PMFs, PDFs, and CDFs

Sophie Chowgule

Summary

Probability mass functions (PMFs), probability density functions (PDFs), and cumulative distribution functions (CDFs) are fundamental concepts in statistics. These functions describe how probabilities are distributed across the possible outcomes of random events. PMFs, PDFs, and CDFs are commonly used to model probability distributions, helping to visualize and understand the behavior of random processes. This guide will explore the role of each function, how they differ, and highlight their applications.

Before reading this guide, it is highly recommended that you read [Guide: Introduction to probability],[Guide: Discrete random variables versus continuous random variables], and [Guide: Introduction to Integration].

Introduction

PMFs, PDFs, and CDFs are key tools in the study of probability, used to model and analyze the behaviour of random variables. These functions describe how probabilities are distributed across the possible outcomes of random events. In turn, a probability distribution provides a complete description of how these probabilities are assigned to all the possible values of a random variable, whether discrete or continuous. Understanding these functions is important for analyzing data, making predictions, and applying statistical methods to solve real-world problems.

What is a probability mass function (PMF)?

As you have seen in (Fact Sheet: Discrete random variables versus continuous random variables), a discrete random variable can take on a countable number of distinct outcomes. For example, rolling a die can result in only one of six possible outcomes. A **probability mass function (PMF)** assigns probabilities to each individual outcome of a discrete random variable, helping to determine the likelihood of a specific event occurring. In the case of the six-sided die, the PMF assigns a probability of $1/6$ to each outcome, reflecting that each outcome is equally as likely. When applied to the entire discrete random variable, a PMF describes how the total probability is distributed across all possible outcomes. More formally:

Definition of a PMF

A **probability mass function** or **PMF** is a function $p(x)$ that, when applied to a discrete random variable X, returns the probability that X is equal to a specific value x. The PMF can be expressed as:

$$p(x) = P(X = x)$$

where $P(X = x)$ is the probability that X equals x.

For a PMF to be considered a valid probability distribution for a random variable, it must satisfy two main conditions:

- **Non-negativity**: The probability assigned to each possible outcome must be greater than or equal to zero, that is:

$p(x) = P(X = x) \geq 0$ for all values of x.

- **Honesty condition**: The sum of probabilities of all possible outcomes of a discrete random variable must be equal to one:

$$\sum_x p(x) = \sum_x P(X = x) = 1$$

Note: The symbol $\sum$ is called the **sigma notation** and represents the sum of values, for example, adding the probabilities from all possible values of a random variable X. For more see Guide: Introduction to sigma notation

Tip

These conditions are derived from the laws of probability. For more, see [Guide: Introduction to probability].

Example 1

You are given a fair six-sided die. Let the discrete random variable X represent the result of rolling the die, and x represent the possible outcomes: $1, 2, 3, 4, 5$, and 6. Since the die is fair, each outcome has an equal probability of $1/6$, so the PMF for this scenario is given by:

$$p(x)$$

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Figure 1: PMF of rolling a fair six-sided die as in example 1.

Non-negativity: From the table you can see all $P(X = x) = 1/6 \geq 0$, so each probability is positive, satisfying the non-negativity requirement.

Honesty:

$$\sum_x P(X = x) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1,$$

confirming that the total probability of the PMF equals 1, meeting the honesty condition. Since the PMF satisfies both the non-negativity and honesty conditions, it is a valid PMF which represents the scenario of rolling a fair six-sided die.

i Example 2

Imagine you flip a fair coin twice. Let the discrete random variable X represent the number of times the coin lands on heads, so the possible outcomes x are $0, 1,$ or 2. Since the coin is fair, with equal probabilities for both heads and tails, the probabilities are determined by counting the outcomes. For example, the probability of both flips landing on heads is $1/4$ since it's one of the four possible outcomes. So, the PMF for this scenario is:

$p(x)$

x	0	1	2
$P(X = x)$	0.25	0.5	0.25

Figure 2: PMF of flipping a fair coin twice as in example 2.

You can see that this PMF also satisfies both key conditions:

Non-negativity: From looking at the table, all probabilities are positive, as $P(X = x) \geq 0$ for all values of x.

Honesty: The sum of probabilities equals 1:

$$\sum_x P(X = x) = 0.25 + 0.5 + 0.25 = 1$$

So, this is a valid PMF representing the number of heads when you flip a fair coin twice.

Example 3

A common example of a PMF is that of the **binomial distribution**. This is a type of PMF used to model scenarios with only two possible outcomes: a success, with probability p, or a failure with probability q. For this formula, x is the number of successes in n number of trials, p is the probability of success in a single trial, and $q = 1 - p$, is the probability of failure. The PMF for a binomial distribution is:

$$p(x) = \binom{n}{x} p^x q^{(n-x)} = \frac{n!}{(n-x)!x!} p^x q^{(n-x)}$$

Binomial distributions are often used to model real life scenarios, such as the probability of heads occurring in multiple fair coin flips. In this example, heads will be considered a success and tails a failure. This time imagine flipping a coin 10 times, where the probability of success (heads) is 0.5. The figure below shows the probability distribution for the number of heads:

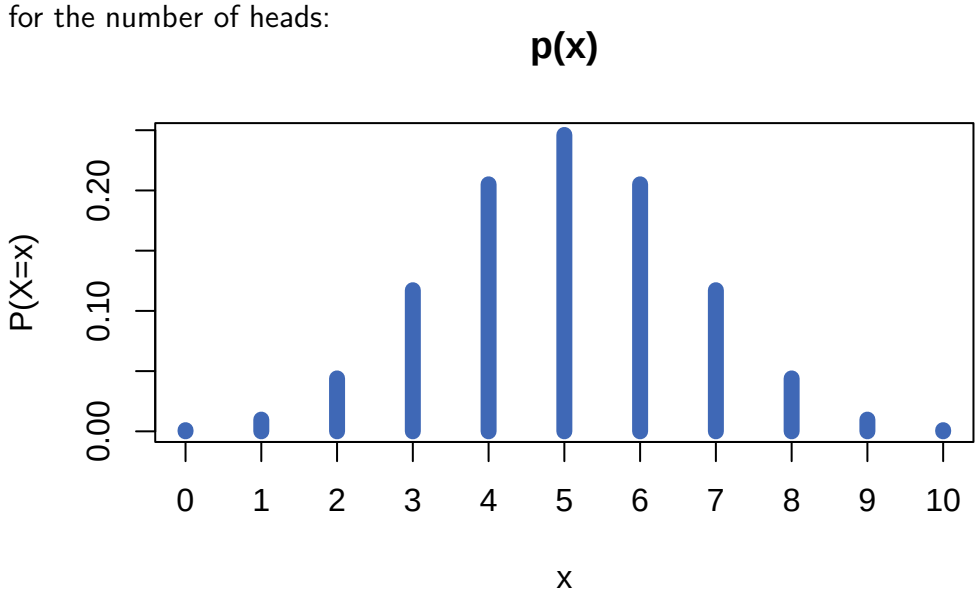

Figure 3: PMF of the binomial distribution for 10 coin flips as in example 3.

You can find that all binomial distributions are valid PMFs. This is because binomial distributions come from the **binomial theorem**, which shows why the sum of all probabilities equal to 1. The binomial theorem states that:

$$\binom{n}{x} p^x q^{(n-x)} = (p+q)^n$$

And since $q = 1 - p$:

$$(p+q)^n = (p+(1-p))^n = (1)^n = 1$$

So, the sum of the probabilities over all possible values of x equals 1. This satisfies the

honesty condition confirming that all binomial distributions are valid PMFs.

What is a probability density function (PDF)?

Unlike discrete random variables, continuous random variables can take on any number of values within a range. For instance, a person's height could be 170cm, 170.1cm or 170.000001cm. Since these values cannot be counted, calculating the probability distribution for continuous random variables requires the use of a **probability density function (PDF)**. Unlike PMFs, PDFs assign probabilities to intervals rather than to specific values and so, are key for determining the likelihood of a random variable falling within a given range.

When applied over all possible values of a continuous random variable, the PDF can be represented as a curve that shows the total probability distribution across all possible outcomes. The probability that X lies within an interval $[a, b]$ is equal to the area under the PDF, $f(x)$ between a and b as shown below:

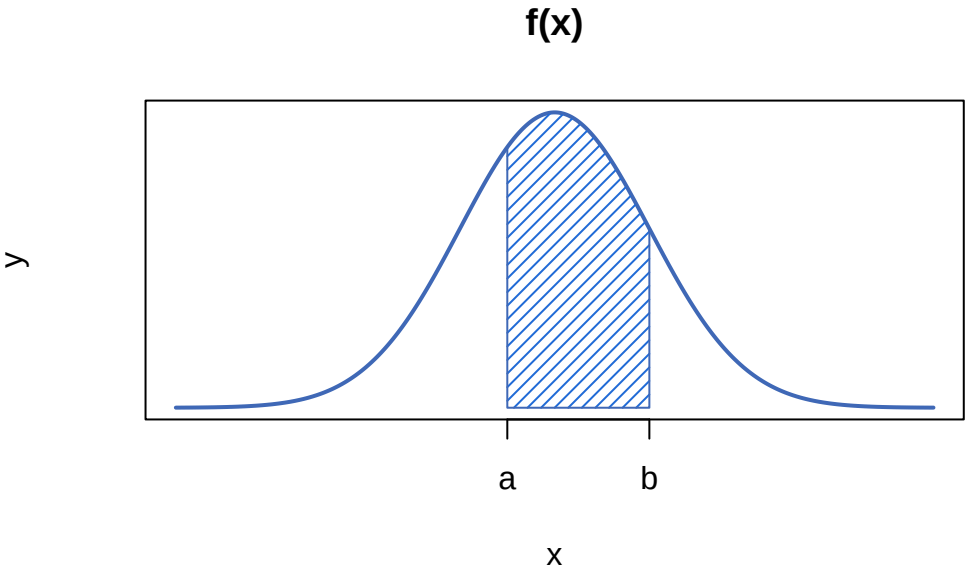

Figure 4: A PDF for the continuous random variable X, where the shaded area represents the probability that X lies between a and b.

Definition of a PDF

A **probability density function** or **PDF**, is a function, $f(x)$, that represents the distribution of probabilities across a continuous random variable X. The probability that X lies within an interval $[a, b]$ is found by integrating the PDF over that interval:

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

where $P(a \leq X \leq b)$ is the probability that X lies between a and b $(a < b)$.

Just like PMFs, PDFs must satisfy two main conditions to be considered valid:

- **Non-negativity**: The PDF must be greater than or equal to zero over its entire range of possible values:

  $f(x) \geq 0$ for all values of x.

- **Honesty condition**: The area under the entire PDF, $f(x)$, must equal 1:

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

  This holds only because of the non-negativity condition, making sure that the total probability remains valid.

Warning

PDFs cannot return probabilities at distinct values. With continuous random variables, there are infinite possible outcomes, so the probability at any specific point is essentially zero. Instead, probabilities for continuous random variables are calculated over intervals, not at individual values, because the area under the curve at a single point is always zero!

$$\int_{a}^{a} f(x)\,dx = 0$$

For more see [Guide: Properties of integration]

Example 4

You are given X, a continuous random variable uniformly distributed on the interval $[0, 1]$. This means that all values between 0 and 1 are equally likely to occur. The PDF for X is given by:

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

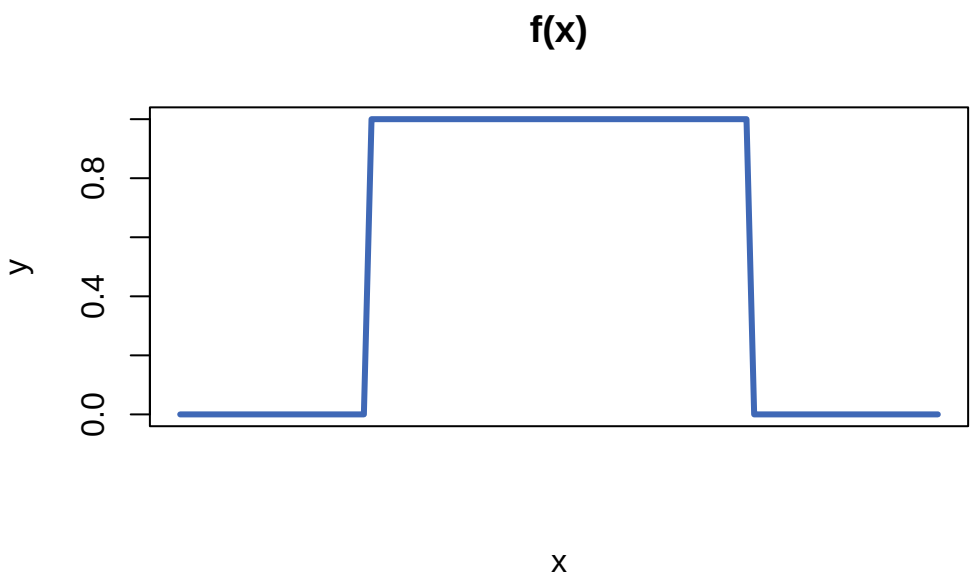

Figure 5: PDF of a uniform distribution between 0 and 1 as in example 4.

To check if this is a valid PDF, you need to confirm that it satisfies the two key conditions:

Non-negativity: $f(x) \geq 0$ for all values of x, as $f(x) = 0$ in $[0, 1]$ and 0 otherwise.

Honesty: The integral of the function $f(x)$ represents the area under the PDF:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} 1\,dx + \int_{1}^{\infty} 0\,dx = \left[\,x\,\right]_{0}^{1} = 1$$

And this satisfies the honesty condition, confirming that the uniform distribution over $[0, 1]$ is a valid PDF. You can find that all uniform distributions are valid PDFs and can confirm this by looking into why their total probability always equals 1:

The formula for a uniform distribution over any interval $[a, b]$ is given by:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

To satisfy the honesty condition, the integral of the PDF over the interval $[a, b]$ must equal 1:

$$\int_{a}^{b} \frac{1}{b-a}\,dx = \frac{1}{b-a}\int_{a}^{b} 1\,dx = \frac{1}{b-a}\left[x\right]_{a}^{b} = \frac{1}{b-a}(b-a) = 1$$

And so you can see that all uniform distributions, whether over $[0, 1]$ or any other interval $[a, b]$ are valid PDFs.

To find the probability that X lies between 0.25 and 0.5, you can calculate the area under

the curve of the PDF within the interval:

$$\int_{0.25}^{0.5} f(x)\,dx = \int_{0.25}^{0.5} 1\,dx = \left[\,x\,\right]_{0.25}^{0.5} = 0.5 - 0.25 = 0.25$$

Therefore, the probability that X lies in the interval $[0.25, 0.5]$ is 0.25.

i Example 5

The **normal distribution** is a widely used example of a probability density function (PDF). It is often employed to model naturally occurring phenomena such as height, weight, and other biological measurements. The general PDF of the normal distribution is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

where σ is the standard deviation and μ is the mean.

The standard normal distribution with a mean of 0 and a standard deviation of 1, is shown below:

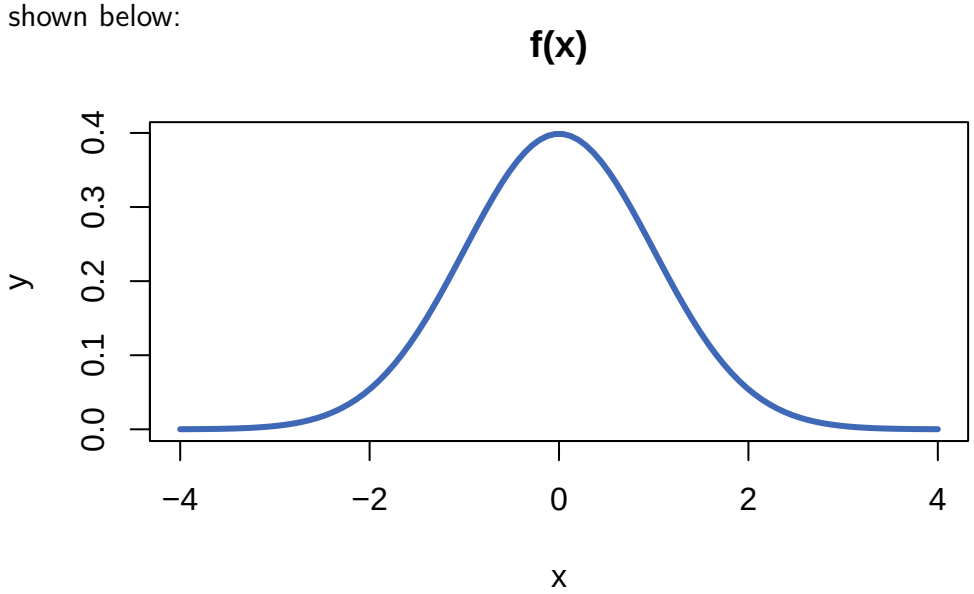

Figure 6: PDF of a normal distribution with mean 0 and standard distribution 1, as in example 5.

You will find that normal distributions share a similar shape, with the peak centered at the mean, and that all normal distributions are considered valid PDFs.

For more see:[Guide: Introduction to probability distributions]

Key differences between PMFs and PDFs

Probability mass function (PMF)	**Probability density function (PDF)**
Finds the probabilities of **discrete random variables**	Finds the probabilities of **continuous random variables**
Probabilities range from **0 to 1** for each exact outcome	Probabilities are calculated over intervals as the probability of an exact outcome is always **0**.
Provides likelihood that X occurs at an **exact value**	Provides likelihood that X lies within an **interval**
Sum of probabilities equals 1	**Integral** over entire domain equals 1

Figure 7: Table comparing the key differences between PMFs and PDFs.

What is a cumulative distribution function (CDF)?

Another key function in the area of probability distributions is the **cumulative distribution function (CDF)**. A CDF returns the probability that a random variable X is less than or equal to a specific value x. CDFs can be derived from both probability mass functions (PMFs) for discrete random variables and probability density functions (PDFs) for continuous random variables.

i Definition of a CDF

A **cumulative distribution function** or **CDF**, is a function, $F(x)$ that returns the probability that X is less than or equal to a variable x.
For a discrete random variable with a PMF $p(x)$, the CDF is given by:

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$$

For a continuous random variable with a PDF $f(x)$, the CDF is given by:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\, dy$$

Where y are the outcomes "less than or equal to" x.

Warning

CDFs are always **non-decreasing**. This is because they deal with cumulative probabilities, which represent the total probability up to a certain point. Since the probability of an event can only increase or remain the same as more outcomes are considered, the probability of a random variable being less than or equal to any value x is always non-decreasing as x increases!

Example 6

Consider you roll a fair six-sided die, as in Example 1. Since this scenario involves a PMF, the cumulative distribution function (CDF) can be derived using the following method. To find the probability of rolling a three or lower, sum the probabilities of rolling each number less than or equal to three:

$$F(3) = P(X \leq 3) = \sum_{x \leq 3} p(x) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Therefore, the probability of rolling a three or lower is 50%.
You can extend this to show that the entire CDF is given by:

$$F(x)$$

x	1	2	3	4	5	6
$P(X \leq x)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{5}{6}$	1

Figure 8: CDF for rolling a fair six-sided die, as in example 6.

Example 7

Imagine you flip a coin twice, like in Example 2. Since this scenario represents a discrete random variable with a PMF, the CDF, can be derived by summing the probabilities of outcomes less than or equal to x, similar to the previous example:

$$F(x)$$

x	0	1	2
$P(X \leq x)$	0.25	0.75	1

Figure 9: CDF for flipping a fair coin twice, as in example 8.
To find the probability that X is greater than x, you can subtract the correlating value in the CDF from the total probability.
For example, to find the probability that X is greater than 1:

$$P(X > 1) = 1 - F(1) = 1 - P(X \leq 1) = 1 - 0.75 = 0.25$$

Therefore, the probability that X is greater than 1 is 0.25.

i Example 8

Consider a continuous random variable X uniformly distributed between 0 and 1, which you can see in example 4. The PDF of X was given by:

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

To find the probability that X is less than or equal to 0.5, use the formula from the definition of the CDF:

$$F(0.5) = P(X \leq 0.5) = \int_{-\infty}^{0.5} f(x)\,dx = 0.5$$

meaning the probability of X being less than or equal to 0.5 is 50%.
On the other hand, to find the probability that X is greater than 0.5, subtract the CDF value at 0.5 from the total probability, 1:

$$P(X > 0.5) = 1 - F(0.5) = 1 - 0.5 = 0.5$$

So, the probability that X is greater than 0.5 is also 50%.

Quick check problems

Further reading

[For more questions on the subject, please go to Questions: PMFs, PDFs, and CDFs]

For more on probability distributions see: [Guide: Introduction to probability distributions]

Version history

v1.0: initial version created 12/24 by Sophie Chowgule

This work is licensed under CC BY-NC-SA 4.0.