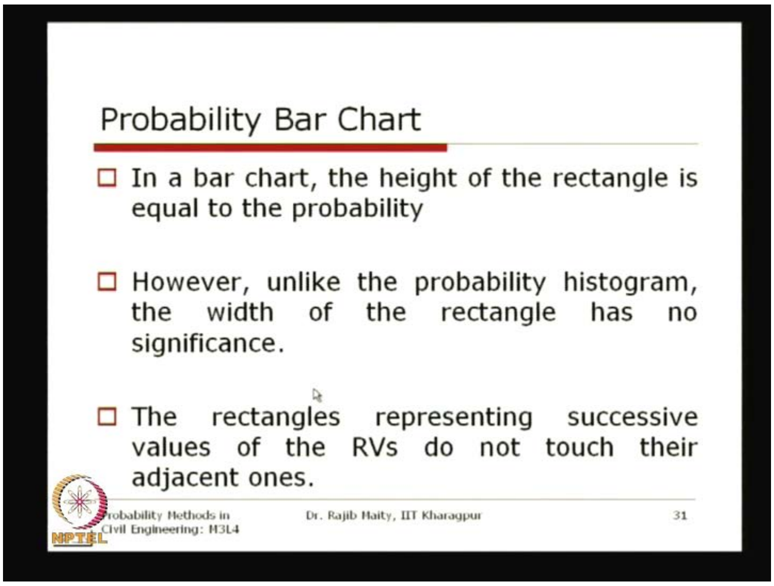
There are otherwise also, how we can graphically represent this one, the first one is the histogram. In a histogram, the area of the rectangle gives the probabilityand thus the height of the rectangle is proportional to the probability. The rectangles representing the successive values of the random variable always touch those adjacent to them sothat there are no gaps in between.

(Refer Slide Time: 52:50)



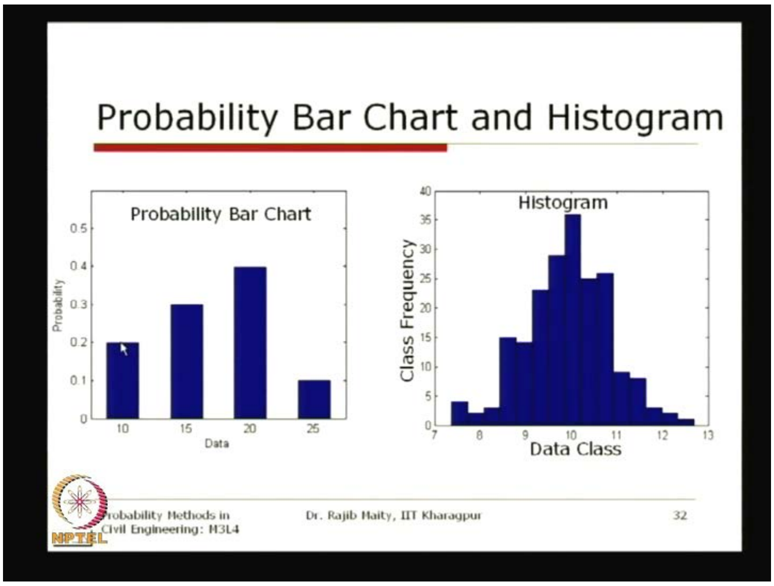
So, here is one example of this histogram is shownsothat any height of this rectangle is proportional to it is probability.Basically,in this range number of that frequency, in this particular class is shown by this numbers in this axis.So, obviously theheight of these rectangles are represented by this histogram.Now, how we can be useful to this one by graphically looking this one, we can see about is distribution how it is looks like and we can fit some our desireddistributionlike this, we can have some idea how the distribution is looks like.

(Refer Slide Time: 53:29)



But there is another one call the probability bar chart. So, in a probability bar chart the height of the rectangle is equal to the probability.However, unlike the probability histogram, the width of this rectangle has nomeaning; nosignificance. The rectangles representing the successive values of the random variables do not touchtheir adjacent ones.So, obviously there is no meaning only the height of this bars are important and which is showing yourprobability.

(Refer Slide Time: 53:59)



So, here is onecomparison when we call that probability bar chart. That mean this, at this particular data point, this isthe probability associated with thisone is 2; this can also be scaled, sothatthis height of this rectangle can, we can thattime will say that it is proportional to the probability. So, these are the how we can represent. So, this is the histogram, how we can represent the sample data,just to get the ideahow it is distributed over the entire range of this sample data.

So with this,we will stop today and next class we will see some description of this discrete random variable. So, today what we learned is that some moredescriptors of these random variable and their analogy with this moment of area, area represented by this pdf. And then we have seen that how the graphically, we can represent the sample data to get the idea of their distributionover the entire range of thissample data. So, we will seenext lecture with the description of thisdiscreteprobability distribution. Thank you.