**INDIAN INSTITUTE
OF
TECHNOLOGY
KHARAGPUR**

**NPTEL
National Programme
on
Technology Enhanced Learning**

Probability Methods in Civil Engineering

Prof. Rajib Maity

**Department of Civil Engineering
IIT Kharagpur**

Lecture – 09

Topic

Further Description of Random Variable

Hello and welcome to the forth lecture of our, this current module 3 this module is on random variable.

(Refer Slide Time: 00:27)

And in this lecture, we will cover that further descriptors of random variable. Basically, in the last class what we did we started with the description of CDF and after that, we started some of the descriptors of the random variable, if the in absence of the a proper definition of this pdf and CDF if we know some sample data then from the sample data there can be we can define some descriptors of the random variable and with that we get some idea about how the particular random variable pdf distribution the behavior of the pdf we know.

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

(Refer Slide Time: 01:40)

Outline

□ Skewness, Kurtosis of RVs

□ Analogy with moments of area

□ Moment Generating function

□ Characteristic function

□ Sample Data and Graphical Representation

Probability Methods in Civil Engineering: M3L4   Dr. Rajib Maity, IIT Kharagpur   2

This skewness and kurtosis are the higher moments as we have just described being the last that analogy, that we have given that analogy with these moment so these are higher with respect these are the higher movements with respect to the mean and standard deviation so we will see

that one, first the 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 of all such moments that is basically this mean standard deviation or mean, variance, skewness, Kurtosis this is up to fourth and above also that higher moments 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 order moment of any other how it can be defined with of this two five functions. Then we will start because these are for this for some distribution for if some sample data is available then how we can that that particular sample data we can represent graphically so that we can have some idea about this descriptors with respect to it is graphical representation. So this is we will try to cover today's lecture and we will start with the skewness.

(Refer Slide Time: 03:09)

Measure of Skewness

□ The skewness of a random variable is the asymmetry of its probability distribution. A measure of skewness may be expressed as $E(X - \mu_x)^3$

- For a discrete random variable X with pmf $p_x(x_i)$,

$$E(X - \mu_x)^3 = \sum_{all\ x_i}(x_i - \mu_x)^3 p_x(x_i)$$

- For a continous random variable X with pdf $f_x(x)$,

$$E(X - \mu_x)^3 = \int_{-\infty}^{\infty}(x - \mu_x)^3 f_x(x)dx$$

Probability Methods in Civil Engineering: M3L4   Dr. Rajib Maity, IIT Kharagpur   3

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 – µx. If you recall that in the last class what we discuss that the first moment

when we are taking when we are looking for the central tendency we take the moment with respect to the origin and so it gives that the distance from the origin to that pdf which is nothing but that it is mean in the location of it is mean that wills shown also minute in terms of some graphical representation and now, all the higher order moments if we take it 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 are talking about this is also a moment with respect to this mean and that mean it is the third moment and second moment we discussed in this last class last lecture that it was a variance so this is a measure of symmetry about this mean so how this particular distribution if we see it here then we will come to know that.

(Refer Slide Time: 04:37)
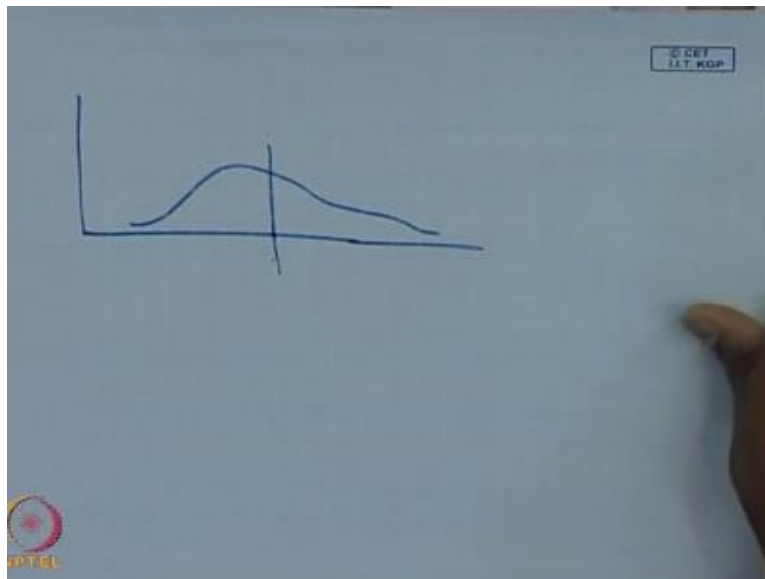
Now, from the origin I know how it is distributed and all so we know that where it is the location of this mean. Now what we are trying to do from with respect to this symmetry that with respect to this mean weather it is symmetric or not so if it is not symmetric then how it is distributed about the mean that is the goal so that is why when we are taking the third moment X - μ so this

sign remains means it will show that if 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 is more this side then it will be this will be negative and if it is this side then it will be positive so what we will do we will just take the expectation of this one.

(Refer Slide Time: 05:29)

Measure of Skewness

□ The skewness of a random variable is the asymmetry of its probability distribution. A measure of skewness may be expressed as $E(X - \mu_x)^3$

- For a discrete random variable X with pmf $p_x(x_i)$,

$$E(X - \mu_x)^3 = \sum_{all\, x_i}(x_i - \mu_x)^3 p_x(x_i)$$

- For a continous random variable X with pdf $f_x(x)$,

$$E(X - \mu_x)^3 = \int_{-\infty}^{\infty}(x - \mu_x)^3 f_x(x)dx$$

Probability Methods in Civil Engineering: M3L4   Dr. Rajib Maity, IIT Kharagpur   3
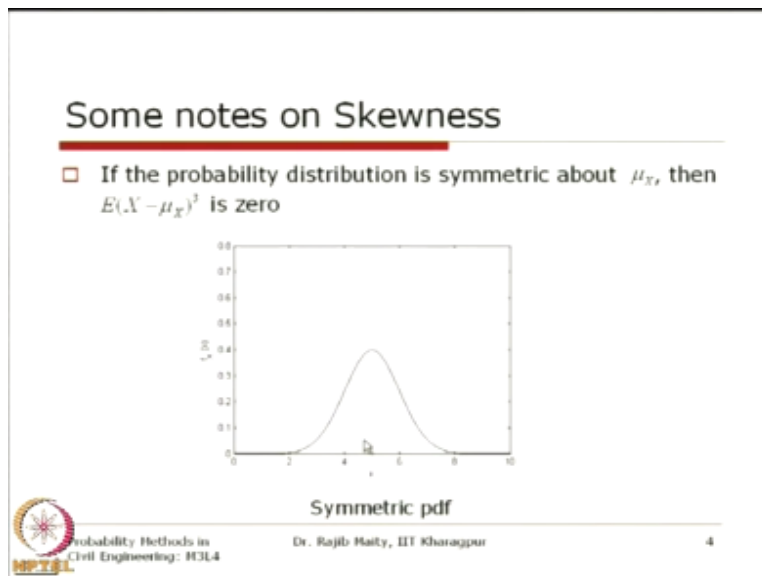
As usual for incase of this discrete random variable and incase of thus in case of continuous random variable for this discreet the expectation we know that. Now this is becoming if you see this is becoming a function of this random variable. So X is our random variable and this $(X - \mu_x)^3$ is a function of our random variable X. So, we know how to take the moment or we know how to take this functions as the expectation of a function that we described in the last lecture with respect to the g x.

So this function will come will be multiplied by the that probability density were the for the discrete it is the probability value for that particular outcome and we will sum up all this things to get the get what is the moment with respect what is the third moment with respect to the mean and similarly for the continuous random variable X for it is pdf if it is pdf is defined by this $f_x(x)$.

Then this expectation of this function again we will be express in terms of a so this summation now is converts to this integration.

And this integration over this entire support of this random variable X, we have taken it from -∞ to +∞. 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 of that variable with respect to mean.

(Refer Slide Time: 07:05)

Some notes on Skewness

- If the probability distribution is symmetric about μ_x, then $E(X - \mu_x)^3$ is zero

Symmetric pdf

 Now if we now see that if the probability distribution is symmetric about this mean then it is this moment will be equal to 0 this will be 0 because whatever now if you see here now if I draw one.
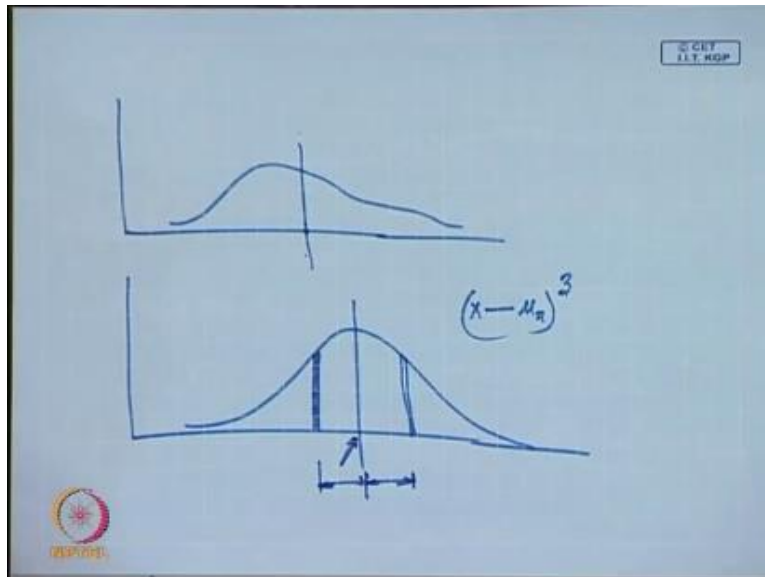
Particular distribution which is symmetric now this is your location of mean now, what we are doing actually we are taking a small area and as this is that distance as what we are taking $(X - \mu_x)^3$. So that whether this is on the negative side with respect to this mean or this is so this is on the positive side with respect to this mean. Now being there being it symmetric with respect to this one being this power 3 this two are cancels out.

So, if this one is symmetric then this moment will be will be exactly equals to 0 so this is what is explained here.
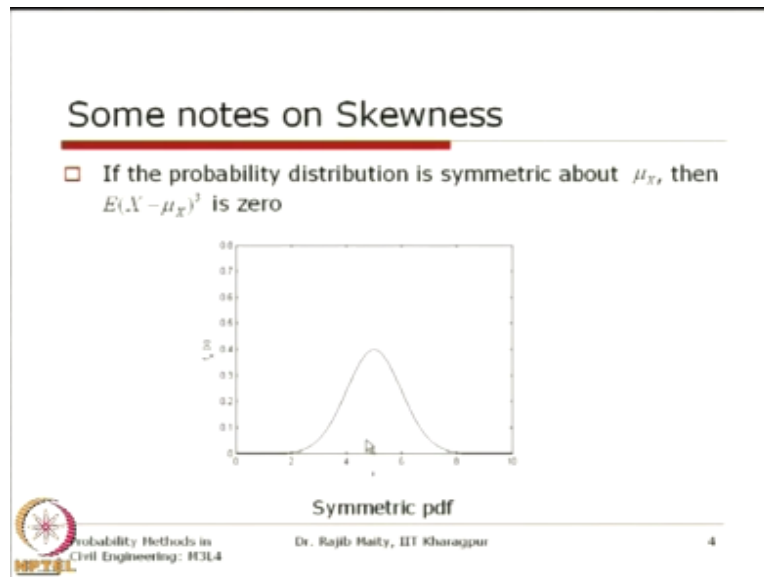
If the probability distribution is symmetric about μ_x then this expectation of $(x - \mu_x)^3$ this is third moment is equals to 0, which is one example is also shown in this diagram. Now we will compare with this symmetric this black one again we have just written which is we know that this is a symmetric or which this third 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 this one is positively skewed how it is distributed.

If for $X > \mu_x$ the values are more widely disperse than for the $X <$ equals to μ_x then this will obviously become positive. So if this is becoming this is becoming positive, then this skewness of this blue pdf is a positively skewed we say it is positively skewed. Now this one is greater than for this zone now, if I just saw if I just draw another one here what is meant by this one.
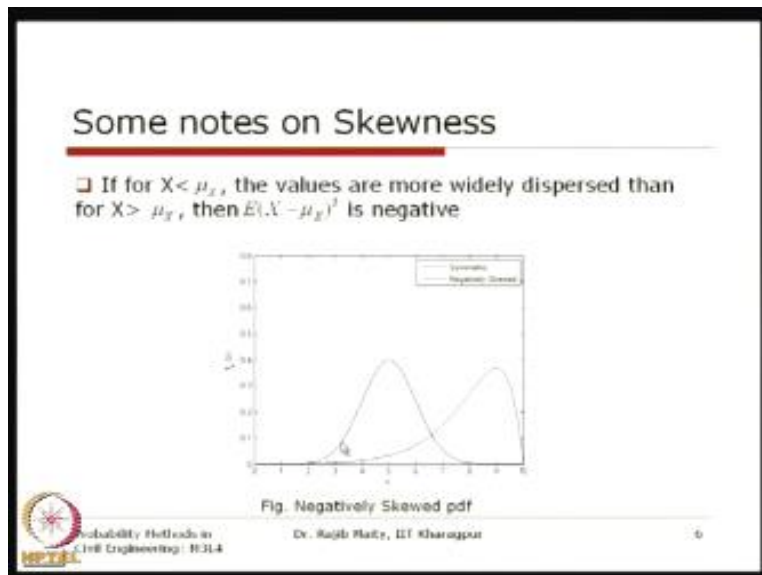
(Refer Slide Time: 09:34)
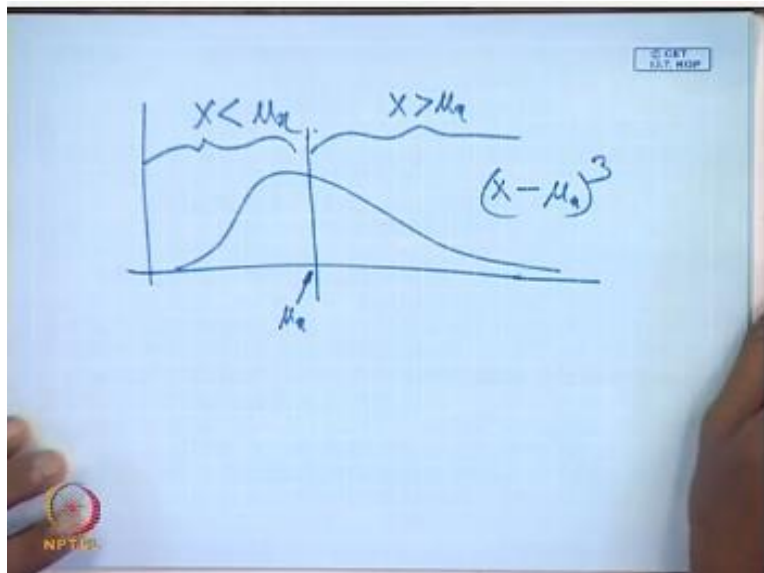
Is now as this side is disperse if we take this mean may be the mean will be somewhere at this location this mean will come here. Now, if we see that data range from the from towards the negative side it was left side of this mean and the right side of this mean that is this is your μ_x. So, this side if I say this is your $X > \mu_x$ now and this side is your $X < \mu_x$, now you can see that this side only widely spread this is widely spread than compared to this one this is lesser this is more dense compared to thus right hand side. Then when we are taking that $X- \mu_x{}^3$ this is giving you more weightage that is why the total summation will be positive in case when it is skewed like this.

(Refer Slide Time: 10:39)
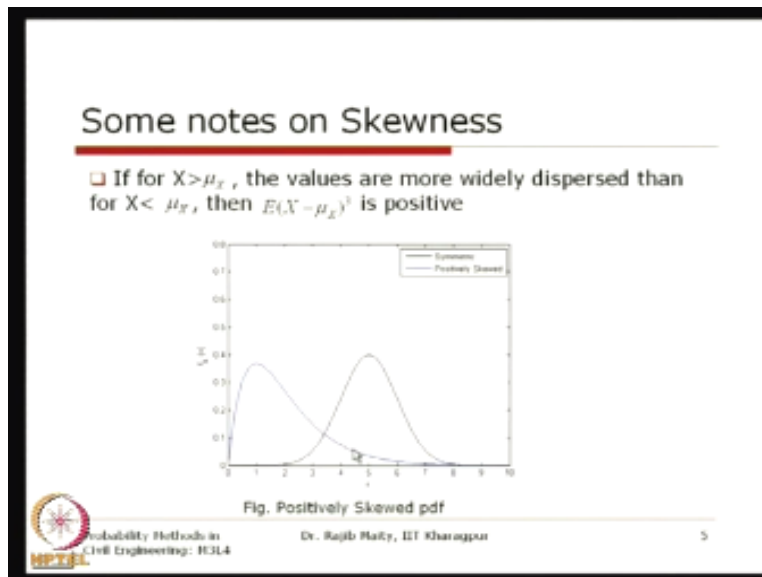
So that example is shown here so somewhere the main location will be in this here where the right side the positive side with respect to the mean will be more disperse than compared to this a left hand side. So that resulting in that so this side it is most this quantity will become positive, so this blue pdf what is shown here that is positively skewed.

(Refer Slide Time: 11:08)

Some notes on Skewness

❏ If for X< μ_x, the values are more widely dispersed than for X> μ_x, then $E(X-\mu_x)^3$ is negative

Fig. Negatively Skewed pdf

Just opposite if we see we will get that what is called the negatively skewed. So in this case it is just opposite that is $X \leq \mu x$ in this zone in this region the values are more widely disperse than for that $X > \mu x$. So somewhere in this location in this point this that mean will come and this side it will be less disperse compare to this the side. So obviously due to this power cube this quantity will become negative. So what is shown here by this blue line this pdf is known as negatively skewed pdf.

Now for up to this second moment what we have seen that it can be viewed as a 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 for our 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 be that convenient so make it convenient what we do is that value 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 becomes dimensionless. 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 area is the convenient.

Coefficient of Skewness

□ Coefficient of Skewness is a convenient dimensionless measure of asymmetry given by

$$\gamma = \frac{E(X - \mu_x)^3}{\sigma_x^3}$$

□ The sample statistic of the coefficient of skewness is given by

$$C_x = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)(n-2)s^3}$$

Where n= number of observations, $\bar{x}$ =sample mean, s= sample standard deviation

Dimensionless measure of a symmetry given by so this is your expectation of this function which is the third order 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 becoming dimensionless.

So this now the if this becomes 0 in case of the perfectly symmetric pdf 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 a symmetric pdf and if it is positive then it is positively skewed, if it is negative then it is negatively skewed. So this denominator is only to make this quantity dimensionless.

Now this is from the now if we have some sample 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 this coefficient of skewness is expressed as 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 up for 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 by this 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 and these are the normalizing constants, so this n divided by n- 1 multiplied by n/2 just to 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 x1 to xn.

(Refer Slide Time: 15:39)

Coefficient of Kurtosis

□ Coefficient of Kurtosis is a convenient non dimensional measure of peakedness given by

$$K = \frac{E(X - \mu_X)^4}{\sigma_X^4}$$

□ The sample statistic of the coefficient of kurtosis is given by

$$K = \frac{n^2 \sum_{i=1}^{n} (x_i - \bar{x})^4}{(n-1)(n-2)(n-3)s^4}$$

Now we will go to this forth moment, forth moment with respect to the mean and which is known as the kurtosis now so the measure of kurtosis is the fourth moment with respect to mean so we are taking the expectation of the function $X - \mu x^4$ So and again to make this one dimensionless what we are using is, we are using that this σx^4 power 4. Now as this again similar to this coefficient of skewness that is we are making we are taking the power of this standard

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? For example thus we have seen that mean what does it mean implies because this is a central tendency that then the coefficient of variation it is the this is indicating that it is dispersion with respect to mean, then skewnees is showing how whether it is most spread towards the right hand side of the mean or the left hand side of the mean. So that is the measure of symmetry.

(Refer Slide Time: 17:00)

Coefficient of Kurtosis

☐ Coefficient of Kurtosis is a convenient non dimensional measure of peakedness given by

$$\kappa = \frac{E(X - \mu_X)^4}{\sigma_X^4}$$

☐ The sample statistic of the coefficient of kurtosis is given by

$$K = \frac{n^3 \sum_{i=1}^{n} (x_i - \bar{x})^4}{(n-1)(n-2)(n-3)s^4}$$

Now what is this kurtosis if we see then we will see that this is also power 4 so as it is power 4 and obviously for the whatever on the left hand side and the right hand side that will be negated and that will make 0. Now this 4 is actually is a measure of peakedness. Now this measure of peakedness means how the peak how does that particular peak of this pdf look like. Now if we see here.

(Refer Slide Time: 17:30)
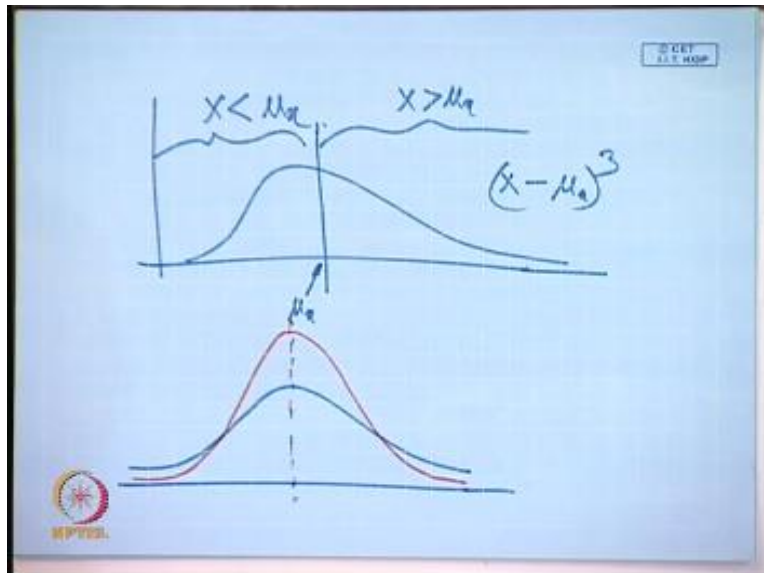
If we see that difference between say this is one so this is also a symmetric distribution and if I draw another one this is also another symmetric distribution and they are mean are also say same, so the first moment is same, second moment is the variance also will be same if we just see the total thing. So we will see that the variance of these two distributions with respect to the mean will be same.

Now the coefficient of the 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 fourth moment when it is normalized by the fourth power of this standard deviation which is known this coefficient of kurtosis, this is known as this measure of peakedness, okay.

So this measure of peakedness with this one having a sample statistics of this kurtosis 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 deviation and this normalizing constant to make this estimate unbiased and consistent. Now this one will be equal to 3 in case of this most mostly used distribution known as this Gaussian distribution or normal distribution.

So for this normal distribution if we calculate this quantity so this 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 standard normal distribution and for that yeah 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 that normal 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 = 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 there afterwards 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 understood what is this implication then third moment with respect to mean forth moment with respect to mean and in this way we can go 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 one it is almost describing almost everything.

But still we will say if we can get that the if we get the measure of this all this kinds of coefficient that is all this moments with respect to mean if we say then we can equivalently we can say that this the all the properties of this p of the pdf is known to us but so in that one we will just see in a minute that how this one is delete how we can say that in a from a single function how can get all this measures all this moments that we will see but before that these moments this analogy of this now why this is called moment that is it is analogy with respect to the area of the pdf will be discuss now so now see.

(Refer Slide time: 21:53)
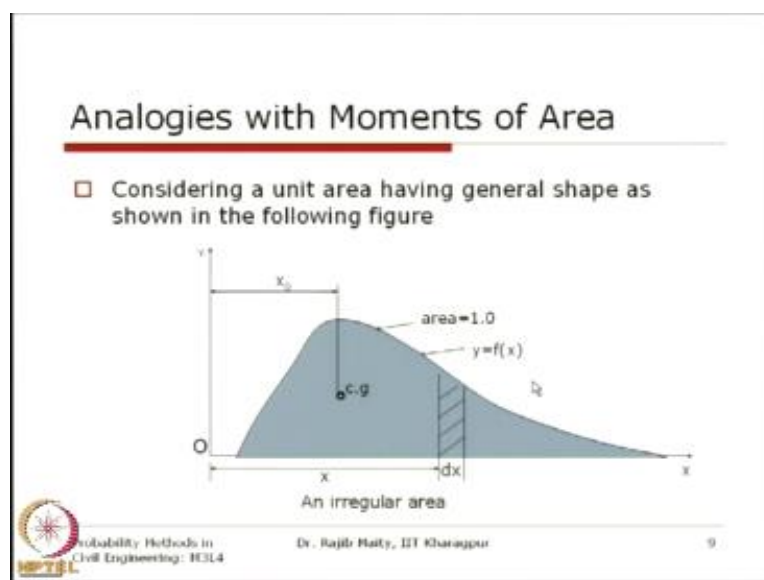
If we say that this is one of the standard pdf that we have shown here now it is mean will be some point where it is cg is now considering a unit area having the general shape as shown in this following figure now this is if I take case a very small area very small length along this X axis which is dx 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 introduce this one this full area then we will get what is the total moment of this area will respect to the origin now if we divide that quantity with respect to this total area then we will get one distance 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 this pdf that this area total area is equals to 1now if we just see this discussion here.

(Refer Slide time: 23:02)

Analogies with Moments of Area

❏ The centroidal distance x_0 of the area is

$$x_0 = \frac{\int_{-\infty}^{\infty} xf(x)dx}{area} = \int_{-\infty}^{\infty} xf(x)dx$$

It is also the first moment (about origin) of the irregularly shaped area.

Comparing it with the expression for the mean or expected value of a continuous random variable, the mean can be referred as the first moment about origin of the pdf of a random variable.

Either the centroidal distance x_0 of this area that is $x_0 =$ so this one what we are doing is that particular area that is f x multiplied by dx.

(Refer Slide time: 23:14)
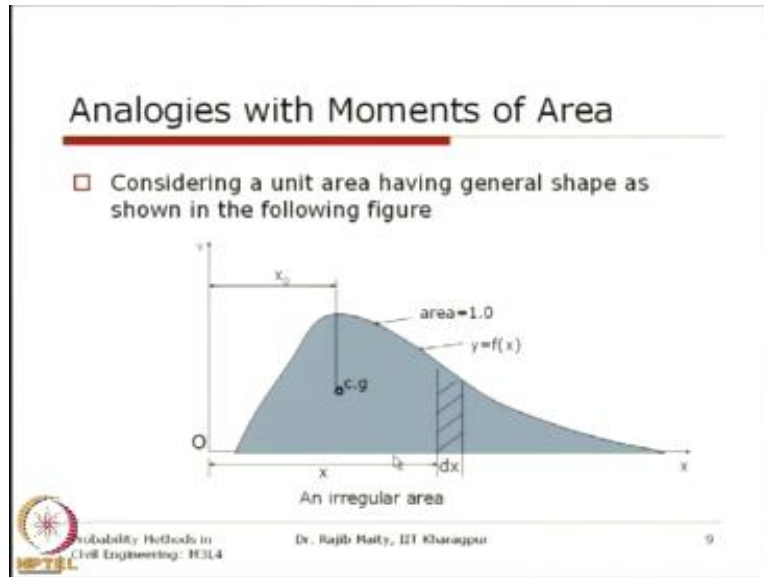
So what is this so this one this height here which is that fx at the location x that multiplied by this small in fact small length dx thus is giving this area that one multiplied by this x is giving with this moment and if we integrate from minus α + α we are getting the total moment and that divided by are obviously is giving the distance of this centroid of this particular are and we know that this total area is 1 so this 1 is equals to x the centroidal distance from the origin is equals to integration - α + α 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 the centroidal 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 with about origin of this irregularly shaped are 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 the pdf with respect to the random variable so this is how we just join the analogy between this analogy between this total 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 is

expectation 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 am sorry so this area multiplied by this distance square will give you that second moment now this second moment the what we are what we have.

(Refer Slide time: 23:14)
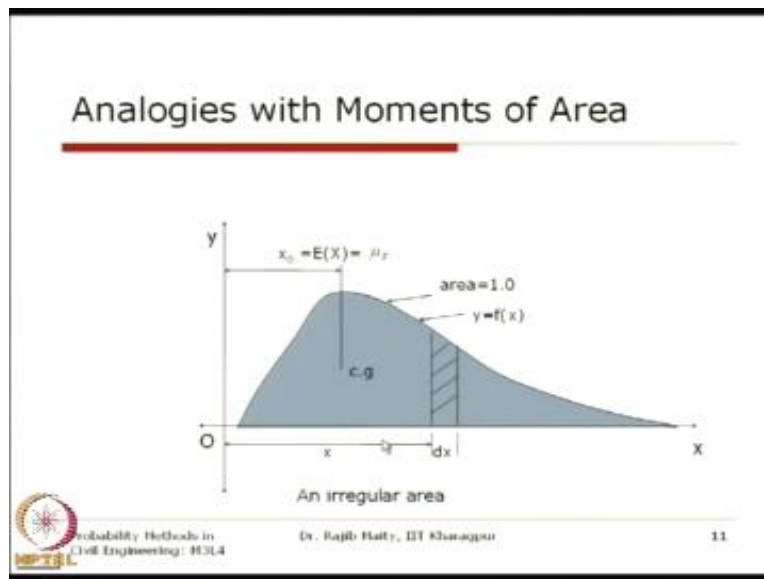
Seen is that with respect to this with the instead of taking with respect to the origin for our convenience, we take the second moment onwards we take that with respect to this location of this mean so this x_0 in case of this first moment this x_0 what we have shown that as a distance from the origin to this centroid of this area is nothing but from this expression which is reflecting that this is nothing but you're the 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 area and multiplied by it is distance and calculate it is moments.

(Refer Slide time: 26:30)

Analogies with Moments of Area

□ The moment of inertia about mean

$$I_r = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x)dx$$

It is also the second moment (about mean) of the irregularly shaped area.

Comparing it with the expression for the variance of a continuous random variable, the variance can be referred as the second moment of the pdf of a random variable about the mean.

□ The first moment of the pdf of a RV about the mean is zero.

Probability Methods in Civil Engineering: H3L4   Dr. Rajib Maity, IIT Kharagpur   12

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 – x so this x - μ 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 was actually that was x – 0 so as we are taking it with respect to the mean we are taking that x - μ x and second moments we are taking it is square multiplied by this small length dx 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 his is also what we see is that this is also now.

Now this one from this earlier expression of this moments this is the second moment about the mean of this irregularly 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 the pdf of a random variable about the mean what we discuss so far and also in the last class we at the towards the end we discuss the first moment with respect to first moment of the pdf of a random variable about the mean is zero.

And this is discussed why it will be zero 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 become zero irrespective of the safe of the pdf.

(Refer Slide time: 28:18)

Analogies with Moments of Area

- In general, the nth moment of a random variable about origin is expressed as
$$E(X^n) = \int_{-\infty}^{\infty} x^n f_X(x)dx$$

- The nth moment of a random variable about mean is expressed as
$$E[(X-\mu_X)^n] = \int_{-\infty}^{\infty} (x-\mu_X)^n f_X(x)dx$$

- Hence the coefficient of skewness can be referred as the third moment of the pdf of a RV about the mean normalized by the cube of standard deviation.

- The coefficient of kurtosis can be referred as the fourth moment of the pdf of a RV about the mean normalized by the fourth power of standard deviation.

So in general if I see that nth moment of A random variable about the origin is expressed as expectation of X^n if I just now we are taking as the with respect to the origin that is why X^n - α to + α X^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- μ X^n = - α to + α from this x - 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 the pdf of 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 is estimates. So, this is the cube of the standard deviation 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 of the standard deviation.

Similarly, for this coefficient of kurtosis this can be referred as the fourth moment of the pdf of the random variable about the mean normalized by it is fourth power of the standard deviation. So, this analogies if you keep in mind, then you will know that why we call this quantity as the moment of the pdf with respect to origin or with respect to mean as the case may be.

(Refer Slide Time: 30:01)

Moment Generating Function

☐ Expectation of e^{sX}, which is a function of the random variable X is known as moment generating function of that random variable X.

☐ Thus, moment generating function, $G_X(s)$ is expressed as

$$G_X(s) = E[e^{sX}] = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx$$

☐ If X is discrete, $G_X(s)$ is expressed as

$$G_X(s) = E[e^{sX}] = \sum_{all\ x_i} e^{sx_i} p_X(x_i)$$

Probability Methods in Civil Engineering: M3L4   Dr. Rajib Maity, IIT Kharagpur   14

Now, as just few minutes back I was discussing that these four things, that we have discussed that is central tendency, variance, variation dispersion, then it 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 5th, 6th, 7th and in this way.

So, every time I have to define that function I have to normalize 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 properties of this pdf that is why, this moment generating function is very important and this is what we are going to discuss next.

Moment Generating Function

☐ Expectation of e^{sX}, which is a function of the random variable X is known as moment generating function of that random variable X.

☐ Thus, moment generating function, $G_X(s)$ is expressed as

$$G_X(s) = E[e^{sX}] = \int_{-\infty}^{\infty} e^{sx} f_X(x)dx$$

☐ If X is discrete, $G_X(s)$ is expressed as

$$G_X(s) = E[e^{sX}] = \sum_{all\ x_i} e^{sx_i} p_X(x_i)$$

Probability Methods in Civil Engineering: M3L4 Dr. Rajib Maity, IIT Kharagpur 14

So, this moment generating function, the expectation of e^{sX} this X is the random variable here. So, this e^{sX} is again a function of the random variable X. So, expectation of e^{sX}, 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 expectation looks like, so this moment generating function which is denoted by G_X s is expressed as $G_X(s)=E[e^{sX}]= -\infty$ to $+\infty$ e^{sX} this function I am taking and that multiplied by this probability density function and integrating with respect to dx $-\infty$ to ∞.

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 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 $\sum$ of all xi e^{sx_i} multiplied by the value of this probability all those respective points.

So this is your so, this is, this expression for the discrete random variable and this is for this continuous random variable is nothing but this moment generating function. Now, we will see

why these things are here, why this how these things are useful and why these things are important.

(Refer Slide Time: 32:59)

Usefulness of Moment Generating Function

☐ First derivative of $G_x(s)$, evaluated at s = 0 results in expected value, which is first moment of the random variable with respect to origin.

$$\left.\frac{dG_x(s)}{ds}\right|_{s=0} = \int_{-\infty}^{\infty} x f_x(x) dx$$

☐ Similarly, second derivative of $G_x(s)$, evaluated at s = 0 results in second moment of the random variable with respect to origin.

And these usefulness of this moment generating function that first derivative of that Gx(s), evaluated as s=0 results in the expected value, which is the first moment of the random variable with respect to the origin. Now, derivative of this $G_x(s)$ and so the, this is the first derivative I am taking and after this derivative I am putting that x=0. So, by this simple calculus, we will get that from this will basically turn to that -∞ to +∞ 1x we will get due to the derivative and e^{sX} when we are putting s=0 that becoming 1 so that is 1 multiplied by $f_x(x)dx$.

Now, this function is nothing but the first moment of that of the pdf. So, when we are taking this derivative, when we are taking the first derivative, the relative function is your first moment with respect to the origin. Similarly, the second derivative of $G_x(s)$, evaluated at s=0 results in the second moment of the random variable with respect to the origin.

(Refer Slide Time: 34:15)

Usefulness of Moment Generating Function

$$\left.\frac{d^2 G_X(s)}{ds^2}\right|_{s=0} = \int_{-\infty}^{\infty} x^2 f_X(x)dx$$

☐ Thus, in general, n^{th} derivative of $G_X(s)$, evaluated at s = 0, results in n^{th} moment of the random variable with respect to origin.

$$\left.\frac{d^n G_X(s)}{ds^n}\right|_{s=0} = \int_{-\infty}^{\infty} x^n f_X(x)dx$$

Probability Methods in Civil Engineering: M3L4   Dr. Rajib Maity, IIT Kharagpur   16

Let us see, so we are taking this derivative double, the double derivative and evaluated with respect to x. So, this x now, is becoming x multiplied by x that is x^2 multiplied e^{sX} s=0, so e^{sX} X=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, the first moment, second moment and in general, if we want to express that the nth moment, n^{th} derivative sorry, n^{th} derivative of $G_x(s)$, evaluated at s=0 equals results in the nth moment of the random variable with respect to origin.

So, this that n^{th} derivative here again so this as, we are taking this n^{th} derivative, this x power is becoming n and which is nothing but the expression for the nth moment of the random variable basically, the pdf of the random variable here. So, this is the, this is how we can use this one. So, if we know this $G_x(s)$. Now, we can take the derivative of whatever the derivative that we need. So, whatever order of this derivative, that we need that the nth order. So, this n can be in general, so n^{th} order derivative, we can take and we put that x=0. So, we will get directly that the nth moment n^{th} moment. So, this is the usefulness of this moment generating function.

Now, there is another one called the characteristic function, this expectation of this e^{isX}. Now, instead of the putting that sX, we are putting one complex variable that is $i=\sqrt{-1}$. So, if we 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 e^{isX}, which is the integration of this $-\infty$ to $+\infty$ e^{isX} fx(x)dx. Now, how we can relate this particular function with respect to the moments that.

(Refer Slide Time: 36:50)

So, we can take that again for this characteristic function also, we can take that n^{th} derivative of that one and we can multiply it by 1 by $(i)^n$. And that function is evaluated at x=0, which is nothing but we will give you that expectation of X^n that is the $-\infty$ to $+\infty$ X^n multiplied by $f_x(x)dx$. Where i=1,this characteristic function use of this characteristic function is mostly is seen in the electrical engineering not much in the civil engineering, but this moment generating function is important in this field.

Now, as we have seen that, we have some descriptors of this random 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 so that how it is displayed. So that we can get some idea about that about it is distribution pack and about it is distribution. So now these graphical representations of this sample 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 screen flow data, rainfall data are the strength of the concrete number of accidents on the particular stretch of high way. So these are some sample data that we can say

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 next to different ways different popular ways, how we can represent the sample data in terms of the graphical views.

(Refer Slide Time: 39:07)

Moments and Characteristic Function

☐ The relation between characteristic function and n^{th} moment of X is **expressed as**

$$\frac{1}{(i)^n} \frac{d^n \phi_X(s)}{ds^n}\bigg|_{s=0} = E(X^n)$$

Where $i = \sqrt{-1}$

Okay 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 this pdf of the random variable about the origin.

That we have seen when we are discussing that representation of this area that is analogy with this moment of the area pdf.

(Refer Slide Time: 39:53)

Then the dispersion asymmetry and peakedness of the pdf of a random variable described in terms of their second moment, third moment and fourth moment of the pdf of the random variable about it is mean this is important. So that when we are taking this, when we are want to know that is dispersion, asymmetry and peakedness, we are taking that moment with respect to it is mean.

And last, we saw that this 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.

Now what just now we are discussing that we will know see how what are the different techniques that this sample data and this sample data how we can represent in terms in graphically, you can represent that sample data. So for convenient representation of data a dataset may be divided into halves or quarters when an ordered dataset is divided into halves the division point is called the median.

So ordered dataset here means that we 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 the median. As you have seen that it is that 50 percent, when the 50percent probability is covered they have called it is median. So and when an ordered dataset is divided into a quarter, that is four different quarters, and then the division points are called the sample quartiles. So these are some terms that will know in detail that is first is that our quartile.

(Refer Slide Time: 41:51)

Quartiles

☐ First Quartile (Q_1):It is a value of the dataset such that one-fourth of the observations are less than this value.

☐ Second Quartile (Q_2):It is a value of the dataset such that half of the observations are less than this value. It is equivalent to the median.

☐ Third Quartile (Q_3):It is a value of the dataset such that three-fourths of the observations are less than this value.

Probability Methods in Civil Engineering: M3L4   Dr. Rajib Maity, IIT Kharagpur   22

So this first quartiles is the Q_1 represent as Q_1, it is the value of the dataset such that one-fourth of the observations are less than this particular value. 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 that one-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 value of that dataset such that the half of the observations are less than this particular value. So this is known as this is known as this quartile which is also equivalent to the median, just what we discuss because it is the half of the data less than that particular value.

And third quartile is similar to the earlier, but this is that three-fourth that is the 75percent of the data when it is less than. So it is a value of the dataset such that three-fourth of the observations is less than this particular value. Now why this things is important, this quartile as important so that 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 some idea that how the distribution of this of the dataset over a particular range.

Whether if you see that there are lot of data is there within the first quartile or within the below 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 data set 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 the 50percent data is there.

So instead of that if we get a distributed fashion like this then it will be helpful to know it is dispersion and it will be more.

(Refer Slide Time: 44:20)

Quartiles

☐ First Quartile (Q$_1$):It is a value of the dataset such that one-fourth of the observations are less than this value.

☐ Second Quartile (Q$_2$):It is a value of the dataset such that half of the observations are less than this value. It is equivalent to the median.

☐ Third Quartile (Q$_3$):It is a value of the dataset such that three-fourths of the observations are less than this value.

Graphically it is done, in terms of the different box plot that will come in a minute before that we will see that in terms of the percentile.

(Refer Slide Time: 44:30)

Percentiles

☐ A $100p^{th}$ percentile value in an ordered dataset is such that $100p\%$ of the observations are equal to or less than this value. This implies that $100(1-p)$ % of the observations are greater than this value. Here p is the cumulative probability so that $0 \le p \le 1$.

☐ The first, second and third quartiles are equal to the 25th, 50th and 75th percentile values in an ordered dataset.
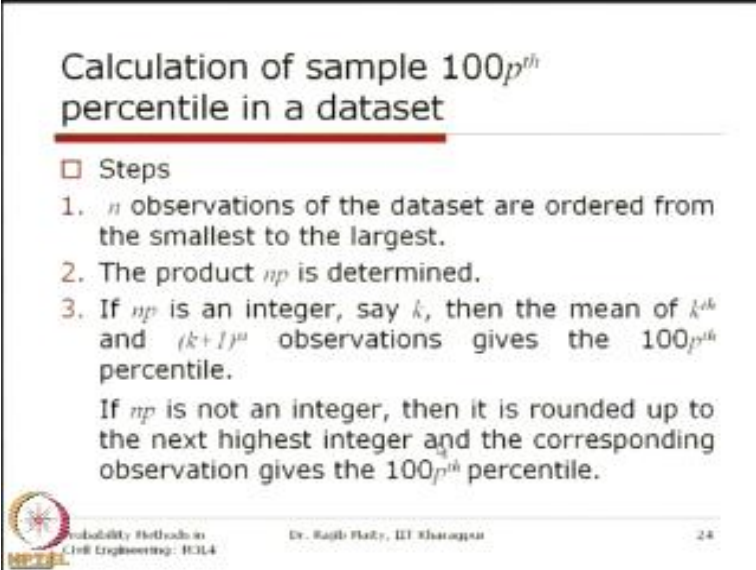
So if when we talk about the percentile of a particular dataset that is available to us, that a 100 multiplied by pth percentile. So, before I go this one this p is 0 is your cumulative probability. So here p, so if we see that p is the cumulative probability so that it varies from the 0 to 1.Now for any percentile value if we so for the p equals to 0.1 means, this quantity will give you the 10th percentile or p equals to 0.6will give you the 60th percentile value.

So in general the 100pth percentile value in an ordered dataset, ordered dataset means, you have just ordered 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 so percentile we are making it more general and define for each and every value 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 quartile with respect to the percentile, then it is very easy to compare that this first, second and the third quartile are equal to the 25th,50th and 75th, percentile values in an ordered dataset respectively.

So this first quartile is nothing but the 25th percentile, second quartile is nothing but the 50th percentile and third quartile is nothing but the 75th percentile. Now this kind of representation if we want to present in a graphical way, then we will see that.

(Refer Slide Time: 46:31)

Calculation of sample 100p^{th} percentile in a dataset

☐ Steps
1. n observations of the dataset are ordered from the smallest to the largest.
2. The product np is determined.
3. If np is an integer, say k, then the mean of kth and $(k+1)^{th}$ observations gives the 100p^{th} percentile.

   If np is not an integer, then it is rounded up to the next highest integer and the corresponding observation gives the 100p^{th} percentile.
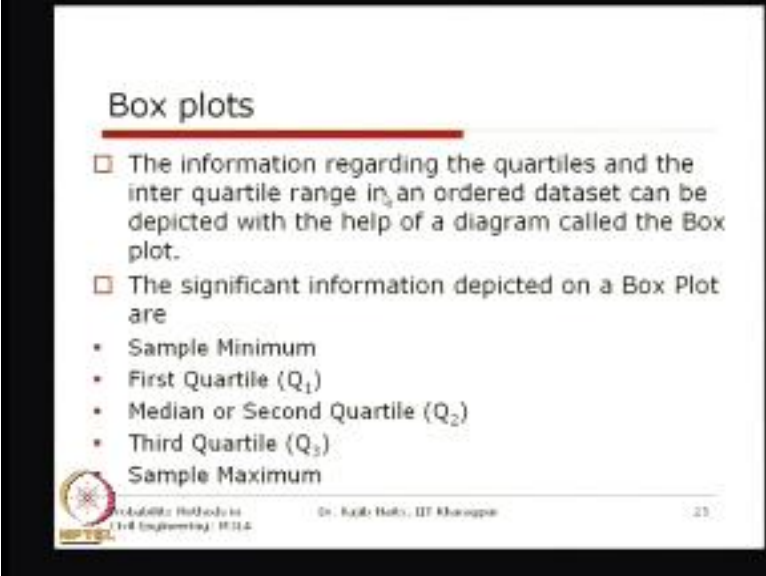
This can be done through a box plot now to prepare that box plot we have to know certain we have to know the percentiles, particularly for those at quartiles that is 25th 50th and 75th and sometimes the 5th and 95th as well. So, to get a particular percentile that is 100pthpercentile, how can we calculate the percentile? So these are the steps involved in that, so suppose that there is total dataset number 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 dataset of size n is arranged first.

The product of n p is determined so if p is known a particular desired percentile 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 become one integer or it can be and it may not be an integer.

So, if it is an integer say that integer is k, then the mean of the k th and k plus first observation. So, from k than immediate next observation this two observations are taken and their mean is calculated and that gives that 100pth percentile. If np is not an integer, then it is rounded up to the next highest integer and the corresponding observation gives that 100 pth percentile.

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 this two values that is kth and in between this k plus first observation. 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 nearest integer value and that particular value we will give the 100pth percentile. So, now in this way whatever the desired percentile is required that is known to us and.

(Refer Slide Time: 48:54)

We can represent that information in terms of the box plot. So, the information regarding the quartile and the inter quartile range in an ordered dataset can be depicted with respect ;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, these five points are essential before we prepare one box 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: 49:55)

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 these two quartiles is represented by a rectangle. So, and the median, so if I just one after another if I just draw it here,
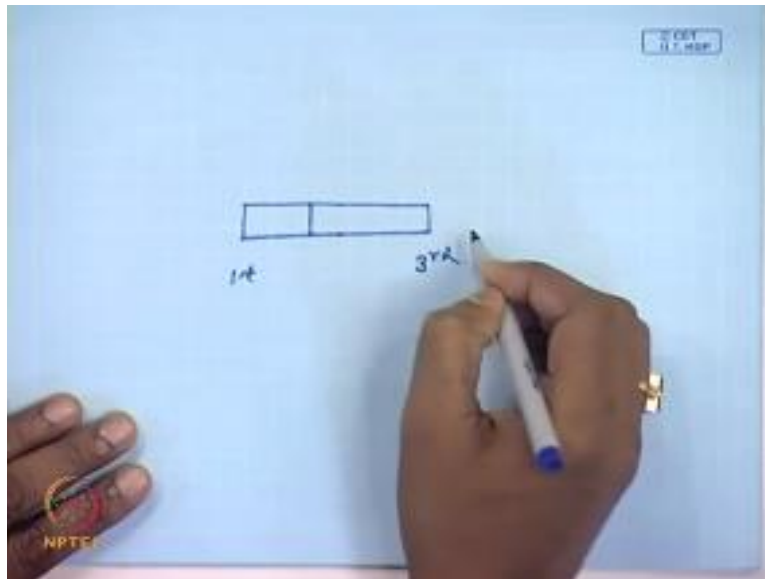
So this one is your, the 1st quartile and this one say is your 3rd quartile and this is represented in terms of a rectangle. Now, the second step is that the median value is represented by a line within this rectangle. Now, this median value, with this one is that it can be either, it is not necessary that we will be exactly middle of this rectangle; it can be any point either this side or towards that side it will be.

(Refer Slide Time: 51:01)

Characteristics of a Box Plot

❑ The range between the first and the third quartiles is represented by a rectangle.

❑ The median value is represented by a line within the rectangle.

❑ The ranges between the first quartile and the minimum value and also between the third quartile and the maximum value are connected by lines. For very large datasets, the 5th percentile and the 95th percentile values may be used in place of the sample minimum and sample maximum respectively.

Then, the range between the first quartile and the minimum value and also between the third quartile and the maximum value are connected by lines. So that the minimum suppose, there this is the minimum value and this the maximum value, then this two are connected by the lines. So this two are connected by these lines. Now, again for the very large dataset that 5thpercentile and 95th percentile values maybe used in the place of the sample minimum and the sample maximum respectively.

(Refer Slide Time: 51:40)
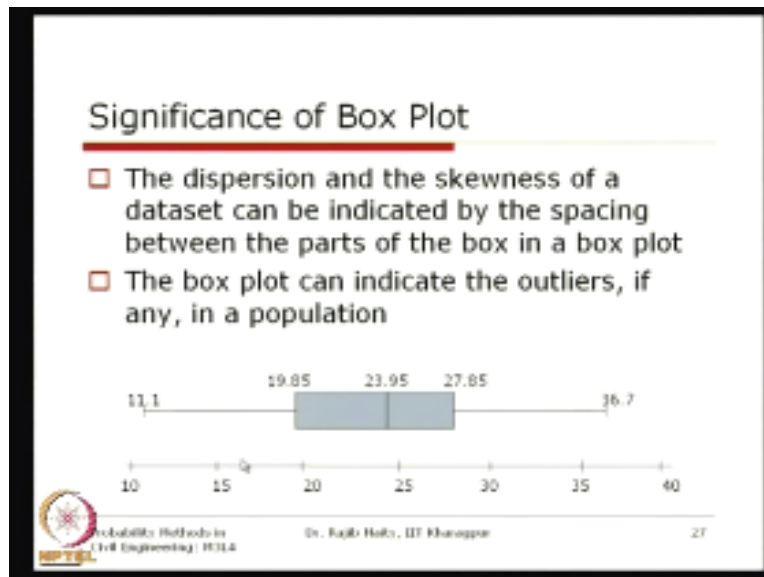
So, here is one representation is shown, the dispersion and the skewers of the dataset as, we just discuss is that can be indicated by the spacing between the parts of the box in a box plot. The box plot can indicate the outliers, if any, in the population. So, here is one example is given, so this is the real axis and where this box plot is shown. So, this is the 11.1 is the sample minimum or the 5thpercentile,19.85 is your first quartile,23.9 is median or second quartile,27.85 is the third quartile and 36.7 is the is 95thquartile or the sample maximum value.
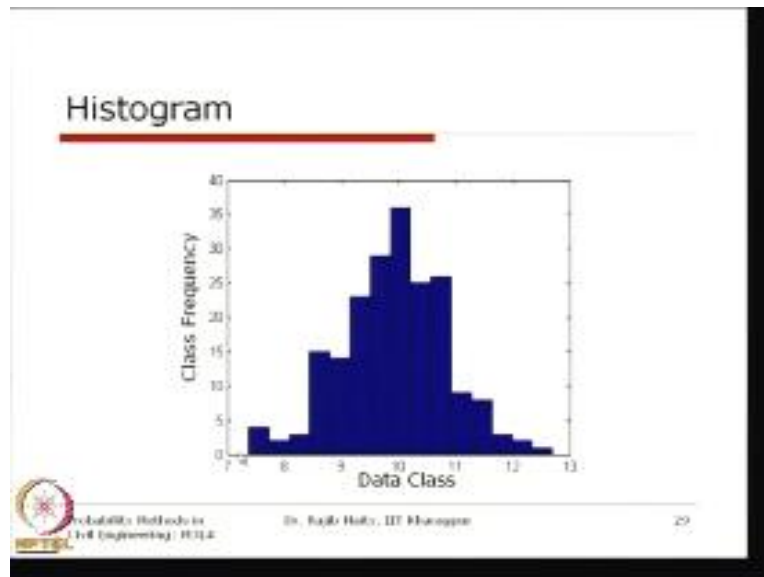
(Refer Slide Time: 52:25)

Histogram

- In a histogram, the area of the rectangle gives the probability and thus the height of the rectangle is proportional to the probability.

- The rectangles representing successive values of the RVs always touch those adjacent to them so that there are no gaps in between.

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 probability and 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 so that there are no gaps in between.

(Refer Slide Time: 52:51)

So, here is one example of this histogram is shown so that 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 the heights 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 desired distribution like this, we can have some idea how the distribution is looks like.

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 no meaning; no significance. The rectangles representing the successive values of the random variables do not touch their adjacent ones. So, obviously there is no meaning only the height of this bars are important and which is showing your probability.

(Refer Slide Time: 53:58)
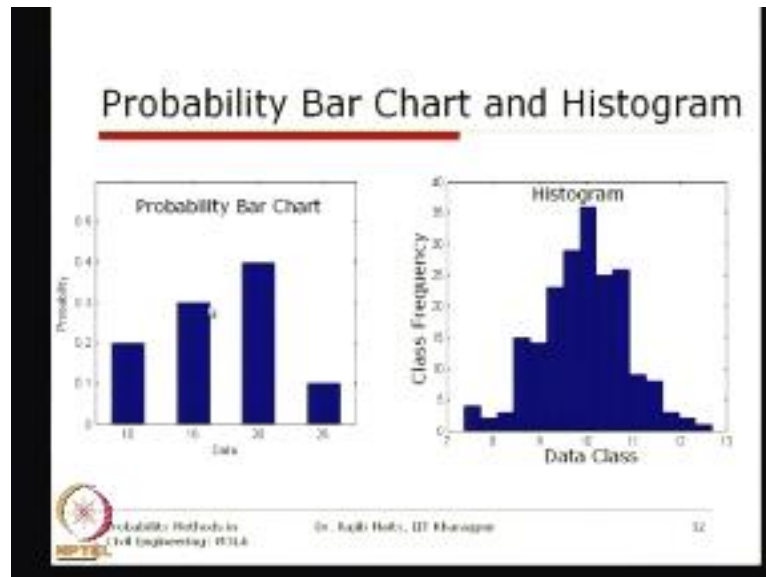
So, here is one comparison when we call that probability bar chart. That mean this, at this particular data point, this is the probability associated with this one is 2; this can also be scaled, so that this height of this rectangle can, we can that time will say that it is proportional to the probability.

So, these are the how we can represent. So, this is the histogram, how we can represent the sample data, just to get the idea how 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 more descriptors 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 distribution over the entire range of this sample data. So, we will see next lecture with the description of this discrete probability distribution. Thank you.

Probability Methods in Civil Engineering

End of Lecture 09

Next: "Discrete Probability Distribution"

In Lec 10

NPTEL Video Recording Team

NPTEL Web Editing Team

Technical Superintendents

Computer Technicians

A IIT Kharagpur Production

www.nptel.iitm.ac.in