PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 03

Lecture - 14

Numerical Examples

So, suppose let f(x) be a function defined by:

$f(x) = c * x$, for $-1 < x < 2$;

$f(x) = 0$, otherwise.

Note that earlier we did x^2, and now we are doing x only. We need to check whether f can be a probability density function (pdf) of a continuous random variable. So, if f can be a pdf, we denote it in short notation as pdf. We will find c and the probability that $x \leq 0$. If it cannot be a pdf, we will justify our answer.

First of all, we will check some of the properties. The most important property we check is that:

$\int_{-\infty}^{+\infty} f(x)\, dx = 1$.

Note that earlier we broke the whole interval, and most of the time, it is 0 outside the range of this random variable. Therefore, we will only integrate where it is non-zero. Thus, we skip all those steps and state that:

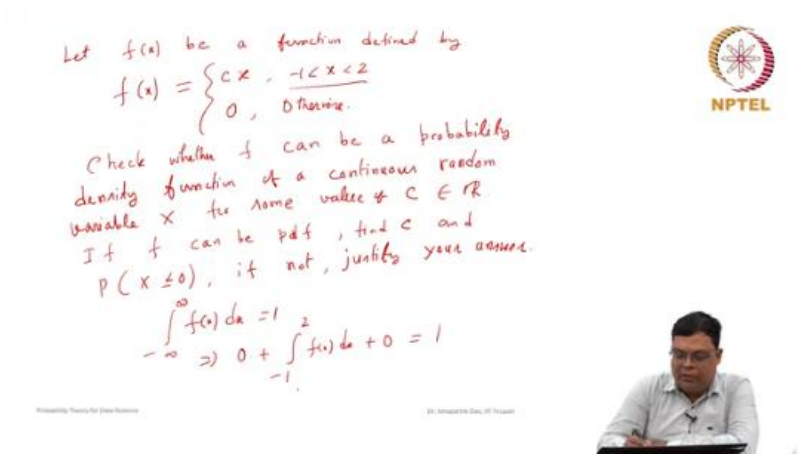
$\int_{-1}^{2} f(x)\, dx + 0 = 1$.

So, from here we can find the solution for c. This is:

$\int_{-1}^{2} c * x\, dx = 1$.

This implies that:

$c * \int_{-1}^{2} x\, dx = 1$.



The integral of x is:

$\int x \, dx = x^2 / 2$.

Evaluating this from -1 to 2 gives us:

$c / 2 * (2^2 - (-1)^2) = c / 2 * (4 - 1)$.

This simplifies to:

$c / 2 * 3 = 1$.

Therefore, we can solve for c, giving us:

$c = 2 / 3$.

This suggests that the function may be a probability density function because when $c = 2/3$, it could satisfy the properties of a probability density function.

Now, we want to find the probability that $x \leq 0$. By definition, the probability that $x \leq 0$ is:

$P(x \leq 0) = \int$ from -1 to 0 of f(x) dx.

We found that $f(x) = (2/3) x$ when $-1 \leq x \leq 2$, and $f(x) = 0$ otherwise. This probability only contributes where the density exists, so we consider the integral from -1 to 0. The integral of $(2/3) x$ from -1 to 0 is:

$(2/3) \int$ from -1 to 0 of x dx,

which results in:

$(2/3) * (x^2 / 2)$ evaluated from -1 to 0.

Calculating this, we have:

$(2/3) * (0^2 / 2 - (-1)^2 / 2) = (2/3) * (0 - 1/2) = (2/3) * (-1/2) = -1/3$.

However, a probability cannot be negative, indicating that we made a mistake in our calculations. The error arises from the assumption that this function can be a probability density function. We need to check the properties of a probability density function more thoroughly. While we checked that:

$\int$ from $-\infty$ to ∞ of f(x) dx = 1,

we overlooked another important property: the function f(x) must always be greater than or equal to 0.

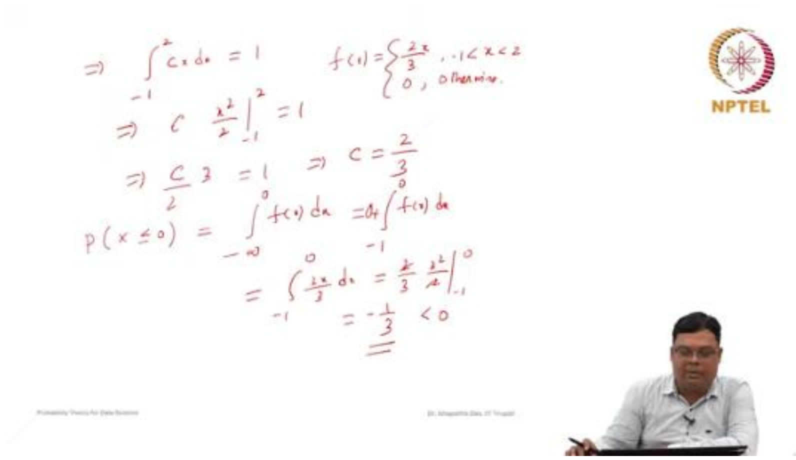
Given that f(x) = c * x, if c is positive, in the interval from -1 to 0, the function takes on negative values. If c is negative, then in the interval from 0 to 2, the function will also take on negative values. Therefore, there cannot exist any value of c that satisfies the condition $f(x) \geq 0$ throughout the interval. Hence, f(x) cannot be a probability density function of the random variable x because it does not satisfy the properties required for a probability density function.

This illustrates that all the properties are very important, and we must check them thoroughly.

Whenever a probability density function is given to us, we have to check whether it is correct or not. Let's do another numerical example. Suppose a random variable x has a probability density function defined as follows:

$f(x) = 1 / 4$, for $-2 < x < 2$; $f(x) = 0$, otherwise.

This is called a uniform random variable because it takes the same probability density value for all values within the interval. If you draw the graph of this probability density function, you will see that it is constant at 1 / 4 between -2 and 2.



So, the graph of this function f(x) shows that it is 0 outside this region, making it a continuous uniform random variable. For a discrete uniform random variable, it would take the same probability for most points. Here, for the continuous case, it takes the same probability density values for all values inside the interval.

Now, we will discuss some questions regarding this density function. The first question is to find the probability that $x < 1$.

The second question is to find the probability that $|x| > 1$. The third question is to find the probability that $2x + 3 > 5$.

These are not very difficult questions. If you have understood the concept, you can follow along as I compute how to find these probabilities using the given probability density function.

To find the probability that $x < 1$, we define it as:

$P(x < 1) = \int$ from $-\infty$ to 1 of f(x) dx.

However, since this density function only exists in the interval from -2 to 2, we can adjust the limits of integration. This means we will integrate from -2 to 1. The integral then becomes:

$\int$ from -2 to 1 of f(x) dx = $\int$ from -2 to 1 of (1/4) dx.

This is a constant, so we can factor it out:

(1/4) * ∫ from -2 to 1 of dx.

The integral of dx from -2 to 1 is simply the length of the interval, which is:

1 - (-2) = 3.

So, the probability that $x < 1$ is:

(1/4) * 3 = 3/4.

Now, regarding the second question: to find the probability that $|x| > 1$, we consider the intervals where this condition holds, which is when $x < -1$ or $x > 1$. Therefore, we need to calculate the probability for these two intervals. The probability of $x < -1$ is:

$P(x < -1)$ = ∫ from -2 to -1 of f(x) dx,

and the probability of $x > 1$ is:

$P(x > 1)$ = ∫ from 1 to 2 of f(x) dx.

Both of these integrals will equal:

(1/4) * (length of the interval).

Calculating the first integral gives us:

(1/4) * (-1 - (-2)) = (1/4) * 1 = 1/4.

For the second integral, we have:

(1/4) * (2 - 1) = (1/4) * 1 = 1/4.

Adding these two probabilities together gives us:

1/4 + 1/4 = 1/2.

Finally, for the third question, to find the probability that $2x + 3 > 5$, we rearrange the inequality:

$2x + 3 > 5$

$2x > 2$

$x > 1$.

To find this probability, we integrate from 1 to 2 of f(x) dx. Thus, the probability is:

$P(x > 1) = \int$ from 1 to 2 of (1/4) dx.

This is:

$(1/4) * (2 - 1) = (1/4) * 1 = 1/4$.

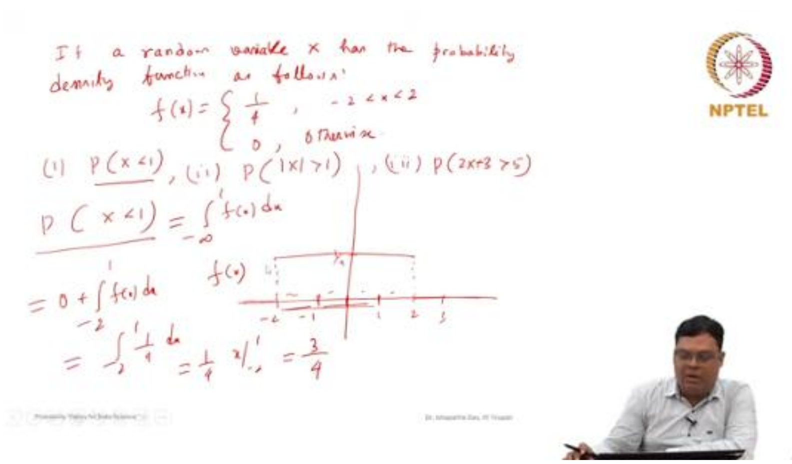
In summary, the probabilities we found are:

$P(x < 1) = 3/4$,

$P(|x| > 1) = 1/2$,

$P(2x + 3 > 5) = 1/4$.

Now, how will we find the probability of $|x| > 1$? Suppose you want to find the probability that $|x| > 1$. You can compute this directly through integration or take the complement.



The complement will be $P(|x| > 1) = 1 - P(|x| \leq 1)$. Why do we do that? Because this is $1 - P(|x| \leq 1)$. First, we find this probability. The probability that $|x| \leq 1$ means that $-1 \leq x \leq 1$. So, $|x|$ will be less than or equal to 1. It's easy to find this probability because the density function is $f(x) = 1/4$ over the interval from -2 to 2. So, we will integrate from -1 to 1. The integral from -1 to 1 of f(x) is:

$\int$ from -1 to 1 of 1/4 dx

Since this density function is 1/4, the integration becomes 1/4 multiplied by the length of the interval from -1 to 1. The length of the interval from -1 to 1 is 1 - (-1) = 2, so the result is:

$(1/4) * 2 = 1/2$

Now we want to find the value of the probability that $|x| > 1$. This is given by:

$P(|x| > 1) = 1 - P(|x| \leq 1) = 1 - 1/2 = 1/2$

Next, we want to find the probability that $2x + 3 > 5$. To find this probability, we set up the inequality:

$2x + 3 > 5$

Subtracting 3 from both sides gives:

$2x > 2$

Dividing both sides by 2 yields:

$x > 1$

To find this probability, we can compute it directly or use the complement. The direct computation gives us the probability that $x > 1$, which is equal to:

∫ from 1 to 2 of f(x) dx

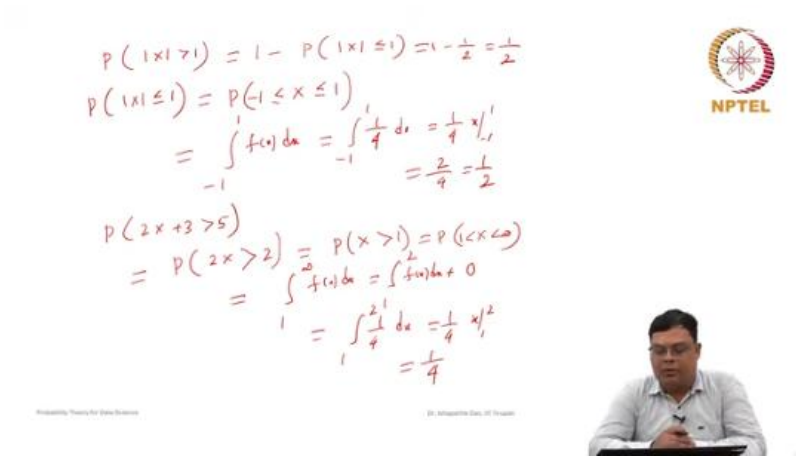
Since $f(x) = 1/4$, we integrate:

∫ from 1 to 2 of 1/4 dx

This is:

$(1/4) * (2 - 1) = (1/4) * 1 = 1/4$

Thus, the probability is 1/4. So, this is an example of a probability density function. Suppose you have been given this probability density function, and you want to find the cumulative distribution function, which we can denote as F(x). How will you find the cumulative distribution function?



The probability density function is defined as 1/4 whenever $-2 < x < 2$, and it is equal to 0 otherwise. To find the cumulative distribution function, we define it as the probability that $X \leq x$ for any real number x. If the real number x is between -2 and 2, then the density exists; otherwise, it is 0. This means that the cumulative distribution function will be:

$F(x) = \int_{-\infty}^{\infty} f(t)\, dt$

We are changing the variable to t because we have already used x for the cumulative distribution function. By integrating, we compute the area under the curve f(t). Now, let's see how this density function is defined. It is 1/4 in the interval from -2 to 2, and 0 elsewhere. If we take any x in this interval, the area under the curve will be computed. If $x < -2$, then the probability will be 0 because there is no density function present.

Thus, $F(x) = 0$ if $x < -2$. When $-2 \leq x < 2$, we will integrate from −2 to x. Therefore, F(x) will be:

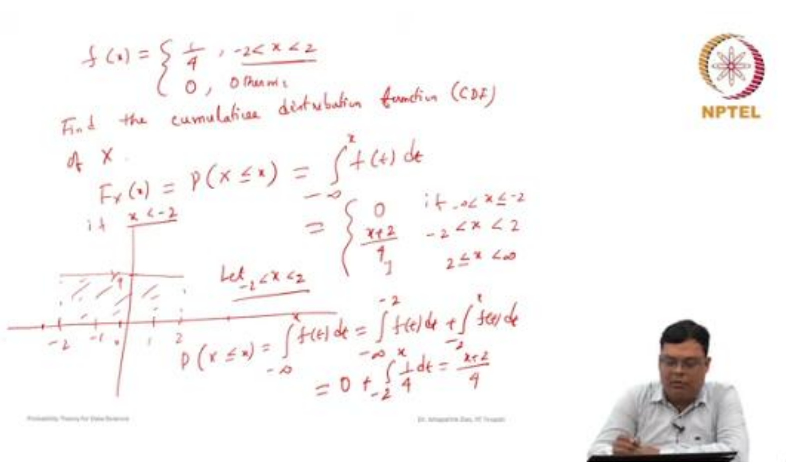
$F(x) = \int_{-\infty}^{x} f(t)\, dt$

We can break this down: the integral from $-\infty$ to −2 is 0, and the integral from −2 to x of f(t) dt. The integral from −2 to x of the density function, which is 1/4, gives us:

$F(x) = (1/4)(x + 2)$

This represents the cumulative distribution function for $-2 \leq x < 2$. For $x \geq 2$, the cumulative distribution function will equal 1. Since the total area under the probability density function is 1, for any value of $x \geq 2$, $F(x) = 1$.

So, again, if you want to see this, whenever $-2 \leq x < \infty$, we have $f(x)$, which is the probability that $X \leq x$. Then what will happen?



This is nothing but the integral from $-\infty$ to x of $f(t)$ dt. Since $x \geq 2$, we need to break this interval into three parts: the integral from $-\infty$ to -2 of $f(t)$ dt, plus the integral from -2 to 2 of $f(t)$ dt, plus the integral from 2 to x of $f(t)$ dt. The first part, from $-\infty$ to -2, will be 0 because the probability density function is 0 in that interval. The second part, from -2 to 2, has a density function value of 1/4. The last part, from 2 to x, is again 0 since the density function is 0 in that range.

So, when you integrate from -2 to 2, you get $t/4$. Applying the limits gives us:

$(2/4) - (-2/4) = 4/4 = 1$.

This means the cumulative distribution function $F(x) = 1$ when $x \geq 2$. To see how this cumulative distribution function looks, it's 0 up to -2. At $x = -2$, the value is still 0, but as x approaches 2, it increases in a straight line to 1. Thus, for the cumulative distribution function, when $x = -2$, it is 0; when $x = 2$, it is 1, and it stays at 1 after that.

This is an example of how to find the cumulative distribution function when given a probability density function. The first part, from $-\infty$ to -2, will be 0 because the probability density function is 0 in that interval. The second part, from -2 to 2, has a density function value of 1/4. The last part, from 2 to x, is again 0 since the density function is 0 in that range. So, when you integrate from -2 to 2, you get $t/4$. Applying the limits gives us:

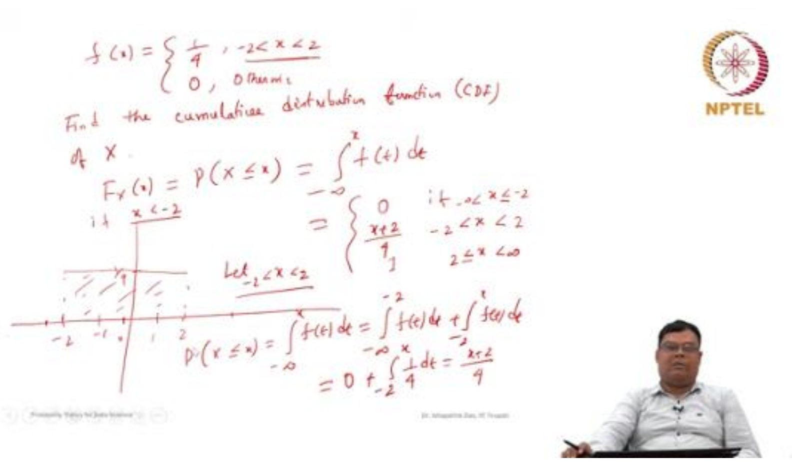
$(2/4) - (-2/4) = 4/4 = 1$.

This means the cumulative distribution function $F(x) = 1$ when $x \geq 2$. To see how this cumulative distribution function looks, it's 0 up to −2. At $x = -2$, the value is still 0, but as x approaches 2, it increases in a straight line to 1. Thus, for the cumulative distribution function, when $x = -2$, it is 0; when $x = 2$, it is 1, and it stays at 1 after that. This is an example of how to find the cumulative distribution function when given a probability density function.

We have a probability density function of a random variable, and we want to determine how to find this probability density function. Most continuous random variables can take values within a finite interval, but they can also take values within an infinite interval, which we will discuss. As we mentioned in the discrete case, random variables can take both finite and infinite values, and we will consider an example here. Let $f(x)$ be a continuous random variable with the probability density function defined as:

$f(x) = c/x^2$ for $x > 1$ and 0 otherwise.

We consider $x > 1$ because including 0 would lead to an undefined situation.



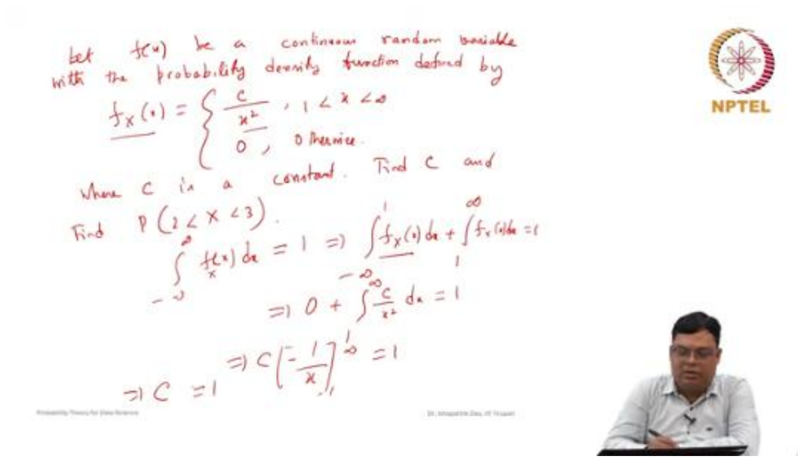
Here, c is a constant, and we need to find the value of c and the probability that x is between 2 and 3. To find c, we need to check some important properties of the probability density function. First, the function $f(x)$ must always be greater than or equal to 0. If c is positive, then $f(x)$ will always be non-negative. The second property we need to check is that the integral of $f(x)$ from $-\infty$ to $+\infty$ must equal 1 for it to be a valid probability density function.

Since f(x) is 0 for $x \leq 1$, we only need to consider the integral from 1 to ∞. The integral of f(x) from 1 to ∞ is given by the integral of c/x^2. We compute this integral by integrating from 1 to some value $a > 1$. This integral is c times the integral of $1/x^2$ from 1 to a. We evaluate the integral, which gives us c times the limit as $a \to \infty$.

As $a \to \infty$, the value approaches c, which must equal 1. Thus, we find that $c = 1$, and therefore the probability density function becomes $f(x) = 1/x^2$ for $x > 1$.

Now, to find the probability that x is between 2 and 3, we integrate f(x) from 2 to 3. This gives us the probability that x lies within this range, and we evaluate the integral to get the final probability value. This process allows us to determine both the value of c and the probability for the specified range.

The constant we found is equal to 1. So, next, what we will discuss is the probability between 2 and 3. We know that the probability density function, f(x), is equal to $1/x^2$ whenever $x > 1$ and equal to 0 otherwise. Now that we know the probability density function, we can compute the probability. The probability that x is greater than 2 and less than 3 will be the integral from 2 to 3 of f(x) dx, which is equal to the integral from 2 to 3 of $1/x^2$ dx.



Evaluating this gives us $-1/x$ from 2 to 3, which results in $1/2 - 1/3$. So, the final probability is 1/6. I hope you have understood and are following the lectures. We have already completed the discussion on the probability mass function, covering important examples and how to find the probability mass function from the probability distribution function when x is a discrete random variable. We also learned how to find the probability distribution function from the probability mass function, along with some numerical examples.

Next, we completed the topic on the probability density function for continuous random variables. We discussed how to define the probability density function and how to compute the cumulative distribution function. We reviewed the properties of the probability density function and worked through numerical examples to find the cumulative distribution function and compute probabilities using the probability density function. So, we've covered the classification of random variables into discrete and continuous types. Next, we will discuss some other important measures known as moments.