PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 03

Lecture - 12

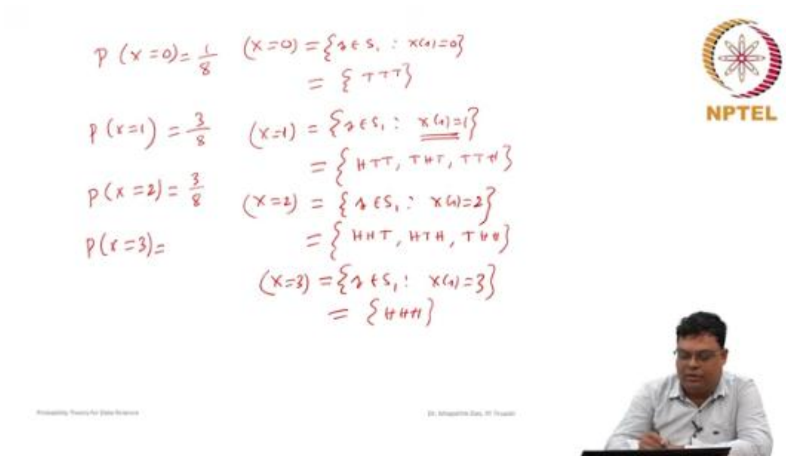
Probability Mass Function

Now, we are interested in finding the probability at a specific point rather than the cumulative probability for values like $x \leq 1$ or $x \leq 2$. So, what will be the probability that f(x) when x = 0? This is the set of all outcomes in the sample space S1 where X(s) = 0. Remember, our sample space contains 8 points, and only one outcome satisfies X = 0. Thus, the probability of X = 0 is 1/8.

Now, what is the probability of X = 1? This is the set of all outcomes in S1 where X(s) = 1. Here, we are considering only the values that satisfy this exact equality, not an inequality. The outcomes that satisfy this are head-tail-tail, tail-head-tail, and tail-tail-head. Since there are three such outcomes out of 8, the probability of X = 1 is 3/8.

Next, let's find the probability of X = 2. This corresponds to outcomes where X(s) = 2, meaning exactly two heads. The outcomes here are head-tail-head, tail-head-head, and head-head-tail. Again, we have three outcomes, so the probability of X = 2 is 3/8.

Finally, we find the probability of X = 3. This is the set of all outcomes in S1 where X(s) = 3, meaning exactly three heads. Only one outcome satisfies this, which is head-head-head. Therefore, the probability of X = 3 is 1/8. So, this is 1/8. You can see that this is exactly the same as the probability of X = 0, the probability of X = 1, the probability of X = 2, and the probability of X = 3.



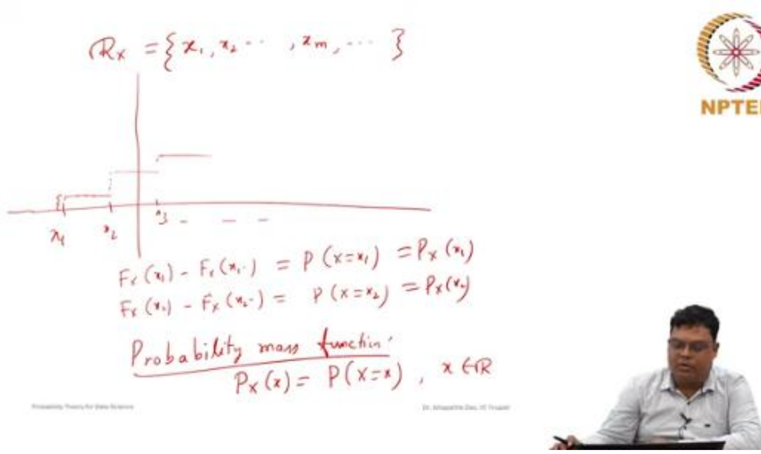
Basically, this jump actually represents the probability at a point. So, this is the probability of $X = 1$, the probability of $X = 2$, and the probability of $X = 3$. So, whenever you write this for discrete random variables, it has either a finite range of X containing only a finite number or it may be countably infinite. If it is a countably finite number, then the range might be something like 1, 2, 3, or any number up to m. If it is countably infinite, then there is a bijective correspondence to a subset of the natural numbers.

So, you can write it as $X_1, X_2, \ldots, X_m$. Because there is an order, we can assume that $X_1 < X_2 < X_3 < \ldots$. Without loss of generality, we can assume the range looks like this. Then, the distribution function for discrete random variables looks like a step function. For example, if we consider points like $x_1, x_2, \ldots$, the distribution function is 0 for values less than x_1.

At x_1, there is a step up, and it remains constant until it reaches x_2, and so forth for x_3. The size of each step corresponds to the probability of the random variable taking those specific values. For instance, the step from x_1 to x_2 reflects the probability that $X = x_1$. Similarly, the step from x_2 to x_3 represents the probability that $X = x_2$. There is a relationship between these probabilities at specific points and the cumulative distribution function.

Sometimes, for convenience, we use the probability at a point instead of the cumulative probability. This is why we define a new function called the probability mass function for discrete random variables. The probability mass function, denoted as $P(X)$, is defined by $P(X) = P(X = x)$. This applies for all $x \in \mathbb{R}$, but usually, this probability will be 0 whenever x is not in the range. There are properties of the probability mass function that we will learn about, and from this relationship, we can find the probability mass function for specific values like $P(x_1)$ or $P(x_2)$.

So, now one thing to consider is that whenever you define R_x as $x_1, x_2, ...,$ f(x) for x_1^- will essentially be 0 because we are considering it without loss of generality, with $x < x_2$. Therefore, f(x) for x_1^- - f(x) for $x_1^-{}_{-1}$ = f(x) for x_1, which is the same as the probability that $X = x_1$. Now, what is $P(X = x_2)$ - f(x) for x_2^-? This is essentially $P(X \leq x_2) - P(X \leq x_1)$. So, this can be represented as $P(X = x_2) - P(X \leq x_1)$.



From the numerical example here, f(x) for 0^- is 0, which is why this is coming out as 1/8; 0^- is 0. Now, f(x) for 1^- is 1/8, which means that f(x) for 0 is also 1/8. For the discrete random variable case, f(x) for x_k^- is equal to f(x) for $x_k^-{}_{-1}$ whenever $k > 1$. Otherwise, when $k = 1$, f(x) for x_1^- is equal to 0. This is why the probability mass function satisfies the equation: $P(x_k)$ = f(x) for x_k^- - f(x) for $x_k^-{}_{-1}$ when $k > 1$. Otherwise, when $k = 1$, it is equal to f(x) for x_1. Essentially, it takes the cumulative sum. If you know the cumulative distribution function, you can find the probability mass function using this method.

Now, how will we find the cumulative distribution function if the probability mass function, P(x), is known to us? The cumulative distribution function f(x) is the probability that $X \leq x$. We can find the cumulative distribution function using the probability mass function.

If you represent this graphically, it becomes clearer. For example, let x_1, x_2, x_3, and x_4 represent discrete values. The probabilities, $P_x(x_1)$, $P_x(x_2)$, and so forth, can be plotted, with values of zero for other values. For the cumulative distribution function, it starts at 0, then takes a jump at each x value. We have already discussed that this value corresponds to $P_x(x_1)$, $P_x(x_2)$, and so on.

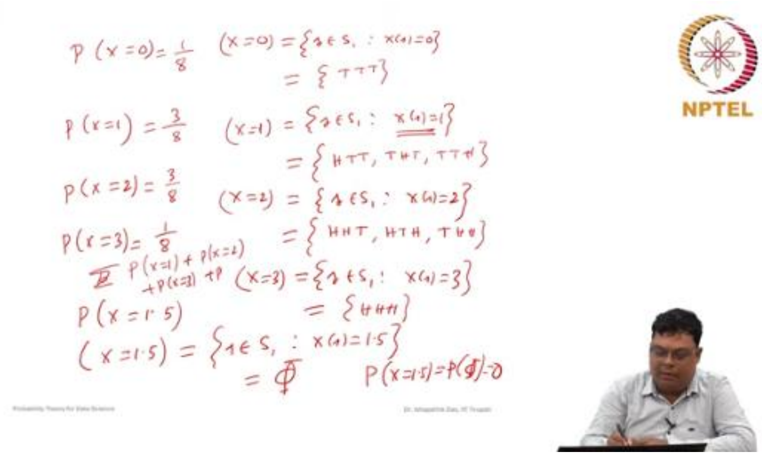
Therefore, $P_x(x_1) = f(x)$ for x_1, while $P_x(x_2) = f(x)$ for x_2 - f(x) for $x_2{}^{-}{}_{-1}$. Let's do a very simple numerical example to clarify this process. Let X be a discrete random variable with the range $R_x = \{-1, 0, 1\}$. The probability mass function, denoted as pmf, is defined as $P_x(x_k) = 1/3$ whenever $x_k \in \{-1, 0, 1\}$; otherwise, it is equal to 0. This establishes some probability distribution. Here, we did not specify any sample space; we are assuming there is some sample space, and X is the random variable taking these values.

Now, we have to check whether this can serve as a valid probability mass function. To do that, we need to know the properties of the probability mass function, particularly that the sum of the probabilities must equal 1, since the probability mass function represents $P_x(x)$ as the probability that $X = x$ for each value of x. Now, you can see from the previous example that we have verified that this will always hold true because it is a probability. So, P(x) will always be ≤ 1 and ≥ 0. Also, P(x) will be 0 if P(x) does not equal x_k for some values of k, such as k = 1, 2, and so on, because the range of X is fixed with values like x_1, x_2, etc.

If X does not belong to this set—meaning it is none of the values x_1, x_2, x_3—then, for instance, in tossing a coin three times, we see that X takes the values 0, 1, 2, and 3. If we consider what happens when X equals a value outside of these, like 1.5, we analyze the set of all s satisfying S_1 such that X(s) = 1.5. None of the outcomes correspond to this because X only takes integer values: 0, 1, 2, or 3, representing the number of heads. Thus, it cannot be equal to 1.5 for any of the points, which leads to the conclusion that it is a null set. Consequently, the probability of X = 1.5 will equal the probability of the null set, which is 0.

So, for any X that does not equal any of the values x_1, x_2, or x_k, the probability will also equal 0. This is an important property to know. Additionally, the sum of $P_x(x)$ for $X = x_k$ will equal 1. If you take this sum, you can see in this numerical example that we have the total values 0, 1, 2, and 3. If you calculate the sum: 1/8 + 3/8 + 3/8 + 1/8, it equals 1.

This can be broken down: 1/8 + 3/8 = 4/8; adding another 3/8 gives 7/8; finally, adding 1/8 totals 1. Thus, the summation of $P_x(x)$ for X = 1 + the probability of X = 2 + the probability of X = 3 + the probability of X = 0 should equal 1. This is because it is a certain event. If you take the summation of the probability of X, it equals the summation of all x_k.



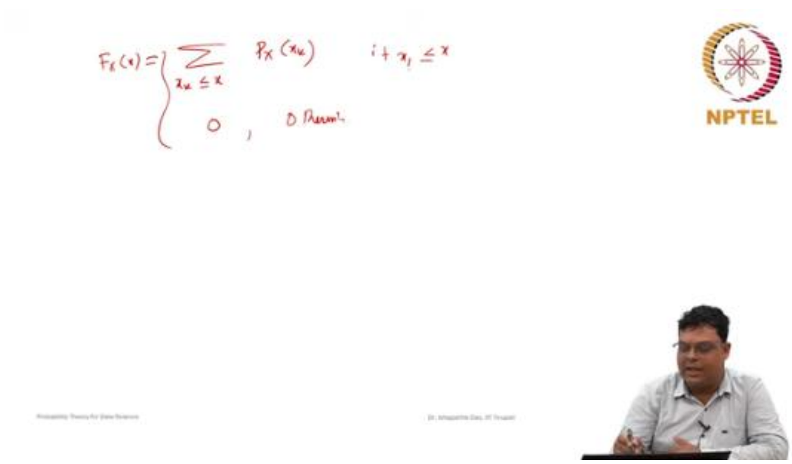
This means we are considering all the possible values that x can take, referring to the probability of a certain event, which is equal to 1. Now, how can we find the distribution function f_x from the probability mass function? Whenever we want to find the distribution function from the probability mass function, we know that if $x \leq x_k$, then the probability of $x = x_k$ is given by $f(x_k) - f(x_{k-1})$. If $k > 1$, we use this formula; otherwise, $P(x = x_1) = f(x_1)$. When you know the cumulative distribution function, you can find the probability mass function.

But suppose you know the probability mass function; then how can you find the cumulative distribution function? Suppose $P_x(x_k)$ is known for any $x_k \in \{x_1, x_2, ...\}$, which forms the range of x. The cumulative distribution function f(x) is defined as the probability that $x \leq x$. Now, if you look at this graph with values x_1, x_2, x_3, etc., x can be anywhere. If x is the minimum value and is less than this minimum value, then the probability will be 0.

For example, if $x < x_1$, from this numerical example, we see that the cumulative distribution function will equal 0 whenever $x < x_1$. This corresponds to a null set; when you consider that $x < x_1$, none of the values will satisfy this relationship. Therefore, it is a null set, and the probability will be 0. Now, if x is somewhere greater than x_1 or x_2, the probability exists only at a point. Other than that, the probability remains constant and equal to 0.

So, this is why, whenever $x \geq x_1$ and $x < x_2$, this value will be equal to $P_x(x_1)$. Now, if $x \geq x_2$ and $x < x_3$, then this value will just be cumulative; we have to add it. So, how does this come together? When you find the probability of $x \leq$ a certain value, you look at all s that belong to the set such that $x_s \leq x$. If x is between x_1 and x_2, and x is a discrete random variable taking only integer values, then this probability will just be for x_1. So, it's basically $P_x(x_1)$. Now, if x is between x_2 and x_3, then the probability of $x \leq x$ will include the probabilities at points x_1 and x_2.

This is how we can find the cumulative distribution function (CDF). In general, we can write the CDF, $F_x(x)$, as the summation of $P_x(x_k)$ for all x_k such that $x_k \leq x$. So, if $x \geq x_1$, this holds true; otherwise, it is equal to 0, because it only counts when it crosses the minimum value, allowing us to add some probability. Otherwise, if $x < x_1$, then none of the values satisfy this inequality; that is why f(x) will be 0. In general, the cumulative distribution function can be found using the probability mass function.

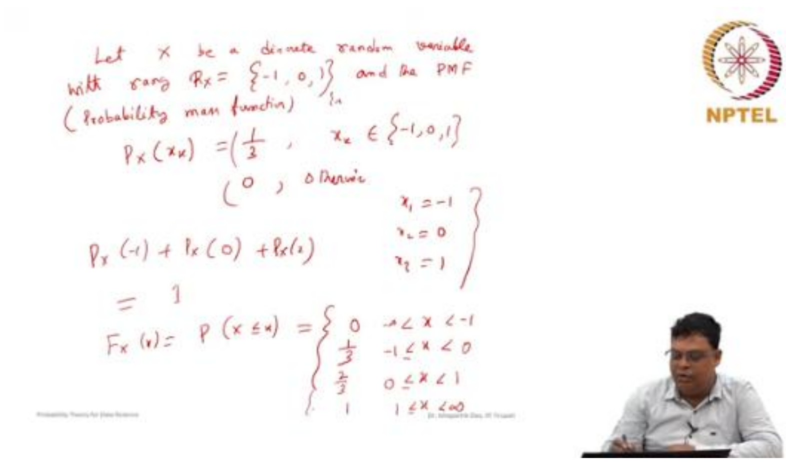


Let's see this example. Suppose x is a discrete random variable with range $R_x = \{-1, 0, 1\}$. The probability mass function is given by this: $x_1 = -1$, $x_2 = 0$, and $x_3 = 1$. In this notation, x_1 is the minimum value, x_2 is the next value, and x_3 is the following value. It is taking only a finite number of values. To qualify as a probability mass function, it must satisfy certain properties.

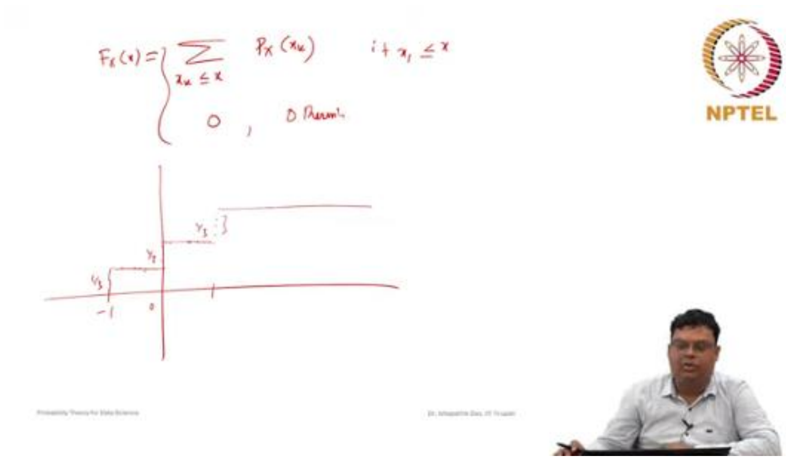
First of all, the values must be between 0 and 1. For example, $P_x(-1) = 1/3$, $P_x(0) = 1/3$, and $P_x(1) = 1/3$. If you take the sum, $P_x(-1) + P_x(0) + P_x(1)$, this will be $1/3 + 1/3 + 1/3$, which equals 1. So, this is a probability mass function. Now, if we want to find the cumulative distribution function, $F_x(x)$, it is defined as the probability that $x \leq x$.

This is equal to 0 whenever $x < x_1$, which is -1. So, if $x < -1$, the probability is 0. When x is between $x_1 = -1$ and $x_2 = 0$, the cumulative distribution function will be $P_x(-1) = 1/3$. Now, when $x \geq 0$ and $x < 1$, the cumulative distribution function will equal $P_x(-1) + P_x(0)$. Since this is a uniform random variable and all probabilities are the same, it is $1/3 + 1/3 = 2/3$.

Finally, for $x \geq 1$ and $x < \infty$, there are no other points. After adding the three points, $1/3 + 1/3 + 1/3 = 1$. This will be a step function. So, if you draw this function, it will be at -1, 0, and 1. The value starts at 0, then reaches 1/3 at $x = -1$, then goes to 2/3 at $x = 0$, and finally reaches 1 at $x = 1$.



The difference between these values is 1/3 each time. This is a probability mass function of a random variable. The probability mass function can satisfy certain relationships, meaning it can be a valid probability mass function. Let us discuss another random variable that may also be infinite. Suppose the range of the random variable x is limited to {1, 2, 3}.

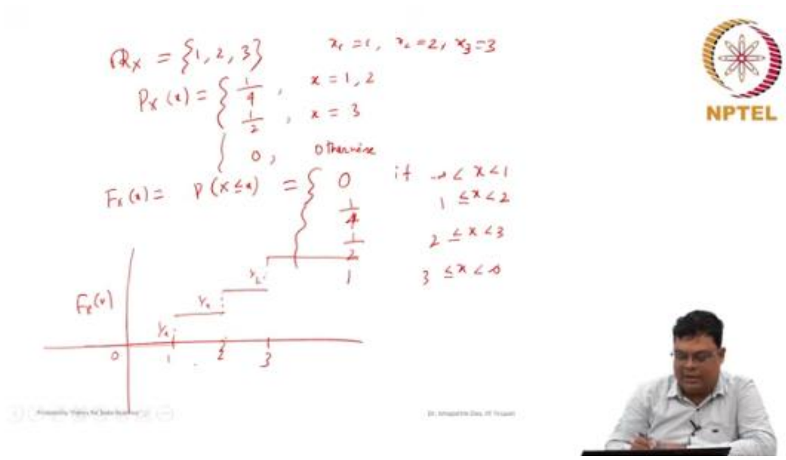


This means x1 = 1, x2 = 2, and x3 = 3. Now, what is the probability mass function? Let X be a discrete random variable with this range. The probability mass function is given by P(x) = 1/4 for X = 1 and X = 2, and P(x) = 1/2 when X = 3. It is equal to 0 otherwise. In this case, we can see that this satisfies the criteria for a probability mass function: it is always between 0 and 1, and it is 0 unless it is in the defined range.

If you sum these values, you get 1/4 + 1/4 + 1/2 = 1. Therefore, this is a valid probability mass function. Now, we want to find the cumulative distribution function from this probability mass function. F(x) is equal to the probability that $x \leq x$. Similarly, we can find this easily. This is equal to 0 if $x < 1$.

When $x \geq 1$ and $x < 2$, we will add the probability of x = 1, which is 1/4. When $x \geq 2$ and $x < 3$, we add 1/4 + 1/4 = 1/2. When $x \geq 3$ and $x < \infty$, this probability equals 1. We can represent this graphically. The points are at 0, 1, 2, and 3.

To summarize how we find the cumulative distribution function: for $x < 1$, it is 0; then it reaches 1/4 at x = 1; it adds to 1/2 at x = 2; and finally, it reaches 1 at x = 3. So, this is the cumulative distribution function $F_X(x)$. This corresponds to the random variable X. Next, we will discuss a discrete random variable that can take on infinite values. For example, consider an infinite series.



If you take an infinite series that is convergent, such as the summation of $1/n^2$ from $n = 1$ to ∞, we know that it converges to some number. This number may or may not be known, but it is actually $\pi^2/6$. You can verify this. Now, we can define a probability mass function using this infinite series. The probability that $X = x_k$ can be defined as some constant multiplied by $1/n^2$.

For $n = 1, 2$, and so forth, this is equal to 0 otherwise. In this case, the range of this random variable is 1, 2, 3, and so on. So, basically, $x_1 = 1$, $x_2 = 2$, and so on. This is countably infinite. Suppose we consider this constant as $6/\pi^2$.

So, instead of a constant, we use $6/\pi^2$ multiplied by $1/n^2$. We want to check whether this is a valid probability mass function. To be a valid probability mass function, the probabilities must be between 0 and 1. You can see that this is always between 0 and 1, and it is 0 if it is outside the range. Additionally, we need to take the sum from $n = 1$ to ∞ of $P(x_n)$.

This is the summation from $n = 1$ to ∞ of $(6/\pi^2)(1/n^2)$. Because $6/\pi^2$ is multiplied by the summation from $n = 1$ to ∞ of $1/n^2$, and this is a convergent series that converges to $\pi^2/6$, we find that: $(6/\pi^2) * (\pi^2/6) = 1$. Therefore, this is a valid probability mass function for a discrete random variable that takes on countably infinite points. You can also find the cumulative distribution function of this random variable. The cumulative distribution function $F(x)$ can be found similarly to the previous example.

It will be 0 if x is less than the minimum value of this range. If x is between 1 and 2, it will equal $P(x = 1)$, which is $6/\pi^2$. For $x \geq 2$ and $x < 3$, it will be the sum of $P(x = 1)$ and $P(x = 2)$, which is $6/\pi^2 + 6/4\pi^2$. This will continue for an infinite interval. If you draw the graph of the cumulative distribution function, your range will be 0, 1, 2, and so on. At 0, the value is 0; at 1, the value is $6/\pi^2$; and at 2, the value will be $6/4\pi^2$, and so on.

This graph will look like this; it will approach 1 asymptotically as $x \to \infty$. So, $F(x)$ converges to 1 as x approaches infinity. This is an example of a discrete random variable. Next, we will discuss continuous random variables and the properties of continuous random variables.