Random variables and probability distributions: discrete and continuous explained

Before discussing further on the topics of binomial and Poisson distributions, we should get our mind clear about the differences between discrete and continuous random variables.

In carrying out any experiment, we generate test results which are basically random by nature. Hence, a random variable provides us a means for describing the experimental outcomes using numerical values. In other words, random variables must assume numerical data for statistical interpretation. The probability distribution for a random variable describes how probabilities are distributed over the values of the random variables. This probability distribution is typically defined in terms of probability density function (pdf) when we refer to the continuous random variables..

A random variable can be classified as being either discrete or continuous depending on the numerical values it assumes.

Discrete random variables

A random variable that may assume either a *finite* number of values or an infinite sequence of values such as 0, 1, 2 …is referred to as a discrete random variable. For example, if we flip a coin and count the number of heads, the number of heads can be any integer value between 0 and plus infinity. But, the outcome is in integer and cannot have an result say 3.5 heads. Hence, the number of heads must be a discrete variable.

Now if we flip a coin twice, we can have four possible outcomes: HH, TT, HT, TH.   The random variable representing the number of heads that results from this statistical experiment can only take on the values of 0, 1, 2, 3 …and so it is a discrete random variable.

The probability distribution for this statistical experiment is shown in the following table:

Number of Heads	Probability
0	0.25
1	0.50
2	0.25

This is an example of a discrete probability distribution which relates each value of a discrete random variable with its probability of occurrence.

The discrete pdf is the probability that the random variable takes the value of x in the form of function $f(x)$:

$$f(x) = Pr[X=x]$$

Therefore, the pdf provides the probability for each value of the random variable. An important advantage of defining a random variable and its probability distribution is that once the probability distribution is known, it will be relatively easy to determine the probability of a variety of events that may be of interest to a decision maker.

In the development of a probability function for any discrete random variable, the following two conditions must be satisfied:

$$f(x) \geq 0$$
$$\Sigma f(x) = 1$$

For example, suppose we roll a dice once in an experiment, the chance (or probability) to have a dot number 1 to 6 faced upwards is 1/6 because there are n = 6 possible values. Hence, the probability function for this discrete uniform random variable is $f(x)$ = 1/6 and $\Sigma f(x) = 1$. In general, we can define the discrete uniform probability function as $f(x) = 1/n$ where n = the number of values the random variable may assume.

Continuous random variables

A random variable that may assume *any* numerical value in an interval or collection of intervals is called a continuous random variable. Any experimental outcomes based on measurement scales such as concentration, weight, temperature, *etc*. can be described by continuous random variables. Examples are aplenty for any laboratory experiments. For example, in an experiment of measuring the temperature of discharged waste gas through a stack, the possible values measured can be in the range of say, 150°C $\leq x \leq$ 200°C. Hence, an infinite number of values are possible for x, including values such as 162.5°C, 175.0°C, 183.33°C, and so on. Hence, we see that it describes a random variable that may assume any value in an interval of values.

A fundamental difference separates discrete and continuous random variables in terms of how probabilities are computed.

For a discrete random variable, the probability function $f(x)$ provides the probability that the random variable assumes a particular value. With continuous random variables, the counterpart of the probability function is the probability density function (pdf), also denoted as $f(x)$.   The difference is that this pdf does not directly provide probabilities. Instead, the area under the graph of $f(x)$ corresponding to a given interval can be proved to provide the probability that the continuous random variable x assumes a value in that interval. Therefore, when we compute probabilities for continuous random variables, we are actually computing the probability that the random variable assumes any value in an interval.

Once a pdf $f(x)$ is identified, the probability for x to take a value between some lower value x_1 and some higher value x_2 can be found by computing the area under the graph of $f(x)$ over the interval from x_1 to x_2.

One easy way to distinguish whether a random variable is discrete or continuous is to think of the values of the random variable as points on a line segment. Choose two points representing values of the random variable. If the entire line segment between the two points also represent possible values for the random variable, then this random variable is continuous.