How to Understand the Difference Between CDF and PDF

Authored by
stats writer

December 30, 2025

In the field of statistics and probability, two fundamental concepts are constantly encountered: the Probability Density Function (PDF) and the Cumulative Distribution Function (CDF). While both are essential for modeling random phenomena, they serve distinct purposes.

The core distinction lies in what each function measures. A PDF describes the relative likelihood of a continuous random variable taking on a specific value, or the probability mass for a discrete variable. In contrast, the CDF provides the probability that a random variable will be less than or equal to a specific input value, offering a cumulative view of the distribution across all possible outcomes up to that point.

Understanding these functions is vital for data analysis and statistical modeling. This comprehensive tutorial provides a clear, simple explanation of the difference between the PDF and the CDF, building from the foundational concept of random variables.

The Foundation: Understanding Random Variables

Before delving into the complexities of Probability Density Functions and Cumulative Distribution Functions, it is necessary to establish a clear understanding of random variables. Simply put, a random variable, often denoted by the capital letter **X**, is a variable whose possible values are numerical outcomes resulting from a random phenomenon or process.

These variables are the essential building blocks of probability distributions. Whether we are flipping a coin, measuring human height, or tracking the time until a machine fails, we are dealing with a random process that yields numerical results described by a random variable. These variables fall into two major categories based on the nature of their possible values: **discrete** and **continuous**.

The choice between using a PDF or a CDF often depends heavily on whether the underlying random variable is discrete or continuous. While the core mathematical concepts are related, their application and interpretation differ significantly across these two categories.

Discrete Random Variables Explained

A discrete random variable is defined as a variable that can take on only a finite or countably infinite number of distinct values. These variables typically arise from counting processes where there are clear, separate gaps between possible outcomes--they jump from one integer value to the next without assuming intermediate fractional or irrational values.

Common examples involve scenarios where specific events are counted. For instance, the number of successful attempts in a series of trials or the count of defects in a manufactured batch are classic examples. These outcomes must be whole, positive integers, such as 0, 1, 2, 3, and so forth, up to a defined maximum or infinity.

Consider the following practical applications of discrete random variables:

The number of times a coin lands on **tails** when flipped 20 independent times. The outcomes can only be integers from 0 to 20.
The number of times a standard six-sided dice lands on the number **4** after being rolled 100 times.
The count of customers entering a store during a one-hour period.

Continuous Random Variables Explained

In contrast to their discrete counterparts, a **continuous random variable** is one that can assume any value within a specified interval or range. This means there is an uncountably infinite number of possible outcomes. These variables typically arise from measurement processes rather than counting processes.

Because measurement involves precision that can theoretically extend indefinitely, the variable can take on fractional, decimal, or irrational values. For example, the height of a person might be 60.2 inches, 65.2344 inches, or even 70.431222 inches. Since the interval between any two measurements can always be subdivided infinitely, there is an infinite continuum of possible values.

Key examples of continuous random variables include physical measurements:

The **height** of a person in meters or inches.
The **weight** of an animal in kilograms or pounds.
The **time** required to complete a task, such as running a mile.

A helpful rule of thumb for distinguishing between the two types of random variables is the concept of **measurement versus counting**. If the outcome is determined by counting distinct events (like coin flips or customer arrivals), the variable is discrete. If the outcome is determined by measuring a quantity (like distance, mass, or time), the variable is continuous. This distinction is crucial for determining whether to use a probability mass function or a Probability Density Function (PDF).

Defining the Probability Density Function (PDF)

The Probability Density Function (PDF), denoted as *f(x)*, is primarily used for **continuous random variables**. It describes the relative likelihood of the variable falling within a specific range of values. Crucially, for continuous variables, the PDF itself does not give the probability of a single exact point; rather, the probability of the variable falling between two points (*a* and *b*) is found by integrating the PDF over that interval.

If we are working with a **discrete random variable**, the equivalent concept is the **Probability Mass Function (PMF)**. The PMF directly gives the probability that the random variable **X** is exactly equal to a specific value *x*, denoted as $P(X = x)$. While many introductory resources may use "PDF" loosely to cover both cases, technically, PMF applies strictly to discrete distributions and PDF applies only to continuous distributions.

Calculating Probabilities for Discrete Variables (PMF Example)

To illustrate the concept of probability distribution for a discrete variable, consider the classic example of rolling a fair, six-sided dice one time. Let *X* be the random variable representing the number the dice lands on. Since each outcome (1, 2, 3, 4, 5, or 6) is equally likely, the probability for each specific event is 1/6. This is determined by the Probability Mass Function (PMF):

$P(X < 1)$: 0

$P(X = 1)$: 1/6

$P(X = 2)$: 1/6

$P(X = 3)$: 1/6

$P(X = 4)$: 1/6

$P(X = 5)$: 1/6

$P(X = 6)$: 1/6

$P(X > 6)$: 0

This example demonstrates that for a discrete distribution, the probability of the random variable taking on any single integer value can be calculated precisely.

The Challenge of Continuous Probability

When dealing with a continuous random variable, the interpretation of the PDF changes significantly. Because a continuous variable can take on an infinite number of values within any range, the probability that the variable will equal any single, exact point is mathematically zero. It is impossible to calculate $P(X = x)$ directly for a continuous PDF, as this calculation is essentially 1 divided by infinity.

Consider the measurement of weight, a continuous variable. If we ask for the probability that a burger weighs exactly 0.25 pounds, we must account for infinite precision (0.2500000... pounds). A burger might weigh 0.250001 pounds or 0.249999 pounds, making the likelihood of hitting the precise target of 0.25 pounds negligible.

Therefore, for continuous variables, the PDF is only used to calculate the probability of the variable falling within a specified **interval** (e.g., the probability that the burger weighs between 0.24 lbs and 0.26 lbs). This calculation requires finding the area under the PDF curve using integration.

Defining the Cumulative Distribution Function (CDF)

The Cumulative Distribution Function (CDF), denoted as *F(x)*, provides a fundamentally different view of the probability distribution. Instead of measuring the likelihood of a specific outcome, the CDF measures the probability that a random variable **X** will take on a value **less than or equal to** a given value *x* (i.e., $P(X \leq x)$).

The CDF is particularly useful because it applies equally well to both discrete random variables and continuous random variables, offering a unified way to describe the distribution across the entire range of possible outcomes. It effectively summarizes the probabilities, accumulating them as *x* increases.

Returning to the dice rolling example, where *X* is the outcome of a single roll, the CDF illustrates how the probability accumulates as we consider higher possible values:

P(X ≤ 0): 0

P(X ≤ 1): 1/6

P(X ≤ 2): 2/6 (Probability of rolling 1 OR 2)

P(X ≤ 3): 3/6

P(X ≤ 4): 4/6

P(X ≤ 5): 5/6

P(X ≤ 6): 6/6

P(X > 6): 0

It is clear from the calculation that the probability of the dice landing on a number less than or equal to 6 is 6/6, or 1, confirming that the cumulative probability covers all possible outcomes of the event.

Essential Properties of the CDF

The Cumulative Distribution Function possesses three mathematical properties that hold true for any distribution, whether discrete or continuous. These properties ensure that the function is a valid representation of probability accumulation:

Lower Bound: The probability that a random variable takes on a value less than the smallest possible outcome is zero. Mathematically, $F(-\infty) = 0$. For the dice example, the probability of rolling a number less than 1 is zero.

Upper Bound: The probability that a random variable takes on a value less than or equal to the largest possible outcome is one. Mathematically, $F(+\infty) = 1$. This signifies 100% certainty that the outcome will fall within the total range of possibilities.

Monotonicity: The CDF must always be **non-decreasing**. As the value of x increases, the cumulative probability $F(x)$ can either stay the same or increase, but it can never decrease. For example, the probability of rolling ≤ 3 (3/6) must be greater than or equal to the probability of rolling ≤ 2 (2/6).

Visualizing the CDF is often helpful, especially for continuous distributions, where it typically forms a smooth, S-shaped curve (a sigmoidal function). You can use an to visualize a cumulative distribution function.

The Mathematical Relationship Between CDF and PDF

While the PDF (or PMF) and the CDF appear to measure different aspects of a distribution, they are intrinsically linked through the fundamental principles of calculus. For continuous random variables, the relationship is defined by integration and differentiation.

Specifically, the CDF is the integral of the PDF. This means that the value of the CDF at any point x, denoted $F(x)$, is equal to the total area under the PDF curve, $f(x)$, from negative infinity up to x. This integral operation accumulates the relative probabilities defined by the PDF to give the cumulative probability.

Conversely, the PDF, $f(x)$, is the derivative of the CDF, $F(x)$. The derivative measures the instantaneous rate of change of the CDF. Since the CDF is always non-decreasing, taking its derivative yields the density function, reflecting how quickly probability is accumulating at that specific point. This inverse relationship confirms that both functions describe the exact same underlying distribution, merely viewed from different mathematical perspectives.

For a rigorous, in-depth explanation of the proof demonstrating why the PDF is the derivative of the CDF, readers should consult advanced statistical textbooks or foundational texts on measure theory and probability.