

Probability and Statistics
Prof. Dr. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture No. #08
Probability Distributions

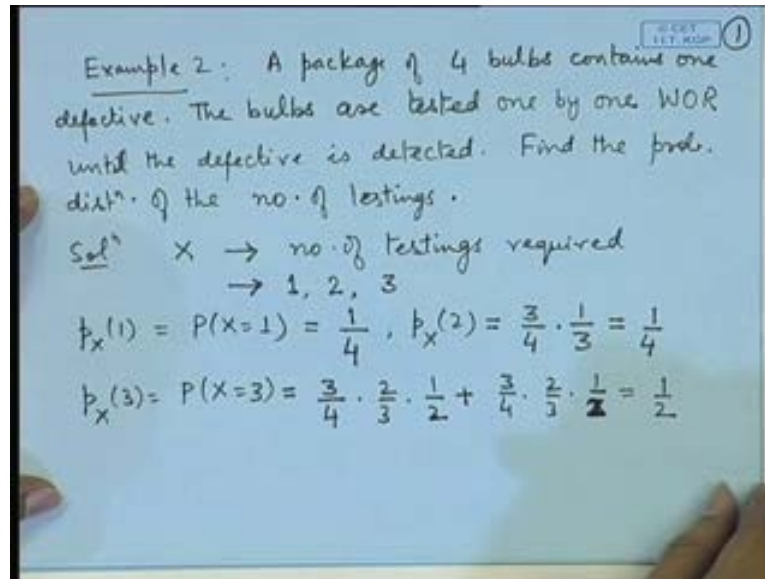
In the last lecture, I have introduced the concept of random variables, so the motivation for defining random variables is that, most of the random experiments and their outcomes, if we look at the complete description of that, it may be quite complex; whereas, we may not be interested in the full description of the phenomena, we may be interested in a particular numerical phenomena connected with the random experiment. Therefore, we define random variable as a real valued function, defined on the sample space.

Now, once we have a probability space ω b P and if we are defining a random variable x on ω ; then naturally the probabilities which are associated with the point in the sample space get translated to the values taken by the random variable. This correspondence between the probability allotment with the values of the random variable is called probability distribution.

In the previous lecture, I introduced the concept of a cumulative distribution function. We showed that it is having a one to one correspondence with the probability distribution of a random variable; and that the same time it gives almost all the information about the random variable.

Now, random variables are of different types, we have the cases where the random variable can take a finite number of values or countably infinite number of values or uncountably infinite values, which we say that lying in an interval. So, the first one, we called a discrete random variable. The allotment of the probabilities for a discrete random variable, we call a probability mass function. So, let me give one more example of a probability mass function.

(Refer Slide Time: 02:16)



Consider the following. A package of 4 bulbs contains one defective; the bulbs are tested one by one without replacement, until the defective is detected. Find the probability distribution of the number of testings. So, let us look at, x is the random variable which denote the numbers of testings. Once we outline a random variable, first thing is we should look at, the set of values that the random variable can take. Here, if we test it once, that itself may be a defective or if that is not so, second one may be defective or if that is not so, the third one may be defective or if that is not so, the fourth one will may be defective.

However, we should look at here that, if the third one is found to be not defective, since, there are only 4 bulbs in the packet. So, the fourth one will be automatically treated as the defective. So, we in reality do not need the 4 testing we need only 1 2 or 3 testings, so what is the probability of x is equal to 1.

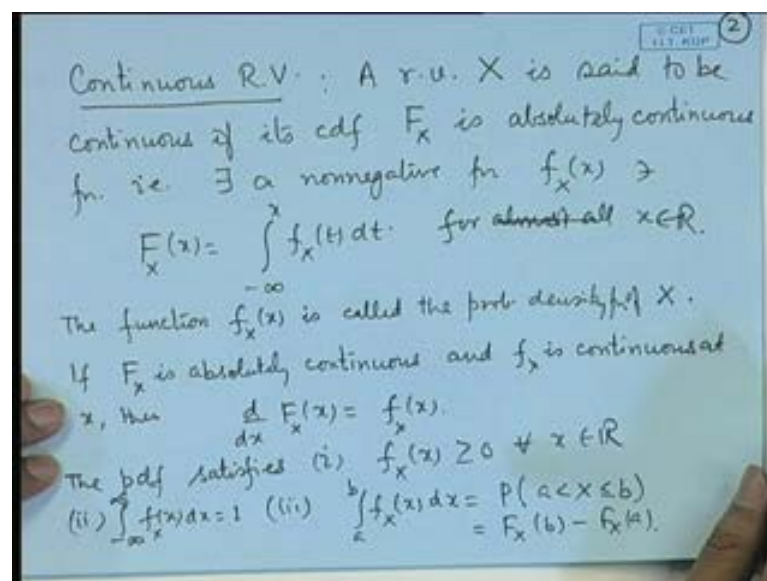
Now, there are 4 bulbs, out of which one is defective, so the probability that the first testing gives you a defective will be 1 by 4. What is the probability, that the second one is defective so that means, the first one must not be defective; now there are only 3 bulbs left out of that 1 is defective, so the probability of that is 3 by 4 into 1 by 3. that is equal to 1 by 4.

If we look at, $P(x=3)$, that is probability of x equal to 3, now by the direct logic, it should be simply 1 minus a probability of x equal to 1; and probability x equal to 2, so it should be half. Let us, see how it can be evaluated directly also.

So, the first one is not defective, second one is not defective and the last one is defective but, this is equal to 1 by 4. We should also take into account the possibilities that, the third one is non defective, because either it is defective or not defective in both the cases, the outcome is known that the fourth one is will be defective or not, so it is actually equal to half. So, this is the probability distribution of the random variable x , which is the number of testings required.

Next let us, denote, define the probability distribution of a continuous random variable. So, what is a continuous random variable, so we say that a random variable x is continuous, if its cdf is an absolutely continuous function.

(Refer Slide Time: 06:22)



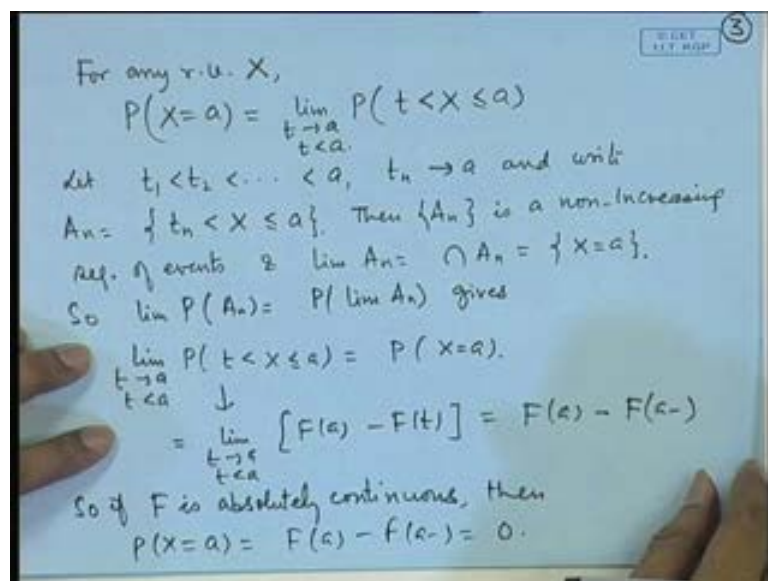
A random variable x is said to be continuous, if its cumulative distribution function F , is absolutely continuous function, that is there exist a nonnegative function, a small f such that, capital F is equal to minus infinity to x $f(t) dt$, for almost all x , basically it should be for all x , belonging to \mathbb{R} . The function f is called the probability density of density function of x or Pdf.

If f is absolutely continuous and f is continuous at x , then d by d x of F x is equal to small f x . The P d f or the probability density function satisfies one, that it is the nonnegative function, secondly, the integral over the full region is equal to 1; and if we integrate the density over an interval a to b , it is denoting the probability of a less than or equal to x less than or equal to b ; that is basically f x b minus f x a .

Now, the continuous random variable is quite different from the discrete random variable, so the first thing that, we noticed here in the discrete random variable, we had a probability mass function. So, P x I it denoted, the probability that the random variable x takes values x I . Here, at the point x we have a density function f x , now this is one should not get mislead and say that, this is probability that capital X equal to small x .

In fact, in the discrete case, the mass function defines the probability of that point in the continuous case, it does not define. In fact, probability of every point is 0. So, that is a important observation to prove that, let us see, firstly that for any random variable x , probability that x is equal to a , is equal to limit, probability t less than X less than or equal to a , t tends to a , where t is less than a . So, if we denote and let us prove this statement, let us consider say t_1 less than t_2 and so on, t_n goes to a and write A_n to be the set, that t_n is less than X less than or equal to a . Then this is a non increasing and limit of A_n will be equal to intersection of A_n , which is basically equal to X is equal to a .

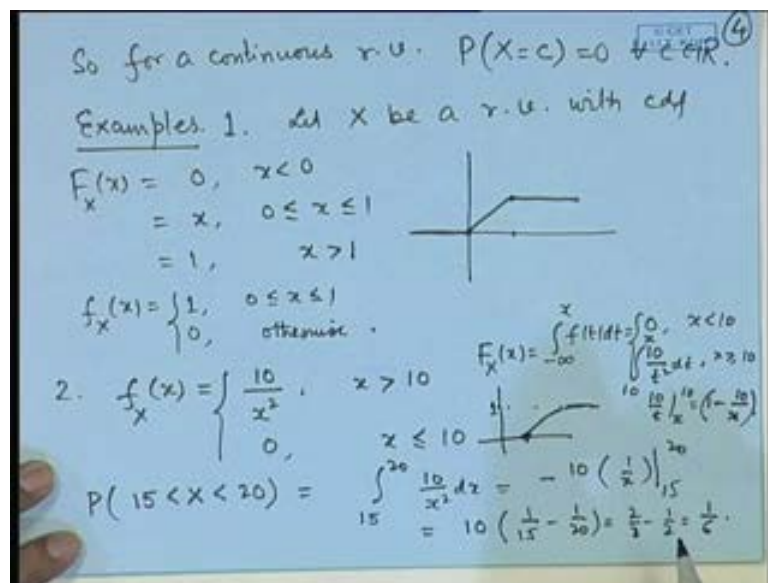
(Refer Slide Time: 10:12)



So, limit of probability of A_n is equal to probability of limit A_n , this statement is equivalent to, limit t tending to a , t less than a , probability t less than X less than or equal to a is equal to probability, X is equal to a . Now, this is actually equal to limit $F(a) - F(t)$, where t tends to a , t is less than a , which is exactly equal to $F(a) - \lim_{t \rightarrow a^-} F(t)$, so if F is absolutely continuous function.

Then probability X equal to a , will be equal to $F(a) - \lim_{t \rightarrow a^-} F(t) = 0$, so for a continuous random variable, probability that X is equal to some value C is 0 for all C .

(Refer Slide Time: 12:46)



(No audio from 12:58 to 13:18)

Let us consider some examples here, let X be a random variable with cdf, F_X is equal to 0 for x less than 0, it is equal to x for 0 less than or equal to x less than or equal to 1, it is 1 for x greater than 1, quite obviously, this is an absolutely continuous function and in fact, you can plot it, for x less than 0 it is 0, for 0 to 1 it is x and thereafter it is 1. So, the density function is given by 1 for 0 less than or equal to x less than or equal to 1 and it is 0 for x lying outside the interval 0 to 1.

Let us consider some more examples, suppose, f_X is given by $10/x^2$ for x greater than 10 and 0 for x less than or equal to 10. I want to find out, what is the probability 15 is less than x less than 20, then it will be equal to integral from 15 to 20,

10 by x square d x, that is equal to minus 10, 1 by x from 15 to 20, that is equal to 10, 1 by 15 minus 1 by 20, so 2 by 3 minus half that is equal to 1 by 6.

If you want to find out the c d f here, then it is minus infinity to x f t d t, now here the positive value of the density starts from the 0.10, therefore, it will be 0 for x less than 10; and thereafter it will become 10 by, integral of 10 by t square; so, that will give me from 10 to x for x greater than or equal to 10. So, it is equal to 10 by t from x to 10 that is equal to 1 minus 10 by x; this is for x greater than or equal to 10. So, if you want, you can plot this function, up to x equal to 10 it is 0; and at x equal to, after that it becomes 1 minus 10 by x, so the function is increasing, because 10 by x is decreasing function as x tends to infinity, this goes to 0; that means, it goes to 1. In fact, if you look at the derivative of this, so it becomes like this basically a concave function, this is suppose 1 here.

(Refer Slide Time: 17:21)

3. $f(x) = \begin{cases} x/2, & 0 \leq x \leq 1 \\ 1/2, & 1 < x \leq 2 \\ (3-x)/2, & 2 < x \leq 3 \\ 0, & \text{otherwise} \end{cases}$ $P\left(\frac{1}{2} < X < \frac{5}{2}\right) = \int_{1/2}^1 \frac{x}{2} dx + \int_1^2 \frac{1}{2} dx + \int_2^{5/2} \frac{3-x}{2} dx$

$\int_{-\infty}^{\infty} f(x) dx = \int_0^1 \frac{x}{2} dx + \int_1^2 \frac{1}{2} dx + \int_2^3 \frac{3-x}{2} dx$

$= \frac{1}{2} \left[\frac{1}{2} + 1 + 3 - \frac{5}{2} \right] = 1.$

$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & x < 0 \\ \int_0^x \frac{t}{2} dt, & 0 \leq x < 1 \\ \int_0^1 \frac{t}{2} dt + \int_1^x \frac{1}{2} dt, & 1 \leq x < 2 \\ \int_0^1 \frac{t}{2} dt + \int_1^2 \frac{1}{2} dt + \int_2^x \frac{3-t}{2} dt, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$

Let me take 1 more application of this suppose, x is a continuous random variable with the density function, say x by 2 for 0 less than or equal to x less than or equal to 1, it is half for 1 less than x less than or equal to 2, it is 3 minus x by 2 for 2 less than x less than or equal to 3, it is 0 at all other points. First thing is we should notice whether, it is a proper probability distribution, then the integral of f x d x over the full range must be equal to 1; obviously, it is nonnegative function. So, this will be equal to integral 0 to 1, x by 2 d x plus integral 1 to 2, 1 by 2 d x plus integral 2 by 2 to 3, 3 minus x by 2 d x.

So, if we integrate each of these terms, it becomes in fact, half can be kept outside, then we will get half plus 1 plus here $3 - x^2$ by 2; so, it becomes $9 - 4$, that is 5 by 2, so clearly this is equal to 1.

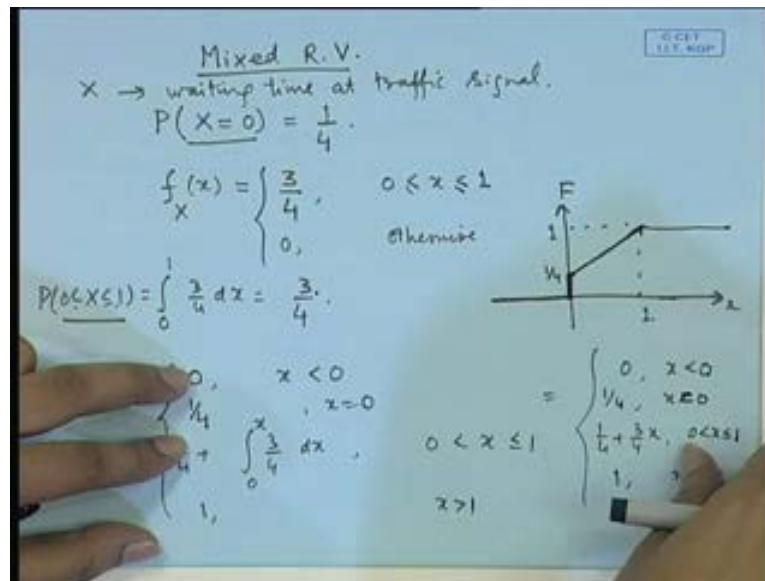
If we want to write down the cumulative distribution function of this, then minus infinity to x $f(t) dt$, so for x less than 0, we are integrating only 0, so this will be equal to 0 in the other regions for 0 to 1, we have to integrate the term t by 2 dt from 0 to x .

If we look at x lying between, so this is for x less than 0, for x between 0 to 1, we have this term for x between 1 and 2, we have 0 to 1, x by 2 or t by 2 dt plus 1 to 2 half dt that is 1 to x , it will be 0 to 1 t by 2 dt plus 1 to 2 half dt plus $3 - t$ by 2 dt from 2 to x ; and it is equal to 1 for x greater than or equal to 3, so if we simplify these terms, we will get here, let us look at this one, it is equal to x^2 by 4, here up to 1 it has been integrated therefore, this should be 1 by 4, plus $x - 1$ by 2; once again here if you had integrated up to 2, then this value would have become 3 by 4, plus the value of this term, that is $3 - t$ whole square by 4, from x to 2, which will be some value.

It is 1 for x greater than or equal to 3, 0 for x less than 0, so here you can see, this function will be having different values x^2 , which is a convex function up to 1, then thereafter it say line and thereafter again it say, parabola kind of function, which is actually in the reverse, so it will go up to 1.

However, in each case if you differentiate it, you will get the density function, if you want to find out probability of a certain set; suppose, I say what is the probability that, x lie between half to 5 by 2, then it will be integral of x by 2, from half to 1 plus integral of half from 1 to 2 plus integral of $3 - x$ by 2 from 2 to 5 by 2, which can be evaluated after thirteenth calculations.

(Refer Slide Time: 22:37)



There is also a possibility, that random variable may be partly discrete or partly continuous, these are known as mixed random variables, let me consider some example, consider for example, a person travelling to his office every day by car; on the way to the office there is a traffic crossing, it may happen that some of the days, that traffic crossing has a green signal will he approaches the traffic light; and he crosses without waiting, on the other days, there is a red light and he has to wait for a certain amount of period.

In other words, he may have if I say x denotes the waiting time at traffic signal, it may happen that the probability that X equal to 0 is say 1 by 4; that means, 25 percent of the time, he is able to go without waiting; on the other hand, if he has to wait, then it is a continuous distribution, he may have to wait for 0 to 1 minute. If we see, the total allotment of the probabilities, in this case the probability of 0 less than X less than or equal to 1, is actually equal to 3 by 4.

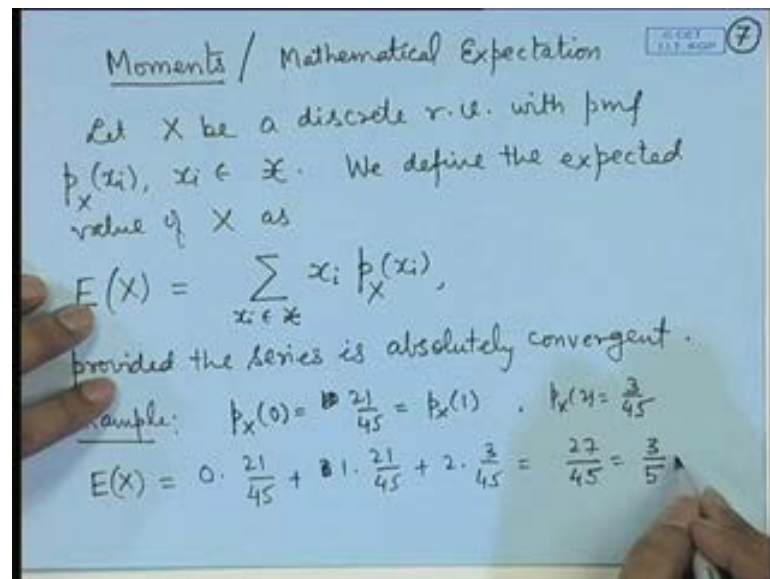
So, probability x equal to 0 and probability of x lying between 0 1 is 3 by 4, so if you add up these 2 it is actually becoming equal to 1 but, here the distribution is not for totally discrete or totally continuous random variable. We can also look at the c d f of this particular random variable, at before x equal to 0 it is 0 at x equal to 0 it is becoming 1 by 4 and there after it is becoming 1 by 4 plus integral 3 by 4 dx from 0 to x .

And of course, it is equal to 1 for x greater than 1, so if we expand it, it is becoming 0 x less than 0 1 by 4 for x less equal to 0; and here it is becoming 3 1 by 4 plus 3 by 4 x for

0 less than x less than or equal to 1, this 1 for x greater than or equal to 1. You notice here that the function is not continuous at x equal to 0 and therefore, in fact, there is a jump point, the right hand limit and the value at x equal to 0 is 1 by 4 whereas, the left hand limit at 0 is 0. So, there is a jump of size 1 by 4 at 0. So, that shows that discrete nature of the random variable at x equal to 0 whereas, after 0, 0 to 1 it is a continuous random variable. So, this an example of a mixed random variable.

If we plot this particular distribution c d f, then up to 0 it is 0, at 0 it is becoming 1 by 4, then from 1 by 4 to 1, it is increasing and it is 1, so in this case it is a continuous random variable, there is a jump here of size 1 by 4.

(Refer Slide Time: 27:09)



Next, we can see that, the concept of moments or mathematical expectation. Let x be a discrete random variable with probability mass function say $P x \ x \ i$ for $x \ i$ belonging to some set x ; so, we define the expected value of x as E of x , this is defined as sigma $x \ i$ into $P x \ x \ i$, where $x \ i$ belongs to x , provided the series is absolutely convergent.

Let us consider some examples, let us take the example of defective computers purchase; that means, a store had certain numbers of computers and a person purchase 2 computers, so in that example we had probability of x equal to 0 was 10 by 21 by 45 and that was same as probability of x equal to 1; and probability x equal to 2 was 3 by 45, let us look at expectation x . So, that will be equal into 0 into 21 by 45 plus 1 into 21 by 45

plus 2 into 3 by 45. So, that is equal to 27 by 45 or 3 by 5; that means, on the average his purchase may include less than 1 defective computer.

(Refer Slide Time: 29:56)

2. $P_X(1) = \frac{1}{4} = P_X(2), P_X(3) = \frac{1}{2}$

$E(X) = 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{2} = \frac{9}{4} > 2$

$P(X = (-1)^{j+1} \cdot \frac{3^j}{j!}) = \frac{2}{3^j}, j=1, 2, \dots$

$\sum_{j=1}^{\infty} |x_j| P(X = x_j) = \sum_{j=1}^{\infty} \frac{3^j}{j} \cdot \frac{2}{3^j}$ is divergent

So $E(X)$ does not exist.

If X is continuous r.v. with pdf $f_X(x)$, then

$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$, provided the integral is absolutely convergent.

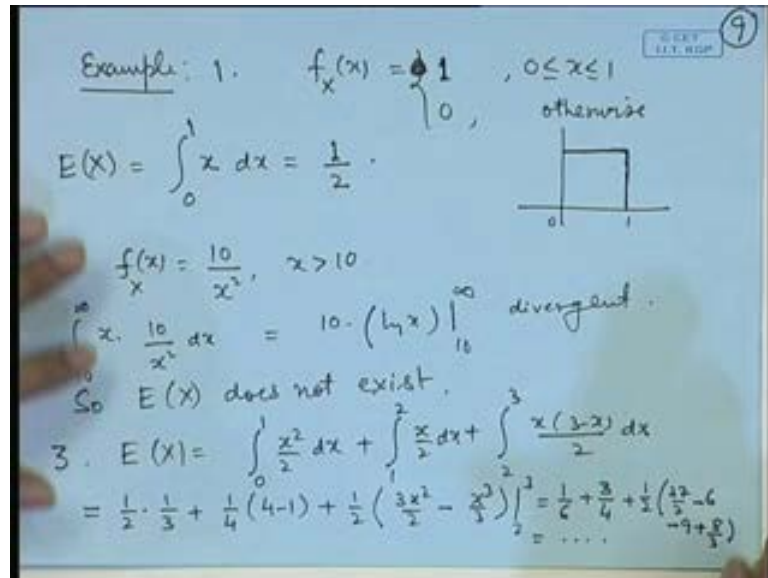
Let us look at the next problem of, let us $P X 1$ was 1 by 4, $P X 2$ was 1 by 4 and $P X 3$ was half, that is a number of trials needed for getting the defective bulb, here expectation of x is equal to 1 into 1 by 4 plus 2 into 1 by 4 plus 3 into 1 by 2, which is equal to 9 by 4, which is actually greater than 2; that means, on the average you will need more than two testings for getting the defective bulb. You notice here that, we have not checked the condition of absolute convergence here, because that number of terms is only finite, the condition is needed actually in order that, the expectation is well defined. Consider say probability that, x is equal to minus 1 to the power, say j plus 1 3 to the power j by j factorial, say j is equal to 2 by 3 to the power j , j is equal to 1 2 and so on.

Let us look at sigma modulus of x_j , probability x is equal to x_j , j is equal to 1 to infinity then, this equal to sigma 3 to the power j by j factorial by j^2 divided by 3 to the power j , now, this is divergent, so expectation x does not exist. Although one may write here, sigma x_j probability x equal to x_j and here, you will get minus 1 to the power j plus 1 2 by j , which is having value $\log 2$ but, it is not absolutely convergent therefore, expectation x does not exist here.

So, this expectation x is also called average value, it is also called mean of X or arithmetic mean of X etcetera. So, several names are there it is also called the first

moment about the origin. If X is a continuous random variable with P d f, $f(x)$ then expectation X is defined as integral minus infinity to infinity $X f(x) dx$, provided the integral is absolutely.

(Refer Slide Time: 33:34)



Let us look at some of the examples the first example here we had

Let us look at some of the examples, the first example here, we had in the continuous case $f(X)$ was 0 for, it was 1 for $0 \leq x \leq 1$ and it is 0 otherwise, so expectation X will become equal to integral x into $f(X)$, that is equal to 1 in this interval and therefore, you will have basically only this term, which is equal to half, which looks reasonable also because, if you plot this distribution from the 0 to 1 interval, it is actually constant value and therefore, the mean value must be middle value, that is half.

let us look at the next example, $f(x)$ is equal to $10/x^2$ for $x > 10$; and skipping the other part here 0, so integral $x \cdot 10/x^2 dx$ for 10 to infinity. So, that is equal to 10 and integral of $1/x$ is $\ln x$ for 10 to infinity, this is divergent, so here expectation x does not exist.

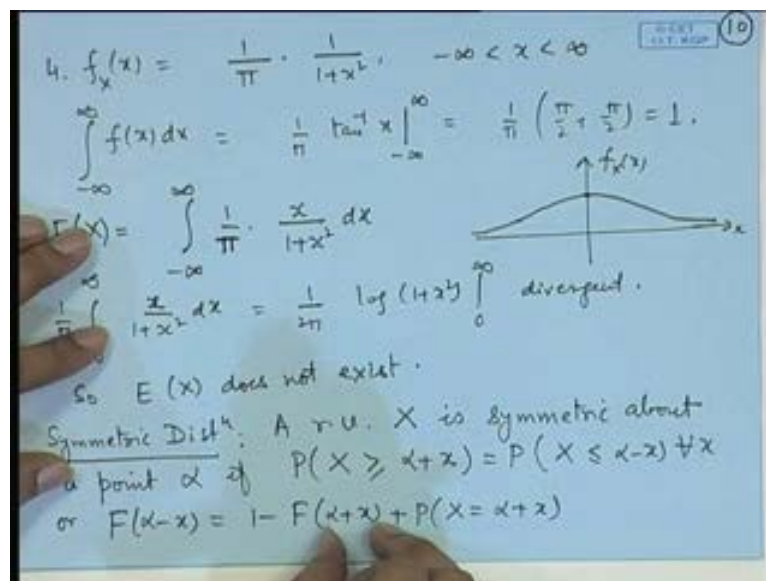
Let us consider another example, $f(x)$ is equal to $x/2$ for $0 \leq x \leq 1$, it is half for $1 \leq x \leq 2$ and $3-x/2$ for $2 \leq x \leq 3$, for this example expectation X will become equal to integral x

square by 2 d x plus integral x by 2 d x for 1 to 2, plus integral 2 to 3 x into 3 minus x by 2, in the remaining portion f x was 0, so there will be no term here.

So, this is equal to half 1 by 3 plus here the integral will become x square by 4, so 1 by 4 4 minus 1, plus 1 by 2 integral of 3 x is 3 x square by 2 minus x square will give you x cube by 3 for 2 to 3, so it is equal to 1 by 6 plus 3 by 4 plus half.

The value what this is 27 by 2 minus 6 minus 9 plus 8 by 3, which can be actually simplify. There may be some other case also where it looks, that the expectation will exist as we have seen here, in the second example the integral itself is divergent and therefore, we are directly concluding that the expectation does not exist; however, there may be a case where it looks that the integral will exist but, actually it does not exist.

(Refer Slide Time: 37:29)



Let us consider the density 1 by pie into 1 by 1 plus x square, so it is a valid probability density as can be seen easily the integral will be tan inverse x, that is equal to 1 by pie **pie** by 2 plus pie by 2, that is equal to 1.

So, it is a valid probability density; however, if I look at expectation of X, a common mistake here is that, here one can think that it is an odd function and it is over a symmetric region, so it should be 0; however, the property of the even function or odd function is applicable, when the integral is convergent, here if we look at integral 0 to

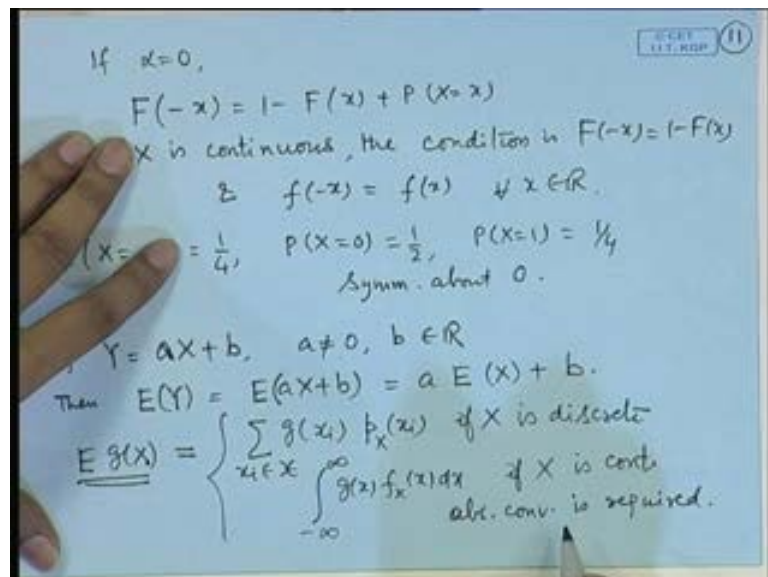
infinity, $1 + x + x^2$, then it is equal to $\log(1 + x^2)$, which is actually divergent.

Therefore expectation x does not exist, if we look at the shape of this curve, actually at x equal to 0, it is $1 + \pi$ and if we look at the shape of this curve, then here it is at X equal to 10 it is 1; and thereafter it is reducing from various curves here for example, we found the expectation to be the middle value here the expectation does not exist, here it is somewhat different.

So, we are tempted to consider something like a concept of symmetry. We can define symmetric distribution as a random variable X is symmetric about a point α , if probability of X greater than or equal to $\alpha + x$ is equal to probability of x less than or equal to $\alpha - x$, for all x .

In other words we can say $F(\alpha - x) = 1 - F(\alpha + x) + P(X = \alpha + x)$ probability of X is equal to $\alpha + x$, if the random variable is continuous, this term will vanish and it will be simply relationship between in the c d f at minus and plus point, suppose, I take α is equal to 0.

(Refer Slide Time: 41:07)



In this definition if we take α is equal to 0, then this condition is reducing to simply f of minus x is equal to $1 - F$ of x plus probability x equal to x , if X is continuous the condition is F of minus x is equal to $1 - F$ of x ; and if it is density, so you will have,

so if you see this distribution, it is symmetric about 0 here the distribution is symmetric about half, etcetera.

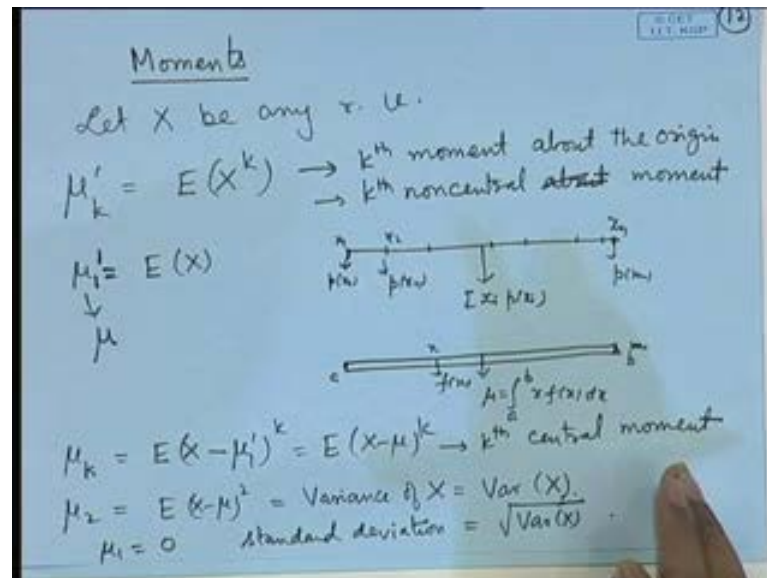
You may also define a discrete symmetric distribution probability X equal to say, minus 1 is $\frac{1}{4}$ probability X equal to 0 is half probability X equal to 1 is $\frac{1}{4}$, so this is symmetric about 0.

We notice here some properties about the expectation function for example, if I define y is equal to $aX + b$, where a is any non 0, real number and b is any real number; then if we consider expectation of y , then it is expectation of $aX + b$, it will become equal to expectation a times expectation of X plus b .

In general we can define expectation of a function of random variable also, for example, if I have expectation of $g(X)$, suppose X is a discrete random variable, then we can consider it as, if X is discrete, we can define it as $\sum g(x) f(x)$, if x is continuous. The definition is subject to the condition that, this summation or this integral must be absolutely convergent, so absolute convergence is required, in order that this expectation of $g(X)$ is well defined, another question arises at this point that, how do we define the expectation of a mixed random variable; so, for a mixed random variable the expectation would be simply, the value multiplied by the probability plus the integral of the density multiplied by the value, so that means, in the discrete and continuous case, we separately evaluated.

Let us look at example of this means random variable, so here expectation X will become equal to $\frac{1}{4} + \int_0^1 x \cdot \frac{3}{4} dx$, so this is equal to $\frac{3}{8}$, which is actually less than half, this could have been the mean, if the random variable was completely defined as 1 for 0 to 1, the probability density function. However, here the $\frac{1}{4}$ probability is taken over by the point x equal to 0 therefore, the average value has average waiting time has reduced and it is now $\frac{3}{8}$, it is not half.

(Refer Slide Time: 45:45)



Next, we discuss the concept of moments of distributions, so let x be any random variable; we define μ_k' is equal to expectation of X to the power k , this is called k th moment about the origin or k th non central moment about, who cares k is non central moment.

Observe here that, if I put k is equal to 1, then μ_1' is nothing but, expectation of X , that is the expected value, so here we are getting **the...**

If I am considering x equal to k equal to 2 k equal to 3 etcetera, so basically we are able to define higher order non central moments. So, if you look at expectation of x in what way it is a moment, so if I consider say a weightless bar with say, weights attached here, at the point say x_1 we attach weight $P \times 1$ at x_2 point we attach the weight $P \times 2$ at the point say x_n we attach the weight $P \times n$.

Suppose, it is attached to 2 hinges with 0 friction and if we consider, the balance point at the point of equilibrium or the center of gravity; that will be $\sum x_i P \times i$, that is the moment of the first moment of this.

In a similar way, if we consider a metallic bar with the density say, $f(x)$ at the point x . Suppose, this is point a , this is point b , then if we consider a to be x $f(x) dx$, then this is denoting the balance point of this are the center of gravity of this bar; so, that is why it is μ_1' is actually the first moment about the origin of the random variable.

We also define μ_k , that is equal to expectation of $x - \mu$ to the power k ; or expectation of $x - \mu$ to the power k , the first one, we can usually denote by μ , the arithmetic mean, this is called k th central moment. In particular μ_2 is called variance of the distribution.

Let us look at the significance of this μ and μ_2 in particular, so μ as, I mentioned, it is denoting the major of central tendency or the centre of gravity or the point of equilibrium of the distribution. We may also be interested in knowing, how the values of the random variable are varying with respect to its mean; to get it we may look at the values of $x_i - \mu$, now if you take the average value of $x_i - \mu$ or $x - \mu$; then expectation of $x - \mu$ is $E(x - \mu)$ which is actually 0. So, this does not give you any information this is basically, because of the fact that the plus deviations and the minus deviations from the mean, they cancel out each other.

So, μ is actually 0; however, to get a better measure of variability we may look at the squared differences, so we look at $(x_i - \mu)^2$ and then, we take the average. So, that is known as the variance of the distribution, we also define a quantity called a standard deviation that is equal to square root of variance of x .

So, this gives a measure of the variability of the distribution of the random variable. It is obvious that, if we are considering k to be positive integral values, then there will be a relation between μ_k and μ_k' , which is expressed as follows, so if we consider say μ_k is equal to expectation of $x - \mu$ to the power k .

(Refer Slide Time: 51:07)

$$\mu_k = E(X - \mu)^k$$

$$= E\left(X^k - \binom{k}{1} X^{k-1} \mu + \binom{k}{2} X^{k-2} \mu^2 - \dots + (-1)^k \mu^k\right)$$

So, using the binomial expansion, this becomes x to the power k minus $\binom{k}{1} x$ to the power $k-1$ plus $\binom{k}{2} x$ to the power $k-2$ minus \dots plus $(-1)^k \mu^k$. So, the first term is $\binom{k}{1} x^{k-1} \mu$, so $\binom{k}{1} x^{k-1} \mu$ and you will have $(-1)^k \mu^k$ to the power k ; so, if you look at this, it becomes μ^k minus $\binom{k}{1} \mu^{k-1} \mu$ plus $\binom{k}{2} \mu^{k-2} \mu^2$ minus \dots plus $(-1)^k \mu^k$.

In particular, we can write say μ_2' is equal to μ^2 minus 2μ plus μ^2 , which is equal to μ^2 minus 2μ plus μ^2 or expectation X^2 minus expectation X whole square. We may also observe one thing here, that since μ^2 is greater than or equal to 0, this implies that expectation of x^2 is always greater than or equal to expectation x whole square.

We also have a relationship between non central moments and central moments in the reverse direction; that means, we may interpret μ_k' as expectation of $(x - \mu)^k$, here we consider this as one term and this as another term. So, this becomes expansion of $\binom{k}{1} x^{k-1} \mu$ minus $\binom{k}{2} x^{k-2} \mu^2$ plus \dots plus $(-1)^k \mu^k$; in particular, μ_2' is equal to μ^2 plus 2μ minus μ^2 . So, 2μ and μ^2 minus μ^2 , now this term is actually 0, so this means that μ_2' is equal to 2μ plus μ^2 minus μ^2 .

We also define absolute moment, so k th absolute moment of X is defined as expectation of modulus X to the power k ; we also consider say, β_k is equal to expectation of modulus $X - \mu$ to the power k , we also define what is known as some factorial moments expectation of $X(X-1)\dots(X-k+1)$. This is called k th factorial moment of X , this is k th absolute moment about origin and this is k th absolute central moment.

In all the definitions of the moments, the basic thing is that these expectations must exist for example, expectation of modulus X to the power k must exist, expectation of modulus $X - \mu$ to the power k must exist; that means, the corresponding integrals are the summations must be absolutely convergent; in some cases a lower order moment may exist, a higher order moment may not exist.

let us take 1 example, if I consider $f(x)$ is equal to say $2/x^3$ for x greater than or equal to 1 and 0 for X less than 1, if we consider expectation of X , then it is equal to $\int_1^{\infty} dx$, that is equal to 2, if we consider expectation of X^2 , then that is equal to $\int_1^{\infty} 2x^2/x^3 dx$, clearly this is divergent. So, a lower order moment may exist but, a higher order moment may not exist.

In the next classes we will define some further characteristics of distributions, basically this moments or a other characteristics actually, like mean or variants, they explain the nature of the random variables value, how the random is taking values over the range of the values with what probabilities, so we will look into these thing in the next lecture thank you.