\PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 02

Lecture – 10

Cumulative Distribution Function

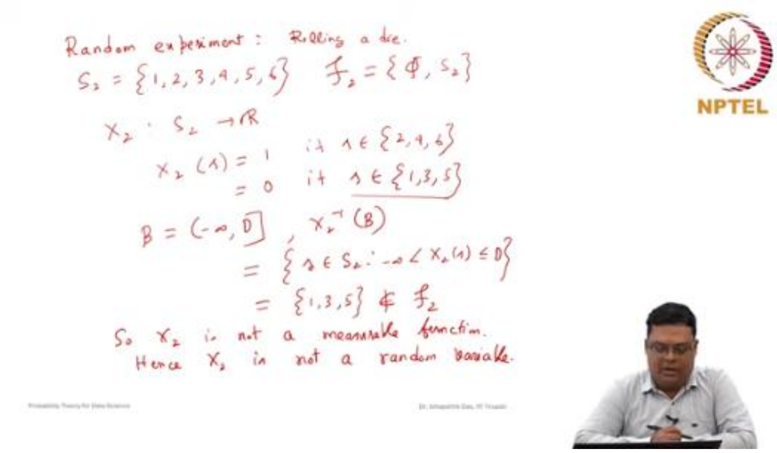
A random variable is a measurable function, $X: S \rightarrow \mathbb{R}$. Here, an example is given for the random experiment of tossing a coin. The sample space S = {head, tail}. We may define the random variable as X(head) = 1 and X(tail) = 0. We have shown that this is a measurable function and hence, a random variable.

There are many examples where we can show that something is not a random variable. For example, let us consider another random experiment, such as rolling a die. Then our sample space, $S_2 = \{1, 2, 3, 4, 5, 6\}$. Now, let's define $\mathcal{C}_2$. Suppose $\mathcal{A}_2 = \{\emptyset, S_2\}$, which is a σ-field, as it includes all unions and complements of these subsets. Now, we define a random variable $X_2: S_2 \rightarrow \mathbb{R}$.

Let $X_2(S) = 1$ if $S \in \{2, 4, 6\}$ and $X_2(S) = 0$ if $S \in \{1, 3, 5\}$. Suppose we take a Borel set, say $B = (-\infty, 1]$. Then, what is $X_2^{-1}(B)$? To find $X_2^{-1}(B)$, we look at all elements in S_2 satisfying the condition $X_2(S) \leq 1$ and $X_2(S) > -\infty$. This condition is satisfied by all elements, as X_2 only takes values 1 or 0. Hence, this inverse set is all of S_2, which belongs to $\mathcal{A}_2$, so no problem.

Now, if we consider the set $B = (-\infty, 0]$, it includes all elements S such that $X_2(S) \leq 0$ and $X_2(S) > -\infty$. Which elements satisfy this condition? Since $X_2(S) = 0$ for $S \in \{1, 3, 5\}$, the set is {1, 3, 5}. However, $\{1, 3, 5\} \notin \mathcal{A}_2$, as $\mathcal{A}_2$ only contains $\emptyset$ and S_2, the σ-field we are considering. Therefore, {1, 3, 5} does not belong to $\mathcal{A}_2$, so X_2 is not a measurable function and hence not a random variable.

So, X_2 is not a random variable. Now, I hope you understand that a random variable must be a measurable function from the sample space S to the Borel σ-field $\mathbb{R}$, meaning it has to be Borel measurable. We say "Borel measurable" because we are working with $\mathbb{R}$ and the Borel σ-field, considering this function with respect to the Borel σ-field. Now, let us discuss some numerical examples. I hope you have understood that events are defined by the random variable.



Random experiment : Rolling a die.

$S_2 = \{1, 2, 3, 4, 5, 6\}$ $\mathcal{F}_2 = \{\Phi, S_2\}$

$X_2 : S_2 \to \mathbb{R}$

$X_2(s) = 1$ if $s \in \{2, 4, 6\}$

$= 0$ if $s \in \{1, 3, 5\}$

$B = (-\infty, 0]$, $X_2^{-1}(B)$

$= \{s \in S_2 : -\infty < X_2(s) \le 0\}$

$= \{1, 3, 5\} \notin \mathcal{F}_2$

So X_2 is not a measurable function.

Hence X_2 is not a random variable.

It is defined as X = x, denoted as S such that X(S) = x. We can have various notations for this. For example, we can define another function, such as X_1, where $x_1 \le x \le x_2$. Let's do a numerical example. In the experiment of tossing a fair coin three times, the sample space S_1 consists of 8 equally likely sample points.

A fair coin means that when we say a coin is fair or unbiased, it means both heads and tails are equally likely; the probability of heads is ½. When you toss the coin three times, the possible outcomes are as follows: if you toss once, it will be heads or tails. The second toss can also be heads or tails, and the third toss can also be heads or tails. Your possible outputs for three tosses will look like this:

{H, H, H}, {H, H, T}, {H, T, H}, {H, T, T}, {T, H, H}, {T, H, T}, {T, T, H}, {T, T, T}.

So, there are a total of 8 observations, or sample points, for this random experiment. Now, if X is the random variable giving the number of heads obtained, we can evaluate it based on the observations. For example, if the observation is H, H, H, the number of heads is 3.

If the observation is H, H, T, the number of heads is 2. Similarly, if the observation is T, T, T, the number of heads is 0. The random variable X is defined as the number of heads obtained. For any point $s \in S_1$, X(s) is the number of heads obtained. So, to summarize:

For the case of H, H, H, X = 3.

For H, H, T, X = 2.

For T, H, H, X = 2.

For T, T, H, X = 1.

For T, T, T, X = 0.

This is how the random variable is defined. So, it can be easily shown that it is a measurable function. We have already discussed what a measurable function is and how to check it. By having a measurable function, we can define an event. An event is a subset of the sample space that belongs to the sigma field.

This means that since X is a measurable function, the inverse of any Borel set belongs to the sigma field C, indicating that it is an event. When we say that $X \leq 1$, it is a subset of S that belongs to C, making it an event as well. Whenever it is an event, we can define the probability of that event. This allows us to talk about the probability of X = 1 or the probability of $X \leq 1$. In general, we can also discuss the probability of $X \leq x$ because it is an event.

Now, if X is the random variable giving the number of heads obtained, we want to find the probability that X = 2 and the probability that $X < 2$. First, we need to understand what $X = 2$ means and what $X < 2$ means, and then we can compute these probabilities. Let’s first consider $X = 2$. This means all $s \in S_1$ satisfy the relationship where $X(s) = 2$. So, what are the points satisfying $X = 2$?

You can see that whenever the number of heads is actually 2, the outcomes HHT, HTH, and THH correspond to two heads. Any other combinations, such as HHH, would have 3 heads, so they do not satisfy this relationship. Other cases like H, T, T have only one head, and T, T, T have none. Therefore, the only combinations where the number of heads is 2 are HHT, HTH, THH. Since this is a fair coin, all eight outcomes in S_1 are equally likely.

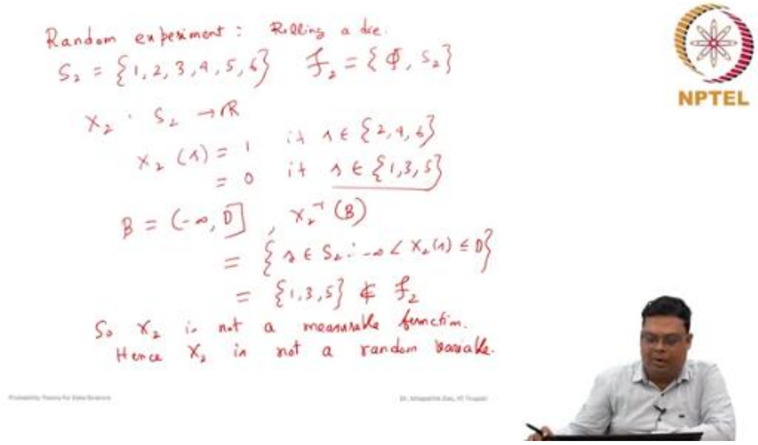
If it were not a fair coin, the probabilities would differ based on the likelihood of heads or tails. Therefore, since all elements in S_1 are equally likely, we can compute the probability that $X = 2$. The probability that $X = 2$ is calculated using the classical approach, assuming all outcomes are equally likely. There are a total of 8 possible outcomes, and $X = 2$ can occur in 3 different ways. Therefore, the probability is $3/8$.

Let A be a subset of S_1 defined by the event $X = 2$. Since the sample points are equally likely, we conclude that the probability of $X = 2$ is $3/8$. Next, we define another set, B, as $X < 2$. This is a subset of S and belongs to the sigma field. To define this, we consider all $s \in S_1$ that satisfy the relationship where the random variable $X(s) < 2$. When we look for points in S_1 satisfying this condition, we find that less than 2 means we can include 0 or 1.

The range of the random variable X is $\{0, 1, 2, 3\}$. Therefore, we are interested in values less than 2, which means we can have either 0 or 1. The outcome corresponding to 0 is TTT, where $X(s) = 0$. For a count of 1 head, we have TTH, THT, and HTT, each yielding 1 head. Any other outcomes will result in a number of heads greater than 1.

Thus, we have 4 elements satisfying the condition, and since all points are equally likely, the probability of event B ($X < 2$) is given by the classical approach: the total possible outcomes is 8, and the favorable outcomes for B is 4. Therefore, the probability of B is $4/8$, which simplifies to $1/2$.

In summary, we have discussed how the random variable is a measurable function, how it can be used to define events, and how we can find the probability associated with these events using the random variable. What is the probability? We can find it. Next, we will discuss the distribution function associated with a random variable.



Random experiment : Rolling a die.

$S_2 = \{1,2,3,4,5,6\}$ $\mathcal{F}_2 = \{\phi, S_2\}$

$X_2 : S_2 \to \mathbb{R}$

$X_2(s) = 1$ if $s \in \{2,4,6\}$

$= 0$ if $\underline{s \in \{1,3,5\}}$

$B = (-\infty, 0]$, $X_2^{-1}(B)$

$= \{s \in S_2 : -\infty < X_2(s) \leq 0\}$

$= \{1,3,5\} \notin \mathcal{F}_2$

So X_2 is not a measurable function.

Hence X_2 is not a random variable.

To understand this clearly, let’s repeat the previous example. Suppose we are conducting a random experiment by tossing a coin. The sample space consists of {H, T}. If we consider the sigma field as the power set of S, we recall that the power set contains all the subsets of S. Let C represent the power set, which includes {H}, {T}, S, and the null set ∅, containing all the subsets of S.

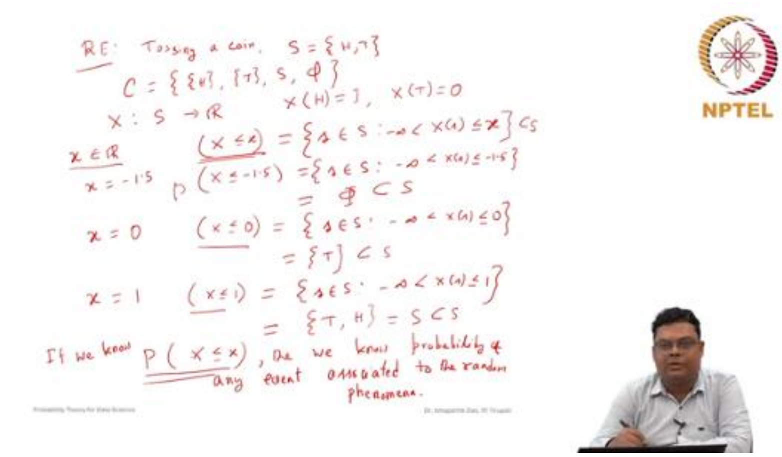
Now, we define a random variable. The random variable X is a measurable function mapping from S to ℝ. We can define any random variable, but let us consider the standard random variable where X(H) = 1 and X(T) = 0. This relates the abstract space to the set of real numbers, and we can also map back to this space using the random variable. Now, consider the set where X ≤ x.

Here, x represents any real number. This set consists of all s ∈ S that satisfy X(s) ≤ x. We understand that the left-hand side is greater than -∞ since X maps from S to ℝ. Let’s examine different values of x. For instance, if x is a negative number, such as -1.5, the set is defined as all s such that -∞ < X(s) ≤ -1.5.

Given that X(H) = 1 and X(T) = 0, none of these values satisfy this relationship, so this set is the null set, ∅, which is a subset of S. Now, consider if x is another value, such as 0. In this case, you can easily show that X ≤ 0 corresponds to a specific set, and I hope you understand this clearly now. You just have to find the values of s where X(s) ≤ 0. The only value satisfying this relationship is when X(T) = 0, which corresponds to T.

Now, if you consider when X = 1, then X ≤ 1 includes all s ∈ S, such that X(s) ≤ 1. This includes all the s satisfying this condition because X(H) = 1 and X(T) = 0. So, this encompasses both T and H, meaning this set is equal to S. All these sets are subsets of S and also belong to C, where C is the power set. By taking the complement of that set or performing any union operation, you can show that for all x ∈ ℝ, you can identify all the subsets of S.

If you can find the probability of these sets, then you can determine the probability of any event associated with the random phenomenon. This means that if you understand these values, you can find the probability of any event related to the random phenomenon. So, knowing this information is very important. If you know this, then we can determine the probability of any events associated with this random variable. We want to define it as a function because it is essentially a function that assigns a value based on any given x.



RE: Tossing a coin. $S = \{H, T\}$

$C = \{\{H\}, \{T\}, S, \phi\}$

$X : S \to \mathbb{R}$ $\quad X(H) = 1, \quad X(T) = 0$

$x \in \mathbb{R}$ $\quad (X \le x) = \{s \in S : -\infty < X(s) \le x\} \subset S$

$x = -1.5$ $\quad P\ (X \le -1.5) = \{s \in S : -\infty < X(s) \le -1.5\}$
$= \phi \subset S$

$x = 0$ $\quad (X \le 0) = \{s \in S : -\infty < X(s) \le 0\}$
$= \{T\} \subset S$

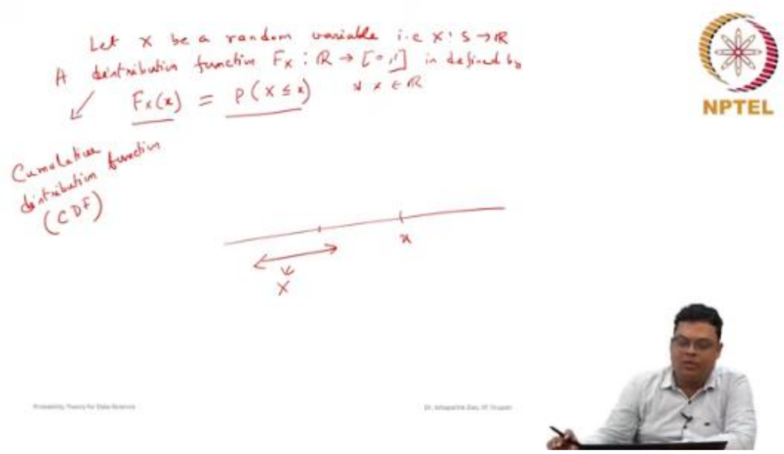
$x = 1$ $\quad (X \le 1) = \{s \in S : -\infty < X(s) \le 1\}$
$= \{T, H\} = S \subset S$

If we know $P(X \le x)$, the we know probability of any event associated to the random phenomena.

This function is denoted by F. For X as a random variable, let X be a random variable. We usually do not specify that X is a function from S to ℝ. For simplicity, we just say X is a random variable, meaning it is a function from S to ℝ. Moreover, it is a measurable function associated with the sample space of the random phenomenon.

Now, the distribution function F, denoted by F(x), is defined for any real number x. This distribution function represents the probability that X ≤ x. Sometimes, this is also known as the cumulative distribution function (CDF). We refer to it as cumulative because we are considering the probability that the random variable takes values less than or equal to x. For any real number x, you want to know the probability that X ∈ (-∞, x].

This means we are taking the probability that X < x, which applies for all x ∈ ℝ. If you understand that this function is the CDF, then you can determine the probability of any event associated with this random experiment or random phenomenon. This is the distribution function. Now let us discuss an example of the distribution function or cumulative distribution function. Let S be the sample space of a random experiment E, and let X be a random variable defined from S to ℝ.



Let X be a random variable i.e $X: S \to \mathbb{R}$

A distribution function $F_X : \mathbb{R} \to [0,1]$ is defined by

$$\underline{F_X(x)} = \underline{P(X \le x)} \quad \forall\, x \in \mathbb{R}$$

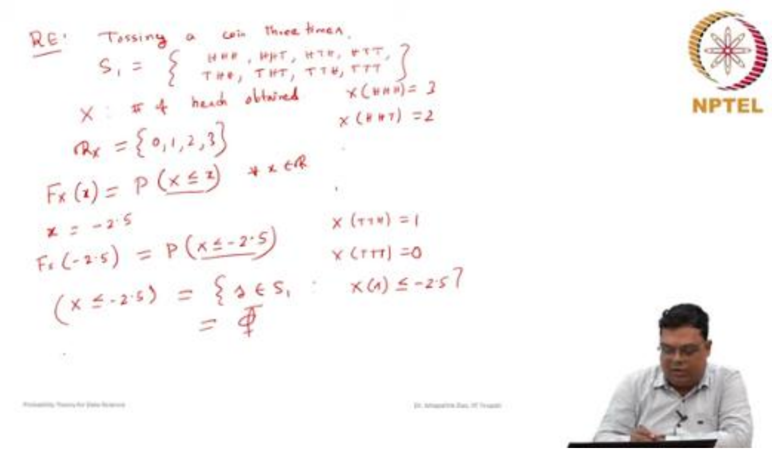
Cumulative distribution function (CDF)

The distribution function, or cumulative distribution function (CDF), is a function F from ℝ to the interval [0, 1], defined by F(x) = P(X ≤ x) for any real number x. Let us consider an example: tossing a coin three times. We will find the cumulative distribution function associated with this random experiment. In this case, we assume the coin is unbiased or fair, meaning there is no specified bias in the probability of heads. The sample space S_1 contains eight elements: HHH, HHT, HTH, THH, HTT, THT, TTH, and TTT.

The random variable X is defined by the number of heads obtained. For example, X(HHH) = 3, X(HHT) = 2, X(HTH) = 2, X(THH) = 2, X(HTT) = 1, X(THT) = 1, X(TTH) = 1, and X(TTT) = 0. Therefore, the range of X is {0, 1, 2, 3}. Now, we want to define and find the cumulative distribution function associated with this random variable X. The definition of the cumulative distribution function is F(x) = P(X ≤ x) for all real numbers x. We need to find the value of this function for some specific values of x.

For example, let's take a negative number, say -2.5. In this case, we find F(-2.5) = P(X ≤ -2.5). To determine this, we need to identify the event corresponding to this random variable. The event X ≤ -2.5 is defined as all s ∈ S_1 such that X(s) ≤ -2.5. Since X only takes non-negative values (0, 1, 2, or 3), there are no values in S_1 that satisfy this condition.

Therefore, this event corresponds to the null set ∅. The probability of X ≤ -2.5 is equivalent to the probability of the null set, which is 0. For any negative real number x, this results in a null set because X only takes positive values. Therefore, if x < 0, the set X ≤ x will also be the null set, leading to a probability of 0. Thus, for this distribution function in this example, F(x) = P(X ≤ x) is equal to 0 if x < 0 and greater than -∞.



RE: Tossing a coin three times.

$S_1 = \{ HHH, HHT, HTH, HTT, THH, THT, TTH, TTT \}$

X : # of head obtained $X(HHH) = 3$

$X(HHT) = 2$

$R_X = \{0, 1, 2, 3\}$

$F_X(x) = P(\underline{X \le x}) \quad \forall x \in \mathbb{R}$

$x = -2.5$ $X(TTH) = 1$

$F_X(-2.5) = P(\underline{X \le -2.5})$ $X(TTT) = 0$

$(X \le -2.5) = \{ s \in S_1 : X(s) \le -2.5 \}$

$= \Phi$

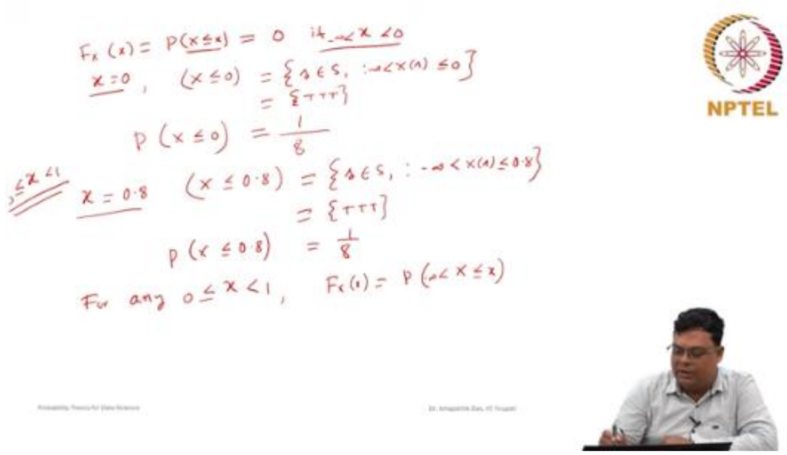
Probability Theory for Data Science Dr. Ishapathik Das, IIT Tirupati

We are not finding the probability over an interval; instead, we are taking a specific value and determining the probability of that event. Since this value is in the interval, the probability remains 0. Now, let us consider when x = 0. We want to find P(X ≤ 0). The set for this is defined as the set of s ∈ S_1 such that X(s) ≤ 0.

For simplicity, we often omit the left-hand side notation, understanding that it implies greater than -∞. In this case, X takes values of 0, 1, 2, and 3, with only TTT (tail, tail, tail) resulting in X = 0. Thus, the probability that X ≤ 0 corresponds to the event containing only one point, TTT. Out of the possible eight outcomes, there is only one way this can occur, which is why the probability is 1/8. Now, if you take any value, suppose x > 0 but x < 1.

For instance, let's say x = 0.8. So, you want to find the probability of this event, which is P(X ≤ 0.8). This is nothing but the set of s ∈ S_1 such that -∞ < X(s) ≤ 0.8. You can see that we are not finding the probability that X = 0.8; we are finding the probability that X ≤ 0.8. The random variable X can take values only in {0, 1, 2, 3}.

So, for X ≤ 0.8, you only have TTT, which gives X = 0. For all other values, X will be greater than 0. Therefore, the probability that X ≤ 0.8 is again 1/8. So, for any small x such that 0 ≤ x < 1, the probability that X ≤ x, denoted as F(x), is equal to P(X ≤ x), which is greater than -∞. So, this is the definition of F_X(x).



$F_X(x) = P(X \leq x) = 0$ if $-\infty < x < 0$

$x = 0$, $(X \leq 0) = \{s \in S, : -\infty < X(s) \leq 0\}$
$= \{TTT\}$

$P(X \leq 0) = \frac{1}{8}$

$0 \leq x < 1$

$x = 0.8$ $(X \leq 0.8) = \{s \in S, : -\infty < X(s) \leq 0.8\}$
$= \{TTT\}$

$P(X \leq 0.8) = \frac{1}{8}$

For any $0 \leq x < 1$, $F_X(x) = P(-\infty < X \leq x)$

NPTEL
Probability Theory for Data Science
Dr. Ishapathik Das, IIT Tirupati

Whenever x ∈ (0, 1), the probability is 1/8. Note that sometimes it can be confusing when we say we are finding x ≥ 0 and x < 1; that is not correct. Basically, we are taking this small x, which is between 0 and 1, and putting it here as x ≤ small x. That is why, for any value you consider between 0 and 1, the probability F(x) that we are finding is 1/8.