In probability theory, a probability density function (PDF), or density of a continuous random variable, is a function that describes the relative likelihood for this random variable to take on a given value.

Probability density function is defined by following formula:

$$P(a \le X \le b) = \int_a^b f(x) d_x$$

Where −

- $[a, b]$ = Interval in which x lies.
- $P(a \le X \le b)$ = probability that some value x lies within this interval.
- d_x = b-a

Probability Density Function (PDF) is used to define the probability of the random variable coming within a distinct range of values, as objected to taking on anyone value.

The Probability Density Function(PDF) is the probability function which is represented for the density of a continuous random variable lying between a certain range of values. It is also called a probability distribution function or just a probability function. However, in many other sources, this function is stated as the function over a general set of values or sometimes it is referred to as cumulative distribution function or sometimes as probability mass function(PMF). But the actual truth is PDF is defined for continuous random variables whereas PMF is defined for discrete random variables.

The probability density function is defined in the form of an integral of the density of the variable density over a given range. It is denoted by f (x). This function is positive or non-negative at any point of the graph and the integral of PDF over the entire space is always equal to one.

Probability Density Function Formula

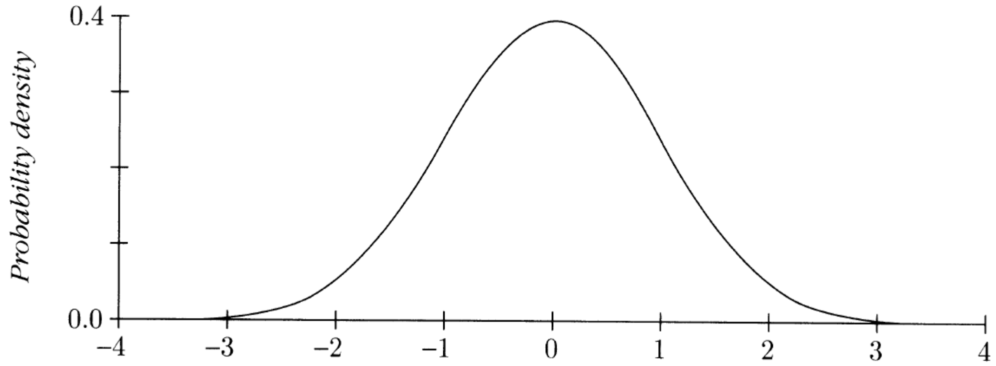
In case of a continuous random variable, the probability taken by X on some given value x is always 0. In this case, if we find P(X = x), it does not work. Instead of this, we require to calculate the probability of X lying in an interval (a, b). Now, we have to calculate it for P(a< X< b). This can be done by using a PDF. The Probability distribution function formula is defined as,

$$P(a < X < b) = \int_a^b f(x)$$

The Normal Distribution

A random variable Z has *standard normal distribution* if Z has as its probability density the *standard normal density*

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad (-\infty < z < \infty)$$



The constant $1/\sqrt{2\pi}$ is put in the definition of the standard normal density so the total area under the standard normal curve $y = \phi(z)$ is 1. This is the first integral in the following box:

Standard Normal Integrals

$$\int_{-\infty}^{\infty} \phi(z)\, dz = 1; \qquad \int_{-\infty}^{\infty} z\phi(z)\, dz = 0; \qquad \int_{-\infty}^{\infty} z^2\phi(z)\, dz = 1.$$

Normal (μ, σ^2) Distribution

If Z has standard normal distribution and μ and σ are constants with $\sigma \geq 0$, then

$$X = \mu + \sigma Z$$

has mean μ, standard deviation σ, and variance σ^2. The distribution of X is called the *normal distribution with mean μ and variance σ^2*, abbreviated normal (μ, σ^2). So X has normal (μ, σ^2) distribution if and only if the standardized variable

$$Z = (X - \mu)/\sigma$$

has normal $(0, 1)$ or *standard normal* distribution. To find $P(c < X < d)$, change to standard units and use the standard normal table

$$P(c < X < d) = P(a < Z < b) = \Phi(b) - \Phi(a)$$

$$\text{where} \qquad a = (c - \mu)/\sigma \qquad Z = (X - \mu)/\sigma \qquad b = (d - \mu)/\sigma$$

Formula for the normal (μ, σ^2) density. For $\sigma > 0$, the formula is

$$\frac{1}{\sigma}\phi((x - \mu)/\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} \qquad (-\infty < x < \infty).$$

Joint Distributions

Given two random variables X and Y defined in the same setting, we can consider their *combined* or *joint* outcome (X, Y) as a random pair of values. By definition, (X, Y) has value (x, y) if X has value x and Y has value y. Thus the event that $((X, Y) = (x, y))$ is the intersection of the events $(X = x)$ and $(Y = y)$, and is usually denoted $(X = x, Y = y)$. So commas mean intersections in statements about random variables.

The range of the joint outcome (X, Y) is the set of all ordered pairs (x, y) with x in the range of X, y in the range of Y, and $P(X = x, Y = y) > 0$. If the range of X is represented by points on a horizontal line, and the range of Y by points on a vertical line, then the range of (X, Y) is represented by a set of points in the plane. Alternatively, the range of (X, Y) may be represented by a set of paths through a tree diagram, as in Chapter 1.

The distribution of (X, Y) is called the *joint distribution* of X and Y. This distribution is determined by the probabilities

$$P(x, y) = P(X = x, Y = y)$$

which must satisfy

$$P(x, y) \geq 0 \qquad \text{and} \qquad \sum_{\text{all } (x,y)} P(x, y) = 1$$

Random Variables with the Same Distribution

Random variables X and Y have the *same* or *identical distribution* if X and Y have the same range, and for every value v in this range,

$$P(X = v) = P(Y = v).$$

Computing probabilities from a joint distribution. Once the joint distribution of X and Y has been calculated, the probability of any event defined in terms of X and Y can be found. Simply sum the probabilities $P(x, y)$ over the relevant set of pairs (x, y):

Independent Random Variables

Random variables X and Y are *independent* if

$$P(X = x, Y = y) = P(X = x)P(Y = y) \quad \text{for all } x \text{ and } y$$

If X and Y are independent random variables, then every event determined by X is independent of every event determined by Y:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Conceptually, independence means that conditioning on a given value of X does not affect the distribution of Y, and vice-versa. Thus the above definition of independence can be re-expressed as follows in terms of conditional distributions:

Conditional Distributions and Independence

The following three conditions are equivalent:

- X and Y are independent;
- the conditional distribution of Y given $X = x$ does not depend on x;
- the conditional distribution of X given $Y = y$ does not depend on y.

Standard Deviation and Normal Approximation

If you try to predict the value of a random variable X by its mean $E(X) = \mu$, you will be off by the random amount $X - \mu$. It is often important to have an idea of how large this deviation is likely to be. Because

$$E(X - \mu) = E(X) - \mu = 0$$

it is necessary to consider either the absolute value or the square of $X - \mu$ to get an idea of the size of the deviation without regard to sign. Because the algebra is easier with squares than with absolute values, it is natural to first consider $E[(X - \mu)^2]$, then take a square root to get back to the same scale of units as X.

Definition of Variance and Standard Deviation

The *variance* of X, denoted $Var(X)$, is the mean squared deviation of X from its expected value $\mu = E(X)$:

$$Var(X) = E[(X - \mu)^2]$$

The *standard deviation* of X, denoted $SD(X)$, is the square root of the variance of X:

$$SD(X) = \sqrt{Var(X)}$$

Intuitively, $SD(X)$ should be understood as a measure of how spread out the distribution of X is around its mean μ. Because $Var(X)$ is a central value in the distribution of $(X - \mu)^2$, its square root $SD(X)$ gives a rough idea of the typical size of the absolute deviation $|X - \mu|$. Variance always appears as an intermediate step in the calculation of standard deviation. Variance is harder to interpret than SD, but has simpler algebraic properties. Notice that $E(X)$, $Var(X)$, and $SD(X)$ are all determined by the distribution of X. That is to say, if two random variables have the same distribution, then they have the same mean, variance, and SD. So we may speak of the mean, variance, and SD of a distribution rather than a random variable.

Parameters of a normal curve. If a histogram displaying the distribution of X follows an approximately normal curve, the curve will be centered near the mean $E(X)$, and $SD(X)$ will be approximately the distance between the center of the curve and its shoulders, where the curve switches from being concave to convex.

The **Normal Distribution** was used as an approximation to the distribution of a sum or average of a large number of independent random variables.

The idea there was to approximate a discrete distribution of many small individual probabilities by scaling the histogram to make it follow a continuous curve.

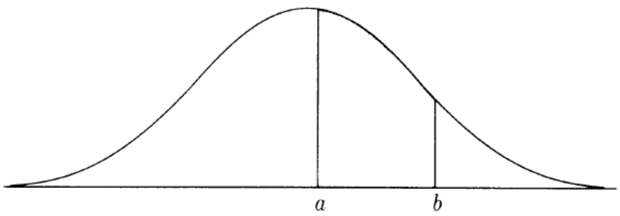
The function defining such a curve is called a ***probability density,* denoted *f(x)*** here.

This function ***determines probabilities over an infinite continuous range of possible values***.

The basic idea is that probabilities are defined by areas under the graph of $f(x)$. That is, a random variable X has density $f(x)$ if for all $a \leq b$

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

which is the area shaded in the following diagram:



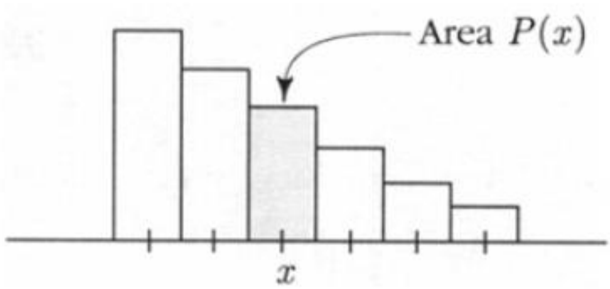
The PDF is used to specify the probability of the random variable falling *within a particular range of values*, as opposed to taking on any one value.

This probability is given by the integral of this variable's PDF over that range—that is, it is given by the area under the density function, but above the horizontal axis and between the lowest and greatest values of the range.

The probability density function is nonnegative everywhere, and its integral over the entire space is equal to 1.

Discrete Distributions

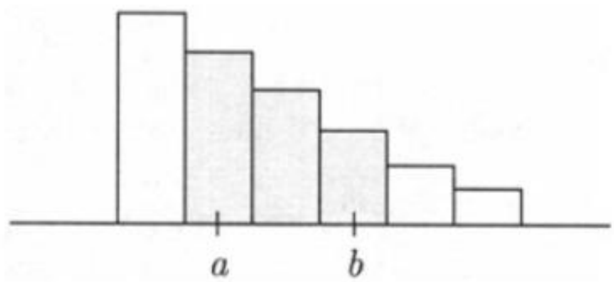
Point Probability:



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

So $P(x)$ is the probability that X has integer value x.

Interval Probability:



$$P(a \leq X \leq b) = \sum_{a \leq x \leq b} P(x)$$

the relative area under a histogram between $a - 1/2$ and $b + 1/2$.