(Def^n)

Random Variable :

A variable, whose values are any definite numbers or quantities that arise as a result of chance factors such that they cannot be exactly be predicted in advance, is called a random variable.

A random variable is a real-valued function defined over a sample space.

Example :

A school consists of 7 teachers of whom 4 are males and 3 are female. A committee of 2 teachers is to be formed. If Y stands for the number of male teachers selected, then Y is a random variable assuming the values 0, 1 and 2. The possible outcomes and the values of the random variable Y are:

Events	Sequence of events	$Y = y$
E_1	Male, male	2
E_2	Male, female	1
E_3	Female, female	0

Types of random variable :

A random variable may be classified as either discrete or continuous depending upon the specific numerical values it can take on

Discrete random variable : (Def^n)

A random variable defined over a discrete sample space (i.e., that may only take on a finite ~~or~~ countable number of different isolated values) is referred to as a discrete random variable.

Example :

(i) Number of telephone calls received in a telephone booth;

(ii) No. of correct answers in 100- MCQ type examination;

(iii) No. of defective bulbs produced during a day's run;

(iv) No. of bubbles in a glass;

(v) No. of under-5 children in a family

Continuous random variable: (Defn)

A random variable defined over a continuous Sample space (i.e., which may take on any value in a certain interval or collection of intervals), is referred to as a continuous random variable.

Example:

(i) Time taken to serve a customer in a bank counter;

(ii) Weight of a six-month old baby;

(iii) Rate of interest offered by a commercial bank;

(iv) Longevity of an electric bulb;

(v) Temperature recorded by the meteorological dept.

(vi) Height, Weight, Time, exam score, temp. etc.

Probability Distribution:

The idea of a probability distribution exactly parallels that of a frequency distribution. Each type of distribution is based on a set of mutually exclusive and exhaustive measurement classes or class intervals.

In a frequency distribution, each measurement class or class interval is paired with a frequency, while in the probability distribution each class or class interval is paired with its probability. Just as the sum of the frequencies must be equal to N (Total number of cases) in a frequency distribution, so must the sum of the probabilities be 1.0 in a probability distribution.

However, there is an important difference between frequency distributions and probability distributions. The latter almost invariably represents hypothetical or theoretical distributions.

A probability distribution is thus an idealization of the way things might be if we only had all the information. However, the frequency distribution usually represents what has actually been seen to be true from some limited number of observations.

The probability distribution dictates what we should expect to observe in a frequency distribution, if some given state of affairs is true.

Thus, if theoretically the random variable X has a probability of 0.30 of taking on a value in the interval between 15-20 (say) then given a frequency distribution of 100 sample observations, we should expect the class interval 15-20 to contain $100 \times 0.30 = 30$ cases. This does not mean that this will necessarily happen for any given sample; we might observe any frequency between 15 and 20 in this particular interval. Nevertheless in the long run, if we sample indefinitely, 30% of all cases sampled should show ~~cases~~ scores in that interval, provided that our hypothetical probability distribution is true.

** Thus a probability distribution is a long-run behavior of a frequency distribution.

Def^n: (Probability Distribution)

Any statement of a function associating each of a set of mutually exclusive and collectively exhaustive classes or class intervals with its probability is a probability distribution.

* A probability distribution will be either discrete or continuous according as the random variable is discrete or continuous.

Discrete Probability Distribution :

A discrete random variable assumes each of its values or numbers with a certain probability. Sometimes these numbers are equally likely, but sometimes some of the possible values of the random variable are more likely to occur than others.

In either case, the function that gives the probability associated with each possible value of the random variable will result in a probability distribution of the random variable.

For example, if a coin is tossed three times, the sample space will consist of eight possible outcomes, as follows:

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

If X denotes the number of heads obtained, then x, by definition, is a discrete random variable. The possible value x of the random variable X and their associated probabilities can be presented in a tabular form as follows:

Values of $X : x$	0	1	2	3
$P(X=x)$	$1/8$	$3/8$	$3/8$	$1/8$

Note that in the above table, each value of X is paired with its probability, and thus the table represents a probability distribution. Since X is discrete, the resulting distribution is a discrete probability distribution.

Further, note that the values of x exhaust all possible cases and the probabilities add up to 1. Also, the probability for any value of X is directly obtained from the above table.

For example, the probability of obtaining exactly 2 heads is the value associated with $X=2$, which is $3/8$.

Symbolically, $P(X=2) = 3/8$

In general, $P(X=x) = f(x)$; for $x = 0, 1, 2, 3$

Scanned with CamScanner

Frequently, it is convenient to represent all the probabilities of a random variable X by a mathematical formula. Such a formula would necessarily be a function of the numerical values x, and we would like to denote it by $f(x)$, $g(x)$, $h(x)$ and so forth. Hence, we can write,

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

In particular

$$f(3) = P(X=3)$$

It is easy to verify that $f(x)$ is the above table can be represented by

$$f(x) = {}^{3}C_{x} \left(\frac{1}{2}\right)^{x} \left(\frac{1}{2}\right)^{3-x}; \text{ for } x = 0,1,2,3$$

In the foregoing example, the set of ordered pairs, each of the form

$\{$ Number of heads (X), probability of $X \}$

or $\{ x, P(X=x) \}$ is called the probability distribution or the Probability Mass Function

(PMF) of the discrete random variable X.

Def^n:

If a random variable X has a discrete distribution, the probability distribution of X is defined as the function "f" such that for any real number x, $f(x) = P(X=x)$.

It is important to emphasize that all functions are not probability mass functions. The function $f(x)$ defined above must satisfy the following conditions in order to be a PMF:

1. $f(x) \geq 0$
2. $\sum_x f(x) = 1$
3. $P(X=x) = f(x)$

Example:

Verify whether the following functions are PMFs or not. (a) $f(x) = \frac{2x-1}{8}$; $x = 0, 1, 2, 3$

Solⁿ:

(b) $f(x) = \frac{x+1}{16}$; $x = 0, 1, 2, 3$

(c) $f(x) = \frac{3x+6}{21}$; $x = 1, 2$

(a) Given that,

$$f(x) = \frac{2x-1}{8} ; \quad x = 0, 1, 2, 3$$

$X : x$	0	1	2	3
$P(X=x) = f(x)$	$-1/8$	$1/8$	$3/8$	$5/8$

Summing up the function over the entire range of X

$$\sum_{x=0}^{3} f(x) = f(0) + f(1) + f(2) + f(3)$$

$$= -1/8 + 1/8 + 3/8 + 5/8$$

$$= 1$$

Since, $P(X=0) = f(0) = -1/8 < 0$, it contradicts the condition (1).

Hence, $f(x)$ is not a PMF.

(b) Given that,

$$f(x) = \frac{x+1}{16} \;;\; x = 0, 1, 2, 3$$

$X : x$	0	1	2	3
$P(X=x) = f(x)$	$1/16$	$1/8$	$3/16$	$1/4$

$f(x) \geq 0 \quad \forall x$

Summing up the function over the entire range of X.

$$\sum_{x=0}^{3} f(x) = f(0) + f(1) + f(2) + f(3)$$

$$= 1/16 + 1/8 + 3/16 + 1/4$$

$$= \frac{1+2+3+4}{16}$$

$$= \frac{10}{16} \neq \underline{1}$$

Hence, $f(x)$ fails to satisfy condition (2), and therefore is not a PMF although for all values of x, $f(x) \geq 0$.

(c) Given that,

$$f(x) = \frac{3x+6}{21} ; \quad x = 1, 2$$

$X : x$	1	2
$P(X=x) = f(x)$	$9/21$	$12/21$

For all x, $f(x) \geq 0$

Summing up the function over the entire range of X,

$$\sum_{x=1}^{2} f(x) = f(1) + f(2) = \frac{9}{21} + \frac{12}{21} = 1$$

Hence, $f(x)$ satisfies all the conditions of being a PMF.

Thus, $f(x)$ represents a PMF.

Scanned with CamScanner

Example: Obtain the probability function for the School-teacher problem (discussed before). If the random variable is the number of male teachers. (Total = 7; Male = 4; Female = 3; Post = 2)

Solⁿ: Here, the random variable X is the number of males.

Events	Sequence of events	$X = x$
E_1	Male, male	2
E_2	Male, female	1
E_3	Female, female	0

From the above table, we see that X takes on values 0, 1, 2.

Now,

$$f(0) = P(X=0) = \frac{{}^3C_2}{{}^7C_2} = 3/21$$

$$f(1) = P(X=1) = \frac{{}^4C_1 \times {}^3C_1}{{}^7C_2} = \frac{12}{21}$$

$$f(2) = P(X=2) = \frac{{}^4C_2}{{}^7C_2} = 6/21$$

Hence, the probability distribution of the r.v X (no. of males) is represented as

$X : x$	0	1	2
$P(X=x)=f(x)$	$3/21$	$12/21$	$6/21$

Since, for all x $f(x) \geq 0$

and $\sum_{x=0}^{2} f(x) = f(0) + f(1) + f(2)$

$= 3/21 + 12/21 + 6/21$

$= \underline{1}$

Hence, $f(x)$ satisfies all required conditions of being a PMF.

Thus, $f(x)$ is a PMF.

—o—

$$\boxed{f(x) = \frac{{}^{4}C_{x}\ {}^{3}C_{2-x}}{{}^{7}C_{2}} \ ; \ x = 0, 1, 2}$$

এভাবে লিখা যায়

Example :

A bag contains 10 balls of which 4 are black. If 3 balls are drawn at random without replacement, obtain the probability distribution for the number of black balls drawn.

Sol^n : Let, the random variable X denotes the number of black balls drawn.

Clearly, X can assume values 0, 1, 2, and 3.

To obtain the probability distribution of X, we need to compute the probabilities associated with 0, 1, 2, and 3.

Since, 3 balls are to be chosen, the number of ways in which this choice can be made is ${}^{10}C_3$.

Thus,

$$f(0) = P(X=0) = \frac{{}^4C_0 \times {}^6C_3}{{}^{10}C_3} = \frac{20}{120}$$

$$f(1) = P(X=1) = \frac{{}^4C_1 \times {}^6C_2}{{}^{10}C_3} = \frac{60}{120}$$

$$f(2) = P(X=2) = \frac{{}^4C_2 \times {}^6C_1}{{}^{10}C_3} = \frac{36}{120}$$

$$f(3) = P(X=3) = \frac{{}^4C_3 \times {}^6C_0}{{}^{10}C_3} = \frac{4}{120}$$

Hence, the tabular form of the probability distribution of X will be as follows:

$X : x$	0	1	2	3
$P(X=x)=f(x)$	$\frac{20}{120}$	$\frac{60}{120}$	$\frac{36}{120}$	$\frac{4}{120}$

Now, if we wish to check whether the distribution is a PMF, then we have to follow

$f(x) \geq 0$ for all x

$$\sum_{x=0}^{3} f(x) = f(0) + f(1) + f(2) + f(3)$$

$$= \frac{20}{120} + \frac{60}{120} + \frac{36}{120} + \frac{4}{120}$$

$$= \underline{1}$$

Thus, $f(x)$ is a PMF.

General form $\longrightarrow$ $\boxed{f(x) = \frac{{}^4C_x \cdot {}^6C_{3-x}}{{}^{10}C_3}}$; $x = 0, 1, 2, 3$

Example: The probability function of a discrete random variable X is given by

$$f(x) = \alpha \left(\frac{3}{4}\right)^x ; \quad x = 0, 1, 2, 3, \ldots, \infty$$

$$= 0 \qquad \text{elsewhere}$$

Evaluate α and find $P(X \leq 3)$.

Soln: Since, $f(x)$ is a probability function,

$$\sum_{x=0}^{\infty} f(x) = 1$$

But,

$$f(0) = \alpha \cdot \left(\frac{3}{4}\right)^0 = \alpha$$

$$f(1) = \alpha \cdot \left(\frac{3}{4}\right) = \alpha\left(\frac{3}{4}\right)$$

$$f(2) = \alpha \cdot \left(\frac{3}{4}\right)^2$$

$$f(3) = \alpha \cdot \left(\frac{3}{4}\right)^3 \quad \text{and so on.}$$

Hence,

$$\sum_{x=0}^{\infty} f(x) = \alpha + \alpha\left(\frac{3}{4}\right) + \alpha\left(\frac{3}{4}\right)^2 + \alpha\left(\frac{3}{4}\right)^3 + \cdots$$

$$= \alpha\left(1 + \frac{3}{4} + \left(\frac{3}{4}\right)^2 + \left(\frac{3}{4}\right)^3 + \cdots\right)$$

$$= \alpha . \left[\frac{1}{1 - 3/4} \right] \qquad \left[\begin{array}{l} 1 + x + x^2 + x^3 + \dots \\ \frac{a}{1-r} \Big| \begin{array}{l} a = 1 \\ r = x \end{array} \end{array} \right]$$

$$= \alpha . \left[\frac{1}{1/4} \right]$$

$$= 4\alpha$$

$$\Rightarrow \sum_{x=0}^{\infty} f(x) = 4\alpha = 1$$

$$\Rightarrow \alpha = 1/4$$

Hence, the complete PMF of X is

$$f(x) = \frac{1}{4} \left(\frac{3}{4}\right)^x ; \qquad x = 0, 1, 2, 3, \dots, \infty$$

Also,

$$P(X \leq 3) = f(0) + f(1) + f(2) + f(3)$$

$$= \frac{1}{4} + \frac{1}{4}\left(\frac{3}{4}\right) + \frac{1}{4}\left(\frac{3}{4}\right)^2 + \frac{1}{4}\left(\frac{3}{4}\right)^3$$

$$= \frac{1}{4}\left[1 + 3/4 + \frac{9}{16} + \frac{27}{64} \right]$$

$$= \frac{175}{256} \quad \underline{Ans}$$

Example : The discrete random variable Y has a probability distribution is shown below:

$Y : y$	-3	-2	-1	0	1
$P(Y=y)=f(y)$	0.10	0.25	0.30	0.15	K

find (i) K ; (ii) $P(-3<Y<0)$

(iii) $P(Y \geq -1)$

Sol^n : (i) Since $f(y)$ is a probability distribution, it must satisfy the condition that

$\cancel{\sum_{y} f(y)=1}$. $\sum_{y} f(y)=1$

Therefore,

$$f(-3)+f(-2)+f(-1)+f(0)+f(1)=1$$

$$\Rightarrow 0.10+0.25+0.30+0.15+K=1$$

$$\Rightarrow K=0.20$$

$$\therefore f(1)=0.20$$

(ii) $P(-3<Y<0) = f(-2)+f(1)$

$= 0.25+0.30$

$= 0.55$

(iii) $P(Y \geq -1) = f(-1) + f(0) + f(1)$

$= 0.30 + 0.15 + 0.20$

$= 0.65$

(Ans)

Scanned with CamScanner

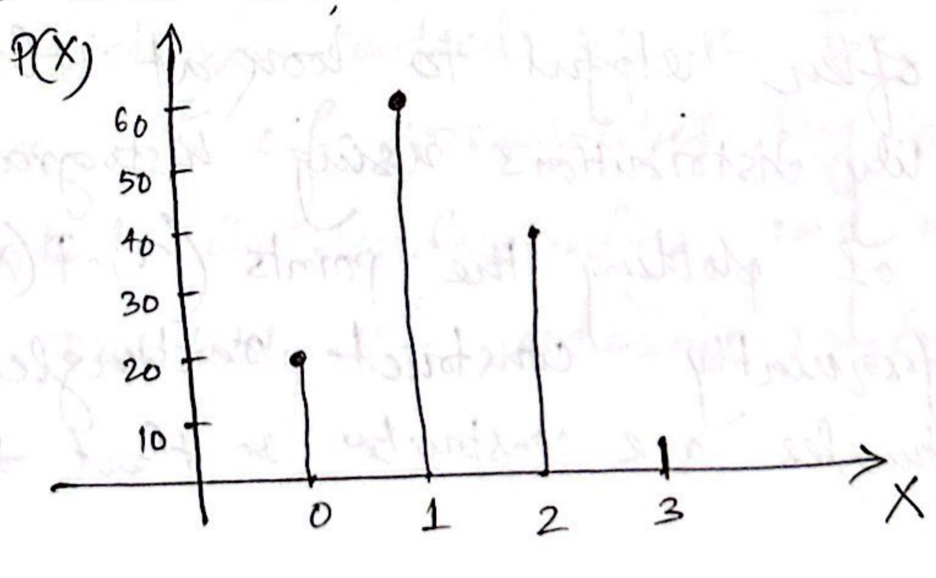
Graphical Representation of a Discrete Probability Distribution :

In addition to specifying a PMF by rule or formula, one can also employ graph. It is often helpful to represent a probability function in graphical form.

For example, let us consider a random variable X with its probability distribution presented in the following tabular form:

$X : x$	0	1	2	3
$P(X=x)=f(x)$	$20/120$	$60/120$	$36/120$	$4/120$

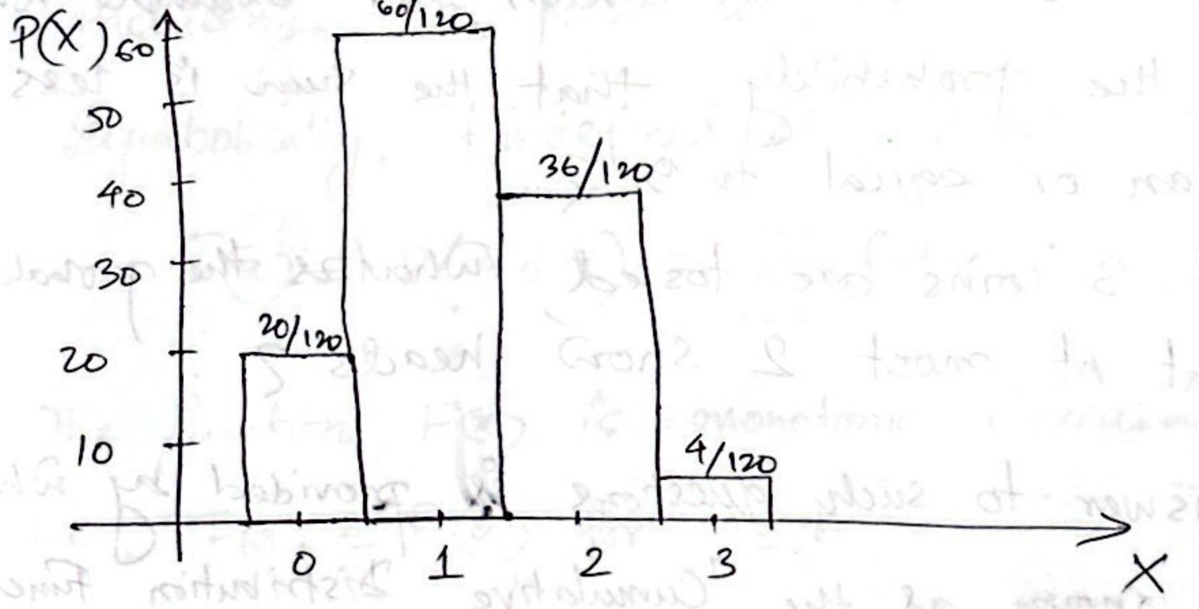
The graphical representation of the abovementioned table would be,



The X-axis of the graph is marked off with different values that the random variable can take on. The Y-axis is a measure of units of probability. Then above each possible value x of X, a vertical line is raised to the height corresponding to the probability $P(X=x)$. It is, however, to be remembered that in many instances, particularly in theoretical statistics, it is much more convenient to specify the distribution of a discrete random variable by its rule (represented in the form of an algebraic formula) rather than by listing in a tabular form or by a graph.

It is often helpful to look at the probability distributions using histograms. Instead of plotting the points $(x, P(X=x))$, we more frequently construct rectangles. Here the rectangles are constructed so that their bases of equal width are centred at each value x and their heights are equal to the corresponding probabilities given by $P(X=x)$.

The bases are constructed so as to leave no space between the rectangles.



Since each base in such histogram has unit width, the $P(X=x)$ is equal to the area of the rectangle centred at x. The concept of representation of probabilities by area is important in understanding and interpreting the probability of continuous random variable.

Discrete Distribution Function:

In many occasions, we are interested to know the probability that a random variable takes on a value less than or equal to a prescribed number x_1, say.

If two dice are thrown, for example, what is the probability that the sum is less than or equal to 5?

If 3 coins are tossed, what is the probability that at most 2 show heads?

Answer to such questions is provided by what is known as the "Cumulative Distribution Function", which applies to both continuous and discrete variables.

Def^n: (Cumulative Distribution Function)

The cumulative distribution function (CDF) or simply the distribution function $F(x)$ of a discrete random variable X with probability function $f(x)$ defined over all real numbers x is the cumulative probability upto and including the point x.

Symbolically, it is defined as follows:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

The function $F(x)$ is a monotonic increasing function.

i.e., $F(a) \leq F(b)$ for $a \leq b$.

And the limit of F to the left is 0 and to the right is 1.

$$\left.\begin{aligned} &\lim_{x \to -\infty} F(x) = 0 \\ \& \; &\lim_{x \to +\infty} F(x) = 1 \end{aligned}\right\}$$

The value of $F(x)$ at any point must be a number in the interval $0 \leq F(x) \leq 1$, because $F(x)$ is a probability of the event $(X \leq x)$.

Example:

Consider the probability distribution of the random variable X as follows.

$X: x$	0	1	2	3
$P(X=x)=f(x)$	$20/120$	$60/120$	$36/120$	$4/120$

Here,

$$F(0) = P(X=0) = f(0) = \frac{20}{120}$$

$$F(1) = P(X \le 1) = f(0) + f(1) = F(0) + f(1)$$

$$\Rightarrow F(1) = \frac{20}{120} + \frac{60}{120} = \frac{80}{120}.$$

$$F(2) = P(X \le 2) = f(0) + f(1) + f(2)$$

$$= F(1) + f(2)$$

$$= \frac{80}{120} + \frac{36}{120}$$

$$= \frac{116}{120}$$

$$F(3) = P(X \leq 3) = f(0) + f(1) + f(2) + f(3)$$

$$= F(2) + f(3)$$

$$= \frac{116}{120} + \frac{4}{120}$$

$$= 1.$$

A tabular representation of $F(x)$ is as follows

$X: x$	<0	0	1	2	3	>3
$P(X=x)=f(x)$	0	$20/120$	$60/120$	$36/120$	$4/120$	0
$F(x)$	0	$20/120$	$80/120$	$116/120$	1	1

A formal way of representing the distribution function $F(x)$ is as follows:

$$F(x) = \begin{cases} 0, & \text{for } x < 0 \\ \frac{20}{120}, & \text{for } 0 \leq x < 1 \\ \frac{80}{120}, & \text{for } 1 \leq x < 2 \\ \frac{116}{120}, & \text{for } 2 \leq x < 3 \\ 1, & \text{for } x \geq 3. \end{cases}$$

Example :

A coin is tossed three times. If X is the random representing the number of heads obtained, find the probability distribution of X and hence obtain $F(x)$.

Sol^n : $S = \{HHH, HHT, HTT, HTH, THT, THH, TTH, TTT\}$

The probability distribution can be presented as

$X : x$	0	1	2	3
$P(X=x)=f(x)$	$1/8$	$3/8$	$3/8$	$1/8$

Therefore,

$$F(0) = f(0) = 1/8$$

$$F(1) = f(0) + f(1) = 4/8$$

$$F(2) = f(2) + F(1) = 3/8 + 4/8 = 7/8$$

$$F(3) = F(2) + f(3) = 7/8 + 1/8 = 1$$

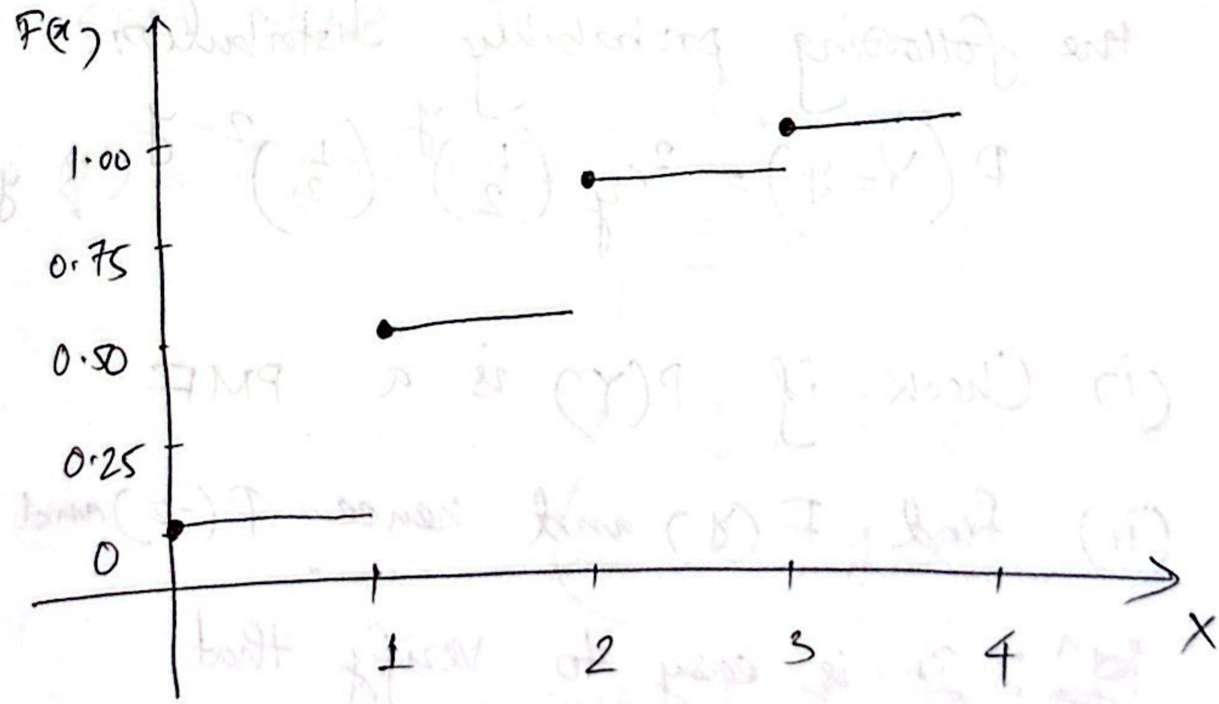
Thus,

$$F(x) = \begin{cases} 0 ; & x < 0 \\ 1/8 ; & 0 \le x < 1 \\ 4/8 , & 1 \le x < 2 \\ 7/8 , & 2 \le x < 3 \\ 1 ; & x \geq 3 \end{cases}$$

In addition to presenting the above probability function in functional form, we can put the above function is tabular form as well.

$X:x$	<0	0	1	2	3	>3
$P(X=x)=f(x)$	0	$1/8$	$3/8$	$3/8$	$1/8$	0
$F(x)$	0	$1/8$	$4/8$	$7/8$	1	1

The distribution function $F(x)$ or CDF can also be shown graphically. A sketch of the above function is shown below:



Note that the graph is of distinctive appearance of a set of steps and for this reason, such a graph is called a "step function"

* Distribution functions for discrete random variables are always step functions because the cumulative distribution function increases only as a countable number of points.

Example:

Suppose a discrete random variable Y has the following probability distribution.

$$P(Y=y) = {}^2C_y \left(\tfrac{1}{2}\right)^y \left(\tfrac{1}{2}\right)^{2-y} \; ; \; y=0,1,2$$

(i) Check if $P(Y)$ is a PMF

(ii) Find $F(y)$ and hence $F(-2)$ and $P(Y \le 1)$

Soln: It is easy to verify that

$$f(0) = P(Y=0) = {}^2C_0 \left(\tfrac{1}{2}\right)^0 \left(\tfrac{1}{2}\right)^2 = 0.25$$

$$f(1) = P(Y=1) = {}^2C_1 \left(\tfrac{1}{2}\right)^1 \left(\tfrac{1}{2}\right)^1 = 0.50$$

$$f(2) = P(Y=2) = {}^2C_2 \left(\tfrac{1}{2}\right)^2 \left(\tfrac{1}{2}\right)^0 = 0.25$$

So, that,

$$\sum_{y=0}^{2} f(y) = \sum_{y=0}^{2} P(Y=y) = f(0) + f(1) + f(2)$$

$$= 0.25 + 0.50 + 0.25$$

$$= 1$$

Also, $f(y) \geq 0, \ \forall \ y = 0, 1, 2$

Hence, $P(Y)$ is a PMF.

The CDF of Y is

$$F(y) = P(Y \leq y) = \begin{cases} 0; & y < 0 \\ \frac{1}{4}; & 0 \leq y < 1 \\ \frac{3}{4}; & 1 \leq y < 2 \\ 1, & y \geq 2 \end{cases}$$

Because the only values of Y that are assigned to positive probabilities are: 0, 1 and 2 and none of these values are less than or equal to -2.

Hence, $F(-2) = 0$.

Also, $P(Y \leq 1) = F(1)$ and thus

$$F(Y=1) = P(Y=0) + P(Y=1)$$

$$= 0.25 + 0.50$$

$$= 0.75$$