Engineering Probability & Statistics (AGE 1150)

Chapter 3: Random Variables and Probability Distributions

Dr. Feras Fraige

Concept of a Random Variable

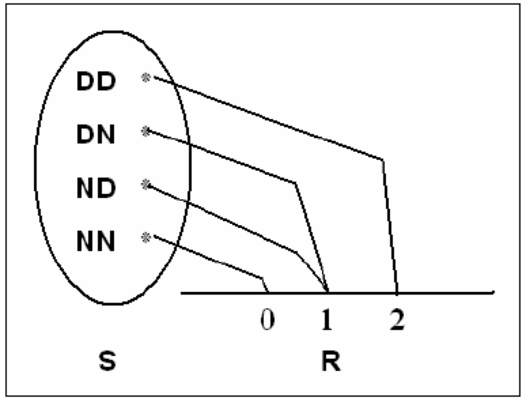
- In a statistical experiment, it is often very important to allocate numerical values to the outcomes.
- Example:
- Experiment: testing two components.
- (D=defective, N=non-defective)
- Sample space: S={DD,DN,ND,NN}
- Let X = number of defective components when

two components are tested.

- Assigned numerical values to the outcomes are shown
- Notice that, the set of all possible values of

the random variable X is {0, 1, 2}.

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



Concept of a Random Variable

Definition:

- 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 $x \in \{x_1, x_n, \ldots\}$
- 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.

Definition 3.2: If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a **discrete sample space**.

Definition 3.3: If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a **continuous sample space**.

Example 3.1: Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y, where Y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Examples

Example 3.2: A stockroom clerk returns three safety helmets at random to three steel mill employees who had previously checked them. If Smith, Jones, and Brown, in that order, receive one of the three hats, list the sample points for the possible orders of returning the helmets, and find the value m of the random variable M that represents the number of correct matches.

***Solution*:** If S, J, and B stand for Smith's, Jones's, and Brown's helmets, respectively, then the possible arrangements in which the helmets may be returned and the number of correct matches are

Sample Space	m
SJB	3
SBJ	1
BJS	1
JSB	1
JBS	0
BSJ	0

Example 3.3: Consider the simple condition in which components are arriving from the production line and they are stipulated to be defective or not defective. Define the random variable X by

$$X = \begin{cases} 1, & \text{if the component is defective,} \\ 0, & \text{if the component is not defective.} \end{cases}$$

Clearly the assignment of 1 or 0 is arbitrary though quite convenient. This will become clear in later chapters. The random variable for which 0 and 1 are chosen to describe the two possible values is called a **Bernoulli random variable**.

Example 3.4: Statisticians use **sampling plans** to either accept or reject batches or lots of material. Suppose one of these sampling plans involves sampling independently 10 items from a lot of 100 items in which 12 are defective.

Let X be the random variable defined as the number of items found defective in the sample of 10. In this case, the random variable takes on the values $0, 1, 2, \ldots, 9, 10$.

Example 3.5: Suppose a sampling plan involves sampling items from a process until a defective is observed. The evaluation of the process will depend on how many consecutive items are observed. In that regard, let X be a random variable defined by the number of items observed before a defective is found. With N a nondefective and D a defective, sample spaces are $S = \{D\}$ given $X = 1$, $S = \{ND\}$ given $X = 2$, $S = \{NND\}$ given $X = 3$, and so on.

Example 3.6: Interest centers around the proportion of people who respond to a certain mail order solicitation. Let X be that proportion. X is a random variable that takes on all values x for which $0 \leq x \leq 1$.

Example 3.7: Let X be the random variable defined by the waiting time, in hours, between successive speeders spotted by a radar unit. The random variable X takes on all values x for which $x \geq 0$.

Discrete Probability Distributions

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

Example:

- 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 "How it comes?"
- Let X= number of heads

Sample point (Outcome)	Probability	Value of X (x)
HH	$P(HH)=P(H)\ P(H)=1/3\times1/3 = 1/9$	2
HT	$P(HT)=P(H)\ P(T)=1/3\times2/3 = 2/9$	1
TH	$P(TH)=P(T)\ P(H)=2/3\times1/3 = 2/9$	1
TT	$P(TT)=P(T)\ P(T)=2/3\times2/3 = 4/9$	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, or probability mass function) of the discrete random variable X.

Definition 3.4: 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. $\sum\limits_{x} f(x) = 1$,
3. $P(X = x) = f(x)$.

Note:

- $P(X \in A) = \sum\limits_{all\ x \in A} f(x) = \sum\limits_{all\ x \in A} P(X = x)$

Example:

- For the previous example, we have:

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

Experiment: tossing a **non-balance coin** 2 times independently.
X= number of heads

- $P(X<1) = P(X=0) = 4/9$
- $P(X\le 1) = P(X=0) + P(X=1) = 4/9+4/9 = 8/9$
- $P(X\ge 0.5) = P(X=1) + P(X=2) = 4/9+1/9 = 5/9$
- $P(X>8) = P(\phi) = 0$
- $P(X<10) = P(X=0) + P(X=1) + P(X=2) = P(S) = 1$

Example

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

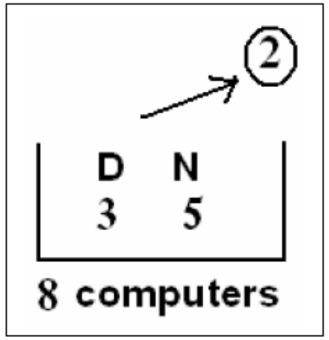
$$n(S) = \binom{8}{2} \text{ equally likely outcomes}$$

- The possible values of X are: x=0, 1, 2.
- Consider the events:

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

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

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



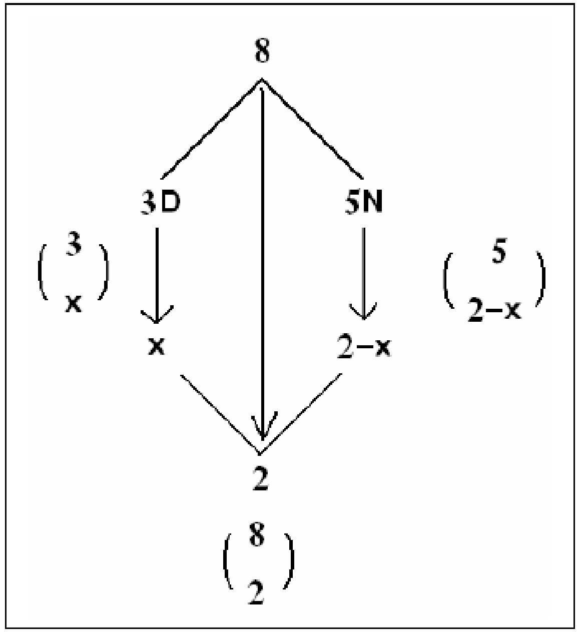
$$f(0)=P(X=0)=\frac{n(X=0)}{n(S)}=\frac{\binom{3}{0}\times\binom{5}{2}}{\binom{8}{2}}=\frac{10}{28}$$

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

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

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 can be given in the following table:

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

- The probability distribution of X can be written as a formula as follows

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

$$\boxed{\text{Hypergeometric Distribution}}$$

Definition:

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

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

Example:

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

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

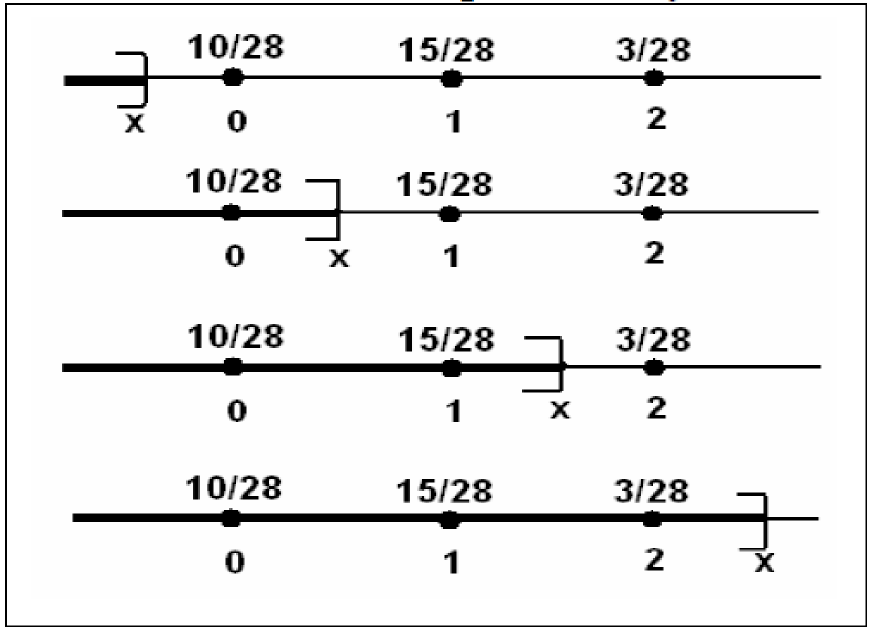
$F(x)=P(X\leq x)$ for $-\infty<x<\infty$

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

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

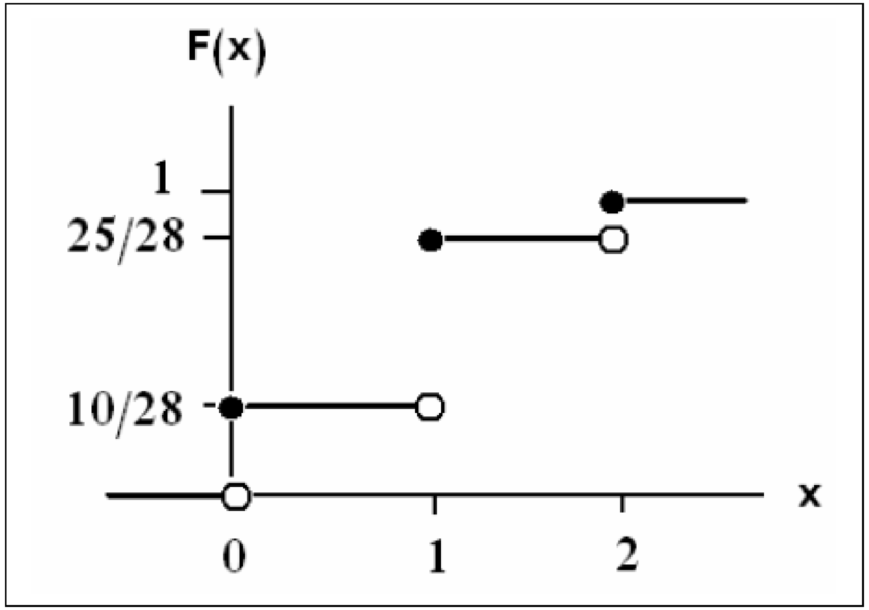
For $1\leq x<2$: $F(x)=P(X=0)+P(X=1)=\frac{10}{28}+\frac{15}{28}=\frac{25}{28}$

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



- The CDF of the random variable X is:

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



Note:

F(–0.5) = P(X≤–0.5)=0

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

F(3.8) =P(X≤3.8)F(2)= 1

Another way to find F(x):

x	<0	**0**	**1**	**2**	>2
f(x)		$\frac{10}{28}$	$\frac{15}{28}$	$\frac{3}{28}$	
F(x)	0	$\frac{10}{28}$	$\frac{25}{28}$	$\frac{28}{28}$	1

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

- **Result:**

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

$P(a \leq X \leq b) = P(a < X \leq b) + P(X=a) = F(b) - F(a) + f(a)$

$P(a < X < b) = P(a < X \leq b) - P(X=b) = F(b) - F(a) - f(b)$

- **Result:**

Suppose that the probability function of X is:

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

- Where $x_1 < x_2 < \ldots < x_n$. Then:
- $F(x_i) = f(x_1) + f(x_2) + \ldots + f(x_n)$; $i=1, 2, \ldots, n$
- $F(x_i) = F(x_{i-1}) + f(x_i)$; $i=2, \ldots, n$
- $f(x_i) = F(x_i) - F(x_{i-1})$

- **Example:**

In the previous example,

$P(0.5 < X \leq 1.5) = F(1.5) - F(0.5) =$

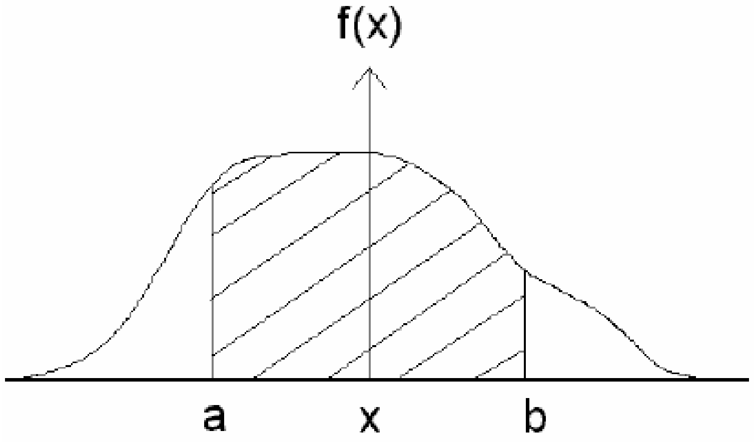
$$\frac{25}{28} - \frac{10}{28} = \frac{15}{28}$$

$P(1 < X \leq 2) = F(2) - F(1) =$

$$1 - \frac{25}{28} = \frac{3}{28}$$

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.



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

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

= 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 3.6:**

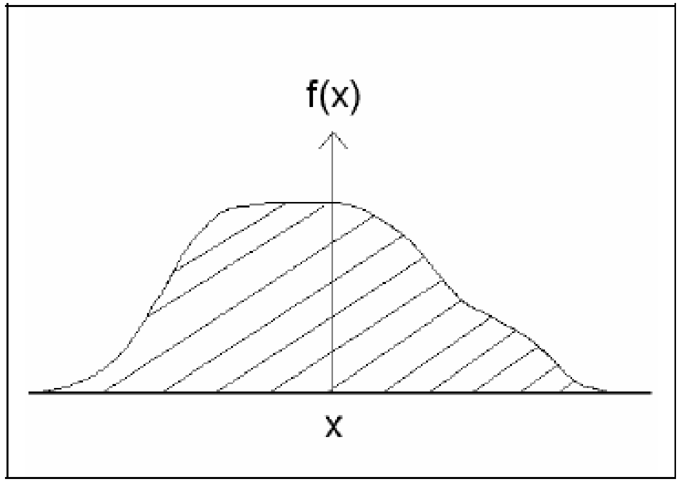
The function f(x) is a probability density function (pdf) for a continuous random variable X, defined on the set of real numbers, if:

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

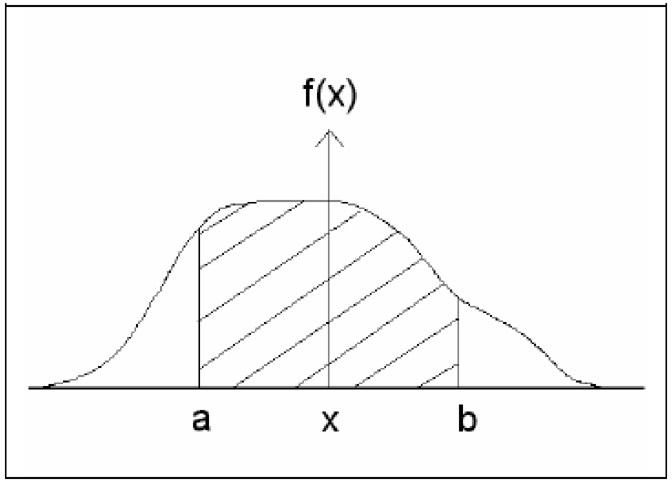
Note:

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. $P(X \in A) = \int_{A} f(x)\,dx$

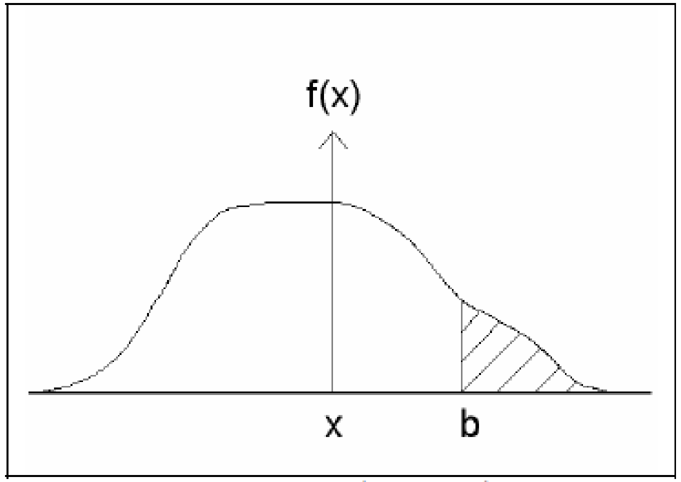


$$Total\ area = \int_{-\infty}^{\infty} f(x)\, dx = 1$$



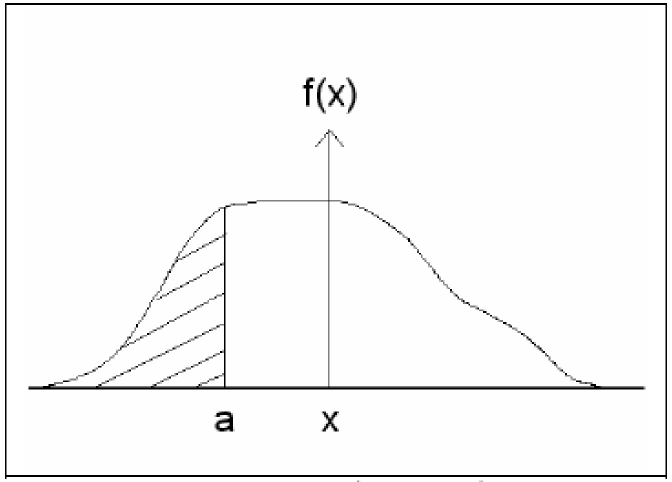
$$area = P(a \le X \le b)$$

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



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

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

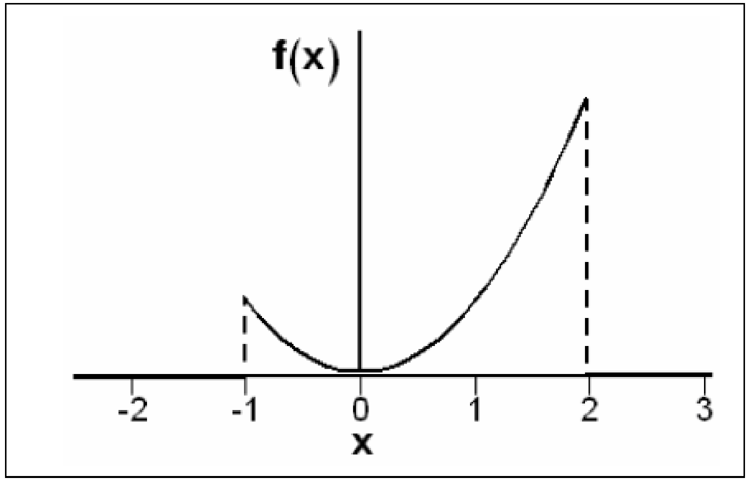


$$area = P(X \le a)$$

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

Example: 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) $f(x) \geq 0$
   (b) $\int_{-\infty}^{\infty} f(x)\,dx = 1$
2. Find $P(0<X\leq1)$

Solution:

- X = the error in the reaction temperature in ^{o}C.
- X is continuous random variable.

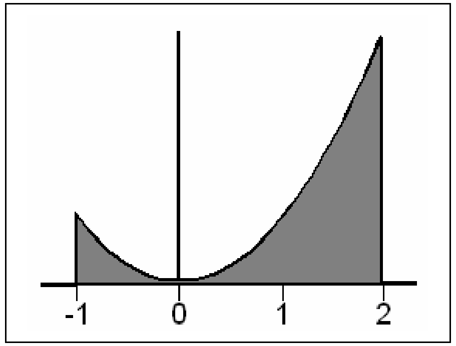
$$f(x) = \begin{cases} \frac{1}{3}x^2 \ ; -1 < x < 2 \\ 0 \ ; \ elsewhere \end{cases}$$

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

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

$= \int_{-1}^{2} \frac{1}{3}x^2\,dx = \left[\frac{1}{9}x^3 \Big|_{x=-1}^{x=2}\right]$

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

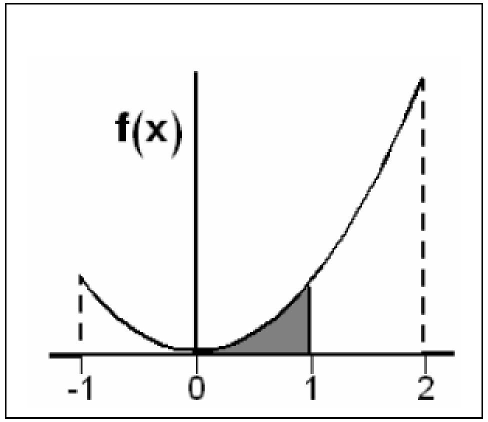


2. $P(0<X\leq 1) = \int_{0}^{1} f(x)\,dx = \int_{0}^{1} \frac{1}{3}x^2\,dx$

$= \left[\frac{1}{9}x^3 \Big|_{x=0}^{x=1}\right]$

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

$= \frac{1}{9}$



- **Definition:**

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

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

- **Result:**

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

Example:

in the previous example,

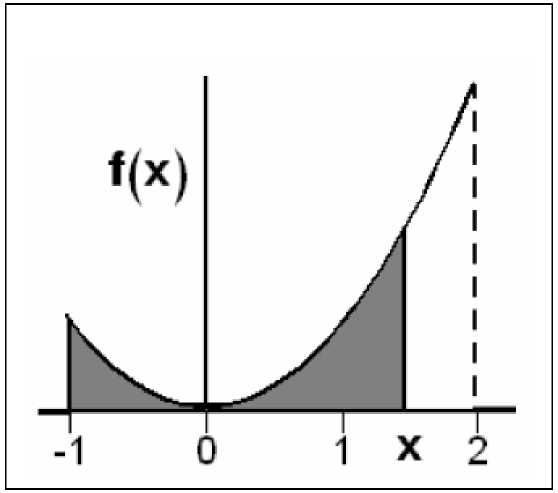
1. Find the CDF
2. Using the CDF, find $P(0<X\leq 1)$.

- **Solution:**

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

(1) Finding F(x):

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



For $x<-1$:

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

For $-1\leq x<2$:

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

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

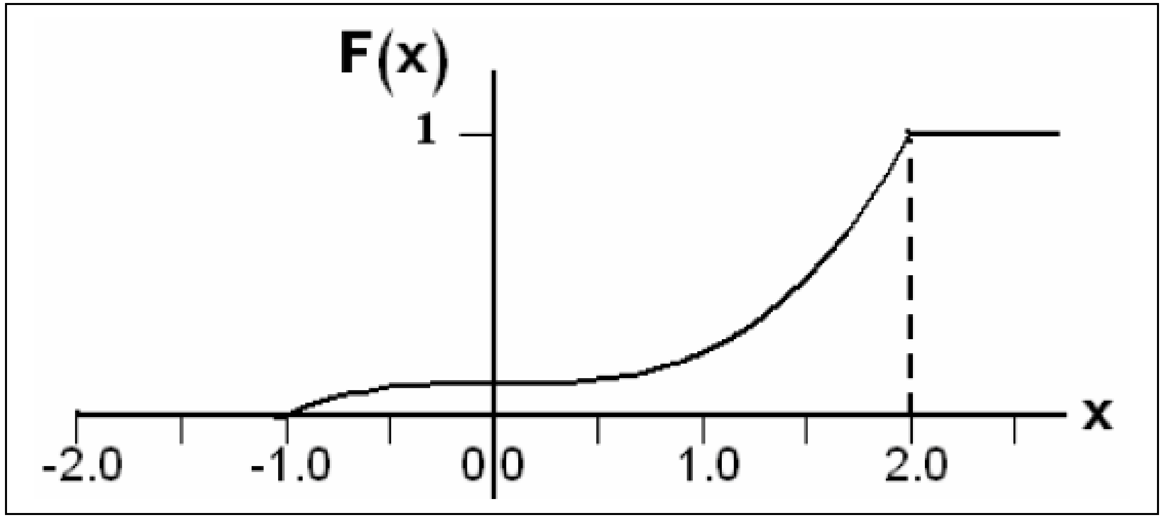
$$=\left[\frac{1}{9}t^3\;\Big|\begin{matrix}t=x\\ t=-1\end{matrix}\right]=\frac{1}{9}(x^3-(-1))=\frac{1}{9}(x^3+1)$$

For $x\geq 2$:

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

Therefore, the CDF is:

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



2. Using the CDF,

$$P(0<X\leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$