Chapter 1
Random Variables and Probability Distributions

1.1 Concept of a Random Variable:

· In a statistical experiment, it is often very important to allocate numerical values to the outcomes.

Example 1:

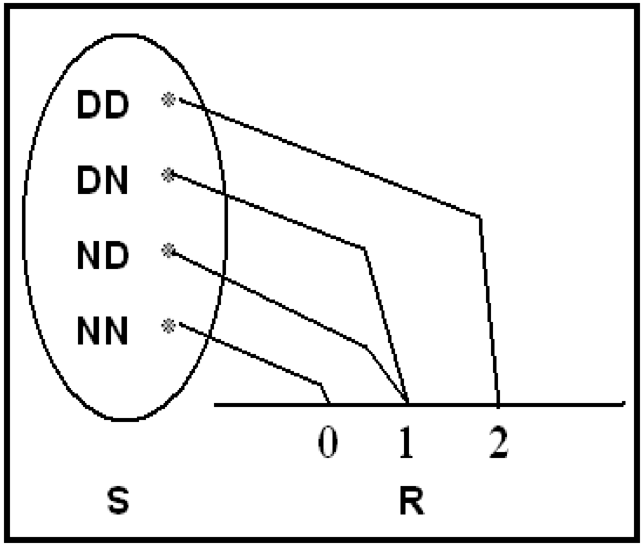
· Experiment: testing two components. (D=defective, N=non-defective)

· Sample space: S={DD,DN,ND,NN}

· Let X = number of defective components when two components are tested.

· Assigned numerical values to the outcomes are:

Sample point (Outcome)	Assigned Numerical Value (x)
DD	2
DN	1
ND	1
NN	0



- Notice that, the set of all possible values of the random variable X is {0, 1, 2}.

Definition 1:

A random variable X is a function that associates each element in the sample space with a real number (i.e., $X : S \rightarrow \mathbf{R}$.)

Notation: " X " denotes the random variable .

" x " denotes a value of the random variable X.

Types of Random Variables:

· A random variable X is called a **discrete** random variable if its set of possible values is countable, i.e.,

.$x \in \{x_1, x_2, \ldots, x_n\}$ or $x \in \{x_1, x_2, \ldots\}$

· A random variable X is called a **continuous** random variable if it can take values on a continuous scale, i.e.,

.$x \in \{x: a < x < b; a, b \in R\}$

· In most practical problems:

○ A discrete random variable represents count data, such as the number of defectives in a sample of *k* items.

○ A continuous random variable represents measured data, such as height.

1.2 Discrete Probability Distributions

- A discrete random variable X assumes each of its values with a certain probability.

Example 2:

- Experiment: tossing a non-balance coin 2 times independently.
- H= head , T=tail
- Sample space: S={HH, HT, TH, TT}
- Suppose P(H)=½P(T) ⇔ P(H)=1/3 and P(T)=2/3
- Let X= number of heads

Sample point (Outcome)	Probability	Value of X (x)
HH	P(HH)=P(H) P(H)=1/3×1/3 = 1/9	2
HT	P(HT)=P(H) P(T)=1/3×2/3 = 2/9	1
TH	P(TH)=P(T) P(H)=2/3×1/3 = 2/9	1
TT	P(TT)=P(T) P(T)=2/3×2/3 = 4/9	0

- The possible values of X are: 0, 1, and 2.
- X is a discrete random variable.
- Define the following events:

Event (X=x)	Probability = P(X=x)
(X=0)={TT}	P(X=0) = P(TT)=4/9
(X=1)={HT,TH}	P(X=1) =P(HT)+P(TH)=2/9+2/9=4/9
(X=2)={HH}	P(X=2) = P(HH)= 1/9

- The possible values of X with their probabilities are:

X	0	1	2	Total
P(X=x)=f(x)	4/9	4/9	1/9	1.00

The function f(x)=P(X=x) is called the probability function (probability distribution) of the discrete random variable X.

Definition 2:

The function f(x) is a probability function of a discrete random variable X if, for each possible values x, we have:

1) $f(x) \geq 0$
2) $\sum_{all\ x} f(x) = 1$
3) $f(x) = P(X=x)$

Note:

$$P(X \in A) = \sum_{\text{all } x \in A} f(x) = \sum_{\text{all } x \in A} P(X = x)$$

Example 3:

For the previous example, we have:

X	0	1	2	Total
f(x)= P(X=x)	4/9	4/9	1/9	$\sum_{x=0}^{2} f(x) = 1$

$P(X<1) = P(X=0)=4/9$

$P(X\leq 1) = P(X=0) + P(X=1) = 4/9+4/9 = 8/9$

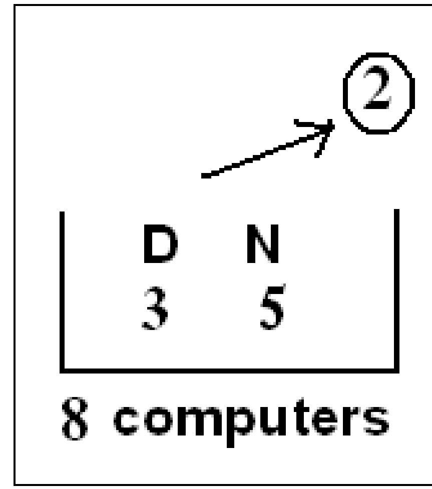
$P(X\geq 0.5) = P(X=1) + P(X=2) = 4/9+1/9 = 5/9$

$P(X>8) = P(\phi) = 0$

$P(X<10) = P(X=0) + P(X=1) + P(X=2) = P(S) = 1$

Example 4:

A shipment of 8 similar microcomputers to a retail outlet contains 3 that are defective and 5 are non-defective.
If a school makes a random purchase of 2 of these computers, find the probability distribution of the number of defectives.



Solution:

We need to find the probability distribution of the random variable: X = the number of defective computers purchased.

Experiment: selecting 2 computers at random out of 8

$n(S) = \binom{8}{2}$ equally likely outcomes

The possible values of X are: x=0, 1, 2.

Consider the events:

$$(X=0)=\{0D \text{ and } 2N\} \Rightarrow n(X=0)=\binom{3}{0}\times\binom{5}{2}$$

$$(X=1)=\{1D \text{ and } 1N\} \Rightarrow n(X=1)=\binom{3}{1}\times\binom{5}{1}$$

$$(X=2)=\{2D \text{ and } 0N\} \Rightarrow n(X=2)=\binom{3}{2}\times\binom{5}{0}$$

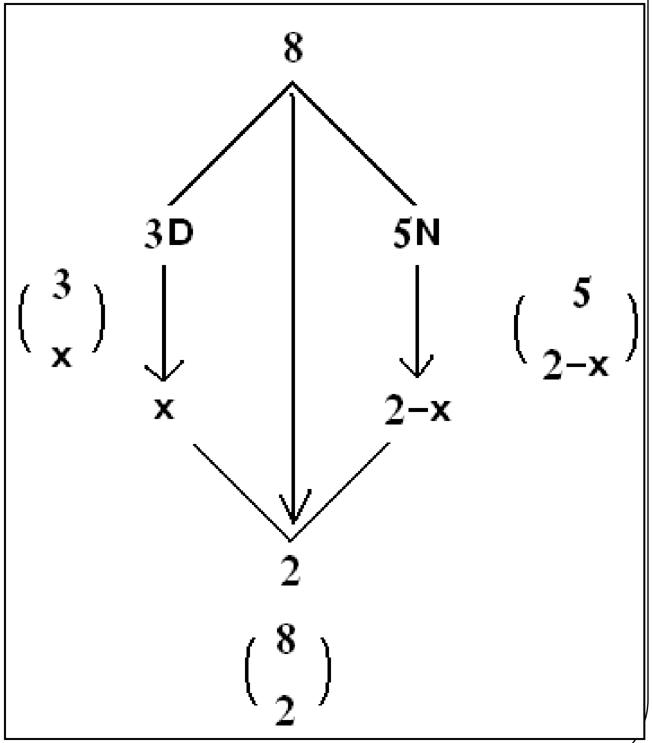
$$f(0)=P(X=0)=\frac{n(X=0)}{n(S)}=\frac{\binom{3}{0}\times\binom{5}{2}}{\binom{8}{2}}=\frac{10}{28}$$

$$f(1) = P(X=1) = \frac{n(X=1)}{n(S)} = \frac{\binom{3}{1}\times\binom{5}{1}}{\binom{8}{2}} = \frac{15}{28}$$

$$f(2) = P(X=2) = \frac{n(X=2)}{n(S)} = \frac{\binom{3}{2}\times\binom{5}{0}}{\binom{8}{2}} = \frac{3}{28}$$

In general, for x=0,1, 2, we have:

$$f(x) = P(X=x) = \frac{n(X=x)}{n(S)} = \frac{\binom{3}{x}\times\binom{5}{2-x}}{\binom{8}{2}}$$



The probability distribution of X is:

x	0	1	2	Total
f(x)= P(X=x)	$\frac{10}{28}$	$\frac{15}{28}$	$\frac{3}{28}$	1.00

$$f(x) = P(X = x) = \begin{cases} \dfrac{\binom{3}{x} \times \binom{5}{2-x}}{\binom{8}{2}}; & x = 0, 1, 2 \\ 0; \ otherwise \end{cases}$$

Hypergeometric Distribution

Cumulative distribution function (CDF), F(x) of discrete R.V.

Definition 3:

The cumulative distribution function (CDF), F(x), of a discrete random variable X with the probability function f(x) is given by:

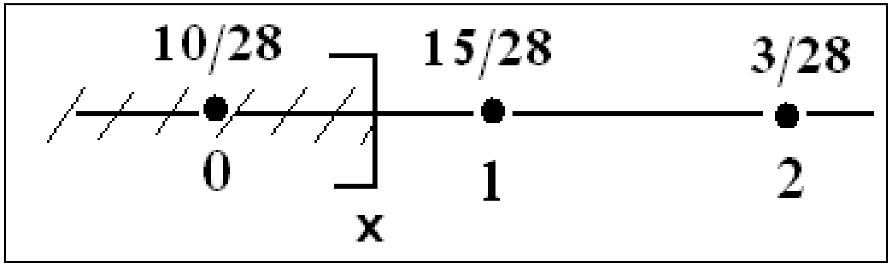
$$F(x) = P(X \le x) = \sum_{t \le x} f(t) = \sum_{t \le x} P(X = t); \quad \text{for } -\infty < x < \infty$$

Example 5:

Find the CDF of the random variable X with the probability function:

X	0	1	2
F(x)	$\frac{10}{28}$	$\frac{15}{28}$	$\frac{3}{28}$

Solution:



$F(x)=P(X\le x)$ for $-\infty<x<\infty$

For $x<0$: $F(x)=0$

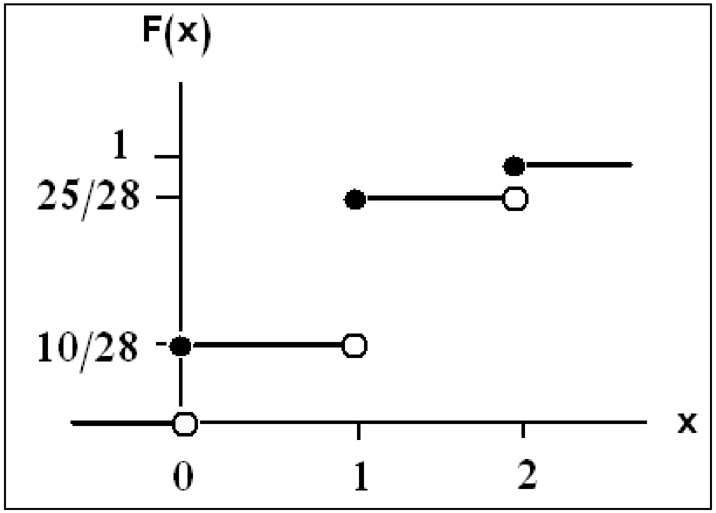
For $0\le x<1$: $F(x)=P(X=0)=\frac{10}{28}$

For $1\le x<2$: $F(x)=P(X=0)+P(X=1)=\frac{10}{28}+\frac{15}{28}=\frac{25}{28}$

For $x\ge 2$: $F(x)=P(X=0)+P(X=1)+P(X=2)=\frac{10}{28}+\frac{15}{28}+\frac{3}{28}=1$

The CDF of the random variable X is:

$$F(x) = P(X \le x) = \begin{cases} 0 \;; & x < 0 \\ \dfrac{10}{28} \;; & 0 \le x < 1 \\ \dfrac{25}{28} \;; & 1 \le x < 2 \\ 1 \;; & x \ge 2 \end{cases}$$



Note:

F(–0.5) = P(X≤–0.5)=0

F(1.5)=P(X≤1.5)=F(1) = $\dfrac{25}{28}$

F(3.8) =P(X≤3.8)=F(2)= 1

Result:

P(a **<** X ≤ b) = P(X ≤ b) – **P(X** ≤ a) = F(b) – **F(a)**

P(a ≤ X ≤ b) = P(a < X ≤ b) + P(X=a) = F(b) – **F(a) + f(a)**

P(a < X < b) = P(a < X ≤ b) – **P(X=b) = F(b) – F(a) – f(b)**

Result:

Suppose that the probability function of X is:

x	x_1	x_2	x_3	$\dots$	x_n
$f(x)$	$f(x_1)$	$f(x_2)$	$f(x_3)$	$\dots$	$f(x_n)$

Where $x_1 < x_2 < \dots < x_n$. Then:

$F(x_i) = f(x_1) + f(x_2) + \dots + f(x_i)$; $i=1, 2, \dots, n$

$F(x_i) = F(x_{i-1}) + f(x_i)$; $i=2, \dots, n$

$f(x_i) = F(x_i) - F(x_{i-1})$

Example 6:

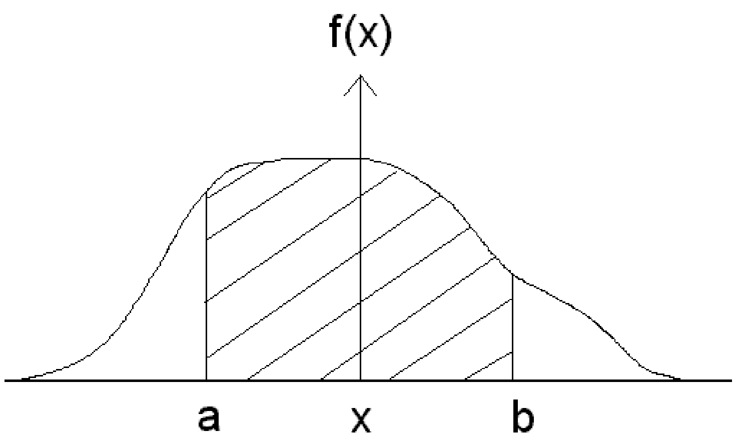
In the previous example,

$P(0.5 < X \leq 1.5) = F(1.5) - F(0.5) = \frac{25}{28} - \frac{10}{28} = \frac{15}{28}$

$P(1 < X \leq 2) = F(2) - F(1) = 1 - \frac{25}{28} = \frac{3}{28}$

1.3. Continuous Probability Distributions

For any continuous random variable, X, there exists a non-negative function f(x), called the probability density function (p.d.f) through which we can find probabilities of events expressed in term of X.



$f: R \rightarrow [0, \infty)$

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

= area under the curve of f(x) and over the interval (a,b)

$P(X \in A) = \int_A f(x)\, dx$

= area under the curve of f(x) and over the region A

Definition 4:

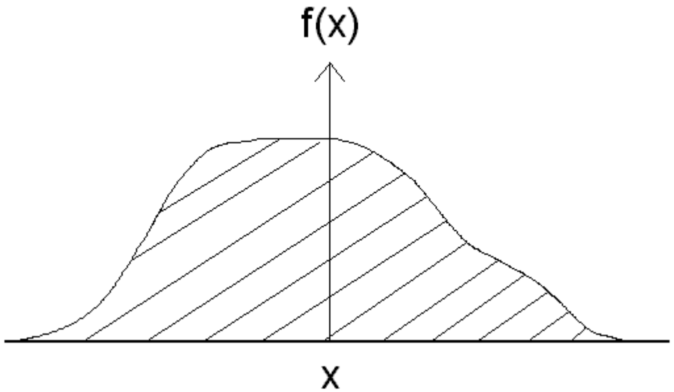
The function f(x) is a probability density function (pdf) for a continuous random variable X, defined on the set of real numbers, if:

1. $f(x) \geq 0 \quad \forall\, x \in R$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$
3. $P(a \leq X \leq b) = \int_{a}^{b} f(x)\, dx \qquad \forall\, a, b \in R;\ a \leq b$

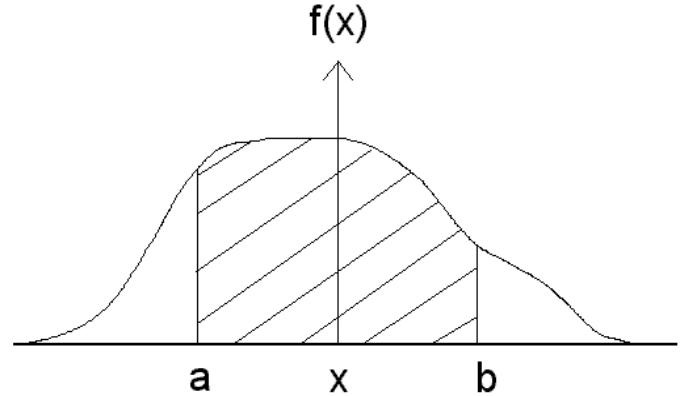
Note:

For a continuous random variable X, we have:

1. $f(x) \neq P(X=x)$ (in general)
2. $P(X=a) = 0$ for any $a \in R$
3. $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$
4. $P(X \in A) = \int_{A} f(x)\, dx$

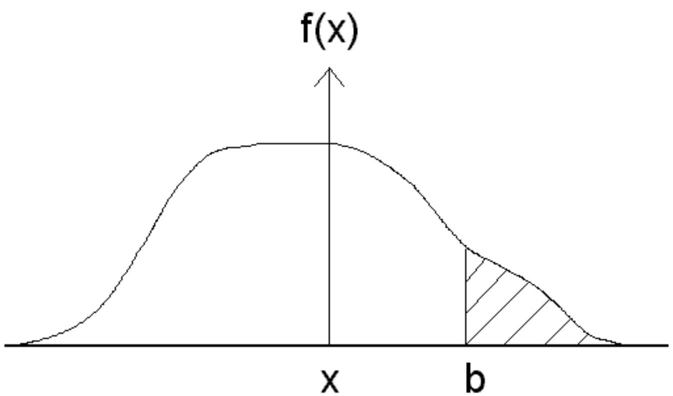


$$\text{Total area} = \int_{-\infty}^{\infty} f(x)\,dx = 1$$



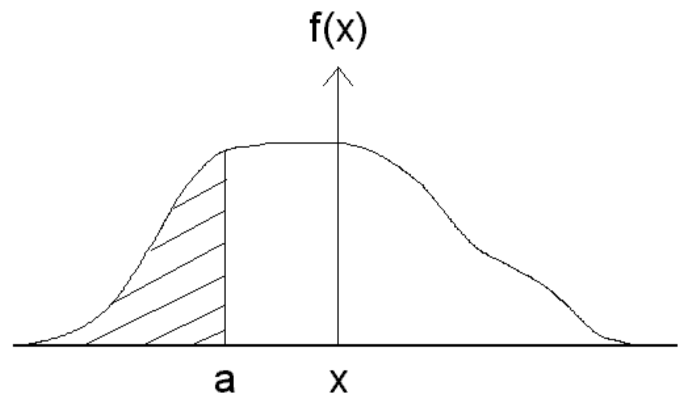
$$area = P(a \le X \le b)$$

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



$$area = P(X \ge b)$$

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



$$area = P(X \le a)$$

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

Example 7:

Suppose that the error in the reaction temperature, in ^oC, for a controlled laboratory experiment is a continuous random variable X having the following probability density function:

$$f(x) = \begin{cases} \frac{1}{3}x^2 \; ; -1 < x < 2 \\ 0 \; ; \; elsewhere \end{cases}$$

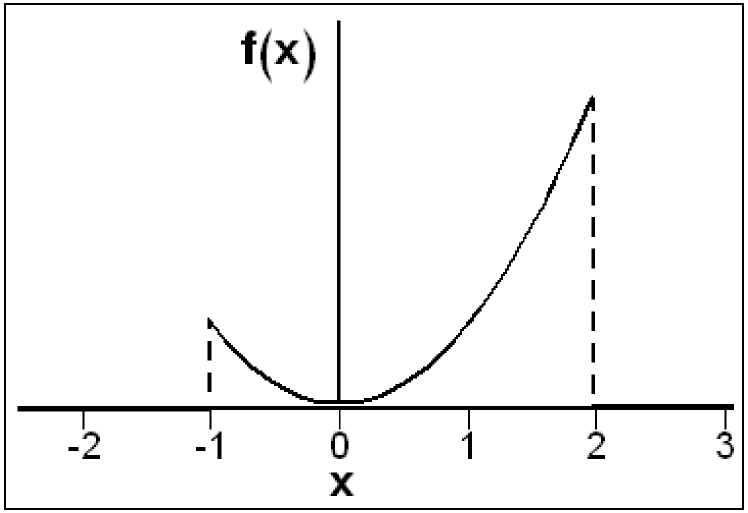
1. Verify that (a) $f(x) \geq 0$ and (b) $\int_{-\infty}^{\infty} f(x)\, dx = 1$
2. Find $P(0<X\leq 1)$

Solution:

X = the error in the reaction temperature in ^oC.

X is continuous r. v.

$$f(x) = \begin{cases} \frac{1}{3}x^2 \; ; -1 < x < 2 \\ 0 \; ; \; elsewhere \end{cases}$$

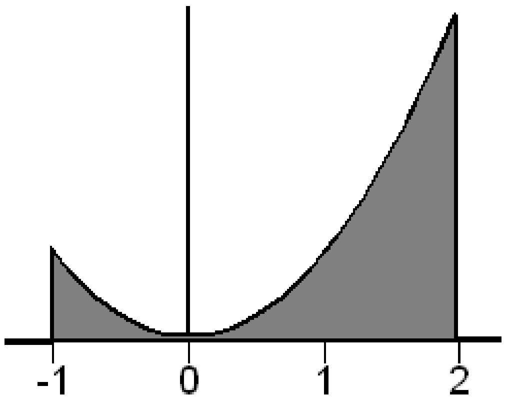


1. (a) $f(x) \geq 0$ because f(x) is a quadratic function.

(b) $$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{-1} 0\,dx + \int_{-1}^{2} \frac{1}{3}x^2\,dx + \int_{2}^{\infty} 0\,dx$$

$$= \int_{-1}^{2} \frac{1}{3}x^2\,dx = \left[\frac{1}{9}x^3 \;\middle|\; \begin{matrix} x=2 \\ x=-1 \end{matrix}\right]$$

$$= \frac{1}{9}(8-(-1)) = 1$$

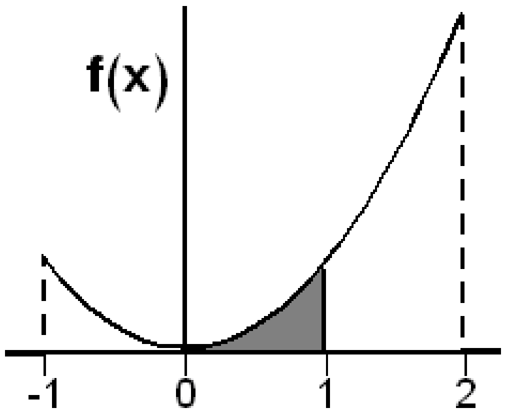


2. P(0<X≤1) = $\int_{0}^{1} f(x)\,dx = \int_{0}^{1} \frac{1}{3}x^2\,dx$

$$= \left[\frac{1}{9}x^3 \;\middle|\; \begin{matrix} x=1 \\ x=0 \end{matrix}\right]$$

$$= \frac{1}{9}(1-(0))$$

$$= \frac{1}{9}$$



The cumulative distribution function (CDF), F(x),

Definition 5:

The cumulative distribution function (CDF), F(x), of a continuous random variable X with probability density function f(x) is given by:

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

Result:

$P(a < X \leq b) = P(X \leq b) - \mathbf{P(X} \leq a) = F(b) - \mathbf{F(a)}$

Example 8:

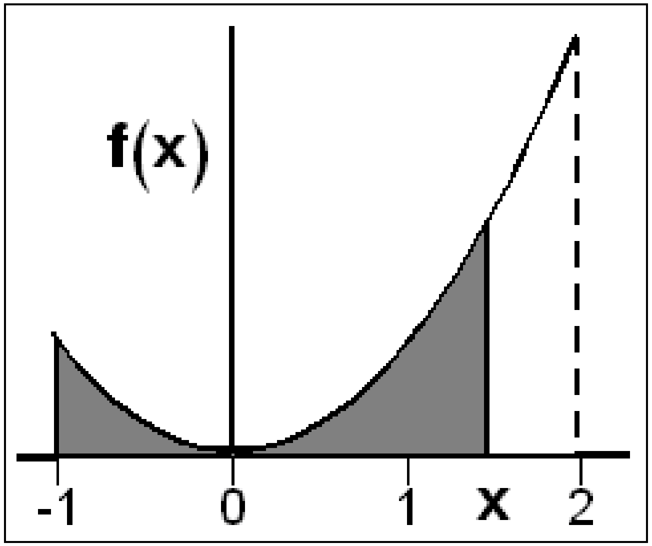
in Example 7,

1.Find the CDF

2.Using the CDF, find $P(0<X\leq 1)$.

Solution:

$$f(x)=\begin{cases}\frac{1}{3}x^2 \; ; -1<x<2\\ 0 \; ; \; elsewhere\end{cases}$$



For x< –1:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} 0\,dt = 0$$

For –1≤x<2:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-1} 0\,dt + \int_{-1}^{x} \frac{1}{3}t^2\,dt$$

$$= \int_{-1}^{x} \frac{1}{3}t^2\,dt$$

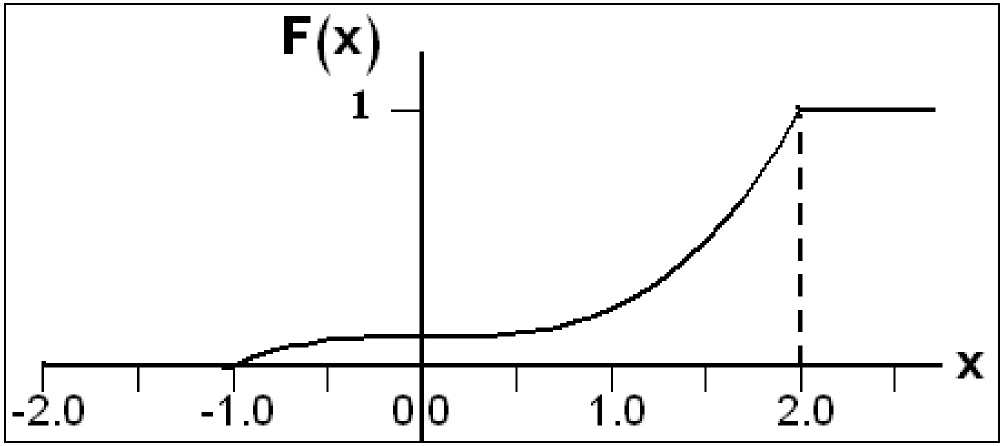
$$= \left[\frac{1}{9}t^3 \;\middle|\begin{matrix} t=x \\ t=-1\end{matrix}\right] = \frac{1}{9}(x^3-(-1)) = \frac{1}{9}(x^3+1)$$

For x≥2:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{-1} 0\,dt + \int_{-1}^{2} \frac{1}{3}t^2\,dt + \int_{2}^{x} 0\,dt = \int_{-1}^{2} \frac{1}{3}t^2\,dt = 1$$

Therefore, the CDF is:

$$F(x) = P(X \le x) = \begin{cases} 0 \;\; ; x < -1 \\ \frac{1}{9}(x^3+1) \;\; ; -1 \le x < 2 \\ 1 \;\; ; x \ge 2 \end{cases}$$



2. Using the CDF,
P(0<X≤1) = F(1) – F(0) = $\frac{2}{9} - \frac{1}{9} = \frac{1}{9}$

Exercise

$$p(1<x<2)=\int_1^2 (1/9)x^2\,dx = \frac{x^3}{27}\Big]_{x=1}^{x=2} = \frac{7}{27} \text{ and}$$

Since X is continous the prob. $p(X=x)=0$:

$$p(1\le X\le 2)=p(1<X\le 2)=p(1\le X<2)$$

$$=p(1<X<2)=\frac{7}{27}.$$

From CDF we can also find this probability

$$F(x)=\int_0^x (1/9)u^2\,du=\frac{x^3}{27}. \text{ Therefore}$$

$$F(x)=\begin{cases} 0, & \text{if } x<0 \\ \dfrac{x^3}{27}, & \text{if } 0<x<3 \\ 1, & \text{if } x>3 \end{cases}$$

$$p(1<X<2)=F(2)-F(1)=\frac{2^3}{27}-\frac{1^3}{27}=\frac{7}{27}$$

Exercise

(a) Find C such that the following is probabilty Density function (pdf) :

$$f(x) = \begin{cases} Cx^2, & 0 < x < 3 \\ = o, & otherwise \end{cases}$$

(b) Find CDF and P(1<x<2)

Exercise

Suppose that the error in the reaction temperature in C 0 for a controlled laboratory experiment is a continuous random variable X having the probability density function:

$$f(x) = \frac{x^2}{3} \quad , \quad -1 < x < 2$$
$$= 0 \qquad otherwise$$

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

b. Find P(0 < X ≤ 1).

c. Find P(0 < X < 3)

d. P(X=2), F(x), F(0.5)

Solution

a. $\int_{-1}^{2} \frac{x^2}{3}\, dx = \frac{x^3}{9}\Big|_{-1}^{2} = \frac{1}{9}[8-(-1)^3] = \frac{8+1}{9} = 1$

b. $P(0 < X \le 1) = \int_0^1 \frac{x^2}{3} dx = \frac{x^3}{9}\Big|_0^1 = \frac{1}{9}$

c. $P(0 < x < 3) = \int_0^2 \frac{x^2}{3} dx = \frac{x^3}{9}\Big|_0^2 = \frac{(2)^3}{9} = 8/9$

d. $P(x = 2) = 0$

$F(x) = \int_{-1}^{x} \frac{x^2}{3} dx = \frac{x^3}{9}\Big|_{-1}^{x} = \frac{x^3-(-1)^3}{9} = \frac{x^3+1}{9}$

$F(0.5) = \frac{(0.5)^3+1}{9} = 0.139$

Example

Number of Programs	frequency	P(X=x)	F(x)= P(X≤ x)
1	62	0.2088	0.2088
2	47	0.1582	0.3670
3	39	0.1313	0.4983
4	39	0.1313	0.6296
5	58	0.1953	0.8249
6	37	0.1246	0.9495
7	4	0.0135	0.9630
8	11	0.0370	1.0000
Total	297	1.0000	

Exercise 2

Find probability mass function and probability distribution for the following random variables:

1- X denote the sum of two upper most faces when two dice are thrown:

2- Y denote the no. of heads minus the no. of tails when 3 coins are thrown

Exercise 3

1-The probability mass function is given by X: 1 2 3

f(x): ½ c

Construct a table for Distribution function F(x) after determining C.

Find the following probabilities:

$P(1 \leq X \leq 3)$, $P(X \geq 2)$,$P(X<3)$