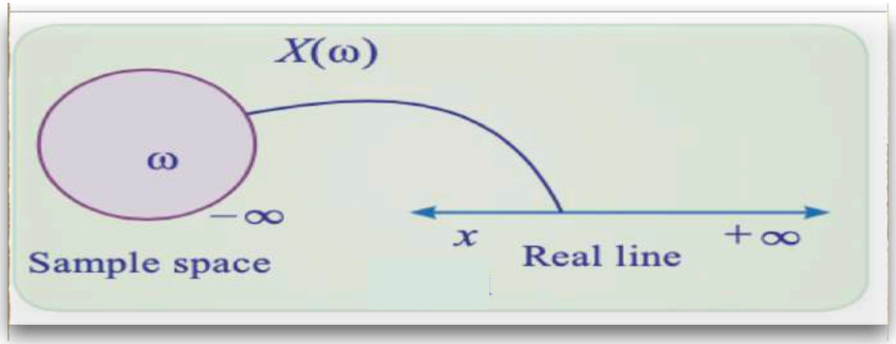
Ramdom Variable

Let S be the sample space of a random experiment. A rule that assigns a single real number to each outcome (sample point) of the random experiment is called random variable.



In other words, a random variable is a real valued function defined on a sample space S that is with each outcome ω of a random experiment there corresponds a unique real value x known as a value of the random variable X. That is $X(\omega) = x$.

Example

Consider the random experiment of rolling a die.
The sample space of the experiment is

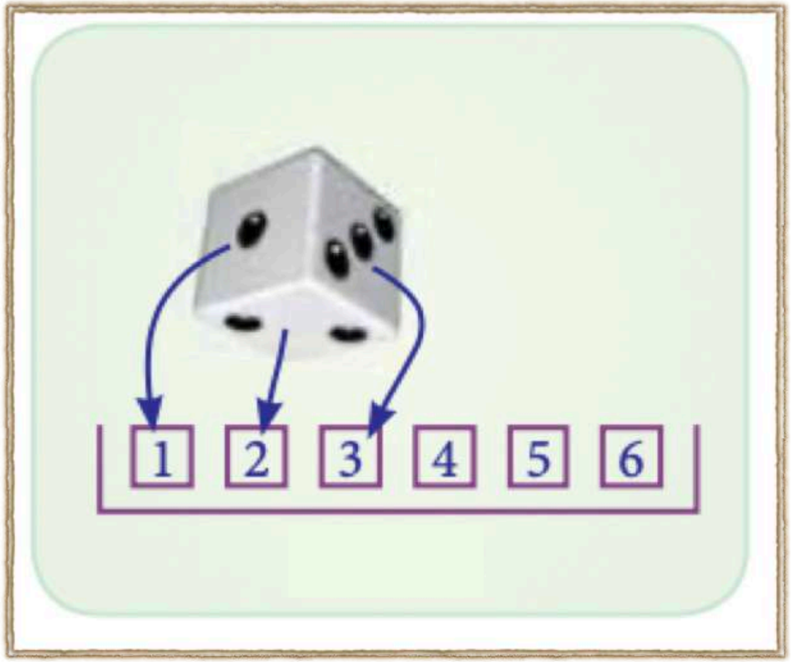
$$S=\{1, 2, 3, 4, 5, 6\}$$

Let X denotes the face of the die appears on top.

The assigning rule is

$X(1) = 1$, $X(2) = 2$, $X(3) = 3$, $X(4)=4$, $X(5)=5$ and $X(6)=6$

Hence the values taken by the random variable X are 1,2,3,4,5,6. These values are also called the realization of the random variable X.



Example

Random experiment : Two coins are tossed simultaneously.

Sample space : $S=\{HH, HT, TH, TT\}$

Assigning rule : Let X be a random variable defined as the number of heads comes up.

Sample Point ω	HH	HT	TH	TT
X (ω)	2	1	1	0

Here, the random variable X takes the values 0, 1, 2 .

EXERCISE

Experiment : Two dice are rolled simultaneously.

Sample space :

Assigning rule : Let X denote the sum of the numbers on the faces of dice

Discrete and Continuous random variables

A random variable is said to be discrete if it takes only a finite or countable infinite number of values.

Example

Consider the experiment of tossing a coin

If X (Head) $=1$, X (Tail) $= 0$

Then X takes the values either 0 or 1 This is a discrete random variable.

Example

Consider the experiment of tossing a coin till head appears.

Let random variable X denote the number of trials needed to get a head. The values taken by it will be $1, 2, 3, ..$

It is discrete random variable taking countable infinite values.

A random variable X is said to be continuous, if it takes values in an interval or union of disjoint intervals.

Example

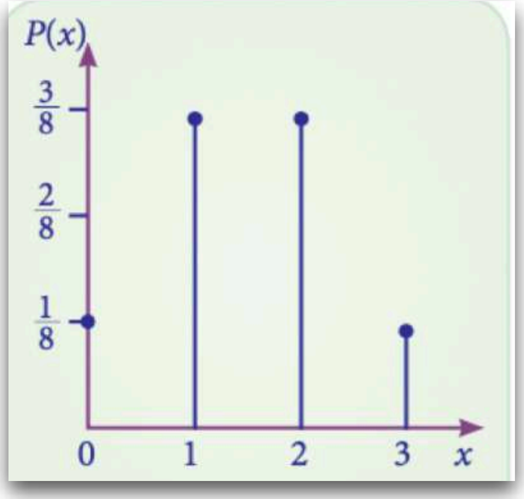
If X is defined as the height of students in a school ranging between 120 cms and 180 cms, Then the random variable X is $\{x \,/\, 120\ cms < x < 180 \text{ cms} \}$ is a continuous random variable.

Example

Let the maximum life of electric bulbs is 1500 hrs. Life time of the electric bulb is the continuous random variables and it is written as $X = \{x \,/\, 0 \leq x \leq 1500\}$

Probability mass function and probability density function

A probability function is associated with each value of the random variable. This function is used to compute probabilities for events associated with the random variables.



The probability function defined for a discrete random variable is called probability mass function. The probability function associated with continuous random variable is called probability density function.

The set of ordered pairs (x, f (x)) is a probability function, probability mass function, or probability distribution of the discrete random variable X if, for each possible outcome x,

1. $f(x) \geq 0$

2. $\sum_{x} f(x) = 1$

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

The cumulative distribution function F (x) of a discrete random variable X with probability distribution f(x) is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \text{ for } -\infty < x < \infty$$

Example: Find a formula for the probability distribution of the random variable X representing the outcome when a single die is rolled once.

- Let X be a random variable whose values x are the possible outcomes when a single die is rolled once. x can take the numbers 1, 2, 3, 4, 5 and 6.
- All possible outcomes are equally likely, i.e. each outcome occurs with probability $1/6$.

Hence, the probability distribution of the random variable X is

x	1	2	3	4	5	6
$f(x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

Example: A shipment of 7 television sets contains 2 de- fective sets. A hotel makes a random purchase of 3 of the sets. If x is the number of defective sets pur- chased by the hotel, find the probability distribution of X. Express the results graphically as a probability histogram.

In order to find the probability distribution of X, we need to find the probability of X being each of the possible values.

Since there are 2 defective sets and the hotel is purchasing 3 sets, <u>X can take values 0, 1 and 2.</u>

(i) If **X=0** then all 3 purchased sets are among the 5 non-defective television sets so:

$$P(X = 0) = \frac{\binom{5}{3}}{\binom{7}{3}} = \frac{10}{35} = \boxed{\frac{2}{7}}$$

(ii) If **X=1** then 2 of 3 purchased sets are among the 5 non-defective television sets, while 1 purchased set is among the 2 defective sets so:

$$P(X = 1) = \frac{\binom{5}{2} \cdot \binom{2}{1}}{\binom{7}{3}} = \frac{10 \cdot 2}{35} = \boxed{\frac{4}{7}}$$

(iii) If **X=2** then 2 of 3 purchased sets are exactly the 2 defective sets, while the remaining one is non-defective so:

$$P(X=1) = \frac{\binom{5}{1} \cdot \binom{2}{2}}{\binom{7}{3}} = \frac{5 \cdot 1}{35} = \boxed{\frac{1}{7}}$$

The denominator in all 3 cases is $\binom{7}{3}$ because that is the number of distinct ways in which 3 television sets can be purchased among 7.

Therefore, we obtain the probability distribution:

x	0	1	2
f(x)	$\frac{2}{7}$	$\frac{4}{7}$	$\frac{1}{7}$

The probability distribution of X, the number of imperfections per 10 meters of a synthetic fabric in continuous rolls of uniform width, is given by

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Construct the cumulative distribution function of X.

The number of imperfections is a **discrete** random variable X.

We know that for a discrete random variable, its cumulative distribution function is defined by:

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

$$\mathbf{F(0)} = \sum_{t \le 0} f(t) = f(0) = \boxed{0.41}$$

$$\mathbf{F(1)} = \sum_{t \le 1} f(t) = f(0) + f(1) = 0.41 + 0.37 = \boxed{0.78}$$

$$\mathbf{F(2)} = \sum_{t \le 2} f(t) = f(0) + f(1) + f(2) = 0.41 + 0.37 + 0.16 = \boxed{0.94}$$

$$\mathbf{F(3)} = \sum_{t \le 3} f(t) = f(0)+f(1)+f(2)+f(3) = 0.41+0.37+0.16+0.05 = \boxed{0.99}$$

$$\mathbf{F(4)} = \sum_{t \le 4} f(t) = f(0)+f(1)+f(2)+f(3)+f(4) = 0.41+0.37+0.16+0.05+0.01 = \boxed{1}$$

Therefore, we obtain the cumulative distribution:

x	0	1	2	3	4
F(x)	0.41	0.78	0.94	0.99	1

Example: Find the probability distribution for the number of jazz CDs when 4 CDs are selected at random from a collection consisting of 5 jazz CDs, 2 classical CDs, and 3 rock CDs. Express your results by means of a formula.

Let X be a random variable which represents the number of jazz CDs among the 4 selected CDs from the collection.

Since there are 5 jazz CDs in the collection of 10 CDs and we're selecting some 4 CDs, the variable X can assume values 0,1,2,3 and 4.

The **total** number of choices that could be made is:

$$\binom{10}{4}$$

since we are choosing some 4 CDs from the collection of 10.

Further on, if our choice contains x jazz CDs, then it also contains $4 - x$ non-jazz CDs, for every $x \in \{0, 1, 2, 3, 4\}$.

The x jazz CDs could have been selected in $\binom{\mathbf{5}}{\mathbf{x}}$ ways, since there are 5 jazz CDs in the collection. Similarly, the $4 - x$ non-jazz CDs could have been selected in $\binom{\mathbf{5}}{\mathbf{4\text{-}x}}$ ways, because there are also 5 non-jazz CDs in the collection.

When divided with the total number of ways to make a selection, we obtain the probability distribution formula:

$$\boxed{P(X = x) = \frac{\binom{5}{x} \cdot \binom{5}{4-x}}{\binom{10}{4}}}$$

The function f(x) is a probability density function (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$ for all $x \in R$

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

3. $P(a < x < b) = \int_{a}^{b} f(x)dx$

The cumulative distribution function F (x) of a continuous random variable X with density function f(x) is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \quad \text{for} \quad -\infty < x < \infty$$

The shelf life, in days, for bottles of a certain prescribed medicine is a random variable having the density function

$$f(x) = \begin{cases} \frac{20,000}{(x+100)^3}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that a bottle of this medicine will have a shell life of

(a) at least 200 days;

(b) anywhere from 80 to 120 days.

(a) Let X be the shelf life for a bottle of the medicine, i.e. the random variable with the density function f.

$$
\begin{aligned}
P(X \geq 200) &= 1 - P(X < 200) \\
&= 1 - \int_{x=0}^{200} f(x) \\
&= 1 - \int_{x=0}^{200} \frac{20000}{(x+100)^3} \\
&= 1 - \left(- \frac{10000}{(x+100)^2} \right) \Big|_0^{200} \\
&= 1 + \frac{10000}{(200+100)^2} - \frac{10000}{(0+100)^2} \\
&= 1 + \frac{10000}{90000} - 1
\end{aligned}
$$

$$
\boxed{P(X \geq 200) = \frac{1}{9}}
$$

(b)

$$
\begin{aligned}
P(80 \leq X \leq 120) &= P(X \leq 120) - P(X < 80) \\
&= \int_{x=0}^{120} f(x) - \int_{x=0}^{80} f(x) \\
&= \int_{x=80}^{120} f(x) \\
&= \int_{x=80}^{120} \frac{20000}{(x+100)^3} \\
&= \left(- \frac{10000}{(x+100)^2} \right) \Bigg|_{80}^{120} \\
&= -\frac{10000}{220^2} - (-\frac{10000}{180^2})
\end{aligned}
$$

$$
\boxed{P(80 \leq X \leq 120) = 0.102}
$$

The waiting time, in hours, between successive speeders spotted by a radar unit is a continuous random variable with cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-8x}, & x \geq 0. \end{cases}$$

Find the probability of waiting less than 12 minutes between successive speeders

(a) using the cumulative distribution function of X;

(b) using the probability density function of X.

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-8x} & x \geq 0 \end{cases}$$

(a) $P(\text{Waiting time less than 12 minutes}) = P(X < 0{\cdot}2) = F(0{\cdot}2) = \underline{\underline{0{\cdot}7981}}$

Convert min to hours

(b) $$f(x) = \frac{d\,F(x)}{dx} = \begin{cases} 0 & x < 0 \\ 8e^{-8x} & x \geq 0 \end{cases}$$

$\therefore$ $P(\text{Waiting time less than 12 minutes}) = P(X < 0{\cdot}2)$

$$= \int_0^{0{\cdot}2} 8e^{-8x}\,dx$$

$$= \frac{8}{8}\left(-e^{-8x}\right)_0^{0{\cdot}2}$$

$$= -\left[e^{-1{\cdot}6} - 1\right]$$

$$= 0{\cdot}7981$$

3.29 An important factor in solid missile fuel is the particle size distribution. Significant problems occur if the particle sizes are too large. From production data in the past, it has been determined that the particle size (in micrometers) distribution is characterized by

$$f(x) = \begin{cases} 3x^{-4}, & x > 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that this is a valid density function.

(b) Evaluate $F(x)$.

(c) What is the probability that a random particle from the manufactured fuel exceeds 4 micrometers?

3.29 Given $f(x) = \begin{cases} 3x^{-4} & x > 1 \\ 0 & \text{elsewhere.} \end{cases}$

a) For $f(x)$ to be valid, $\int_1^\infty f(x)\,dx = 1.$

$$\int_1^\infty 3x^{-4}\,dx \Rightarrow \left[-\frac{1}{x^3}\right]_1^\infty = \underline{\underline{1}}$$

i.e $f(x)$ is a valid density function.

b) $F(x) = \int^x f(x)\,dx.$

$$F(x) = 3\int_1^x x^{-4}\,dx = \left[-\frac{1}{x^3}\right]_1^x = \underline{1 - x^{-3}}$$

c) $P(X > 4) = 1 - P(X \leq 4)$

$$= 1 - \left[-\frac{1}{x^3}\right]_1^4 \Rightarrow 1 + \left[\frac{1}{64} - 1\right].$$

$$= \underline{\underline{0.0156}}$$