Factorial

رياضيات اول تكنلوجي

1

def The product of n consecutive positive integer beginning with 1 denoted by $n!$ and is read as factorial n.

$$n! = 1\cdot 2\cdot 3 \cdot \cdots (n-1)\,n$$

or $$n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$$

Ex:

$$5! = 5\times 4\times 3\times 2\times 1$$
$$= 1\times 2\times 3\times 4\times 5 \qquad = 120$$
or $$= 5\cdot 4\cdot 3\cdot 2\cdot 1$$
$$= 1\cdot 2\cdot 3\cdot 4\cdot 5$$

$$6! = 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1$$
or $$= 6\cdot 5!$$

$$7! = 7\cdot 6!$$

Ex: compute $\frac{12!}{10!\,2!}$

$$\Rightarrow \frac{12\cdot 11\cdot \cancel{10!}}{\cancel{10!}\cdot 2!} = \frac{\overset{6}{\cancel{12}}\cdot 11}{\cancel{2}} = 6\cdot 11 = 66.$$

Ex: if $\frac{1}{8!} + \frac{1}{9!} = \frac{x}{10!}$, find x?

$$\Rightarrow \frac{1}{8!} + \frac{1}{9\cdot 8!} = \frac{x}{10\cdot 9\cdot 8!}$$

$$\Rightarrow \frac{1}{8!} + \frac{1}{9}\cdot\frac{1}{8!} = \frac{x}{10\cdot 9}\cdot\frac{1}{8!}$$

$$\Rightarrow \frac{1}{8!}\left[1 + \frac{1}{9}\right] = \frac{1}{8!}\left[\frac{x}{10\cdot 9}\right]$$

$$\Rightarrow 1 + \frac{1}{9} = \frac{x}{10\cdot 9}$$

$$\frac{10}{9} = \frac{x}{90}$$

$$9x = 900$$

$$\boxed{x = 100}$$

permutation

$$p(n,r) = \frac{n!}{(n-r)!} \quad , \quad n \geq r$$

Ex $p(9,4) = \frac{9!}{(9-4)!}$

$$= \frac{9!}{5!}$$

$$= \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5!}{5!}$$

$$= 3024.$$

Ex $p(6,3) - p(5,2)$

$$\Rightarrow \frac{6!}{(6-3)!} - \frac{5!}{(5-2)!}$$

$$\Rightarrow \frac{6!}{3!} - \frac{5!}{3!}$$

$$\Rightarrow \frac{6! - 5!}{3!} = \frac{720 - 120}{6}$$

$$= \frac{600}{6} = 100$$

* Find The value of n ?

$$P(n,5) = 42\,P(n,3)$$

$$\Rightarrow \frac{n!}{(n-5)!} = 42\,\frac{n!}{(n-3)!}$$

$$\Rightarrow \frac{n(n-1)(n-2)(n-3)(n-4)\cancel{(n-5)!}}{\cancel{(n-5)!}} =$$

$$\frac{42 \cdot n(n-1)(n-2)\cancel{(n-3)!}}{\cancel{(n-3)!}}$$

$$\Rightarrow \frac{n(n-1)(n-2)(n-3)(n-4)}{1} = \frac{42\,n(n-1)(n-2)}{1}$$

$$\Rightarrow \cancel{n}\cancel{(n-1)}\cancel{(n-2)}(n-3)(n-4) = 42\,\cancel{n}\cdot\cancel{(n-1)}\cancel{(n-2)}$$

$$\Rightarrow (n-3)(n-4) = 42$$

$$\Rightarrow n^2 - 7n + 12 = 42$$

$$\Rightarrow n^2 - 7n + 12 - 42 = 0$$

$$\Rightarrow n^2 - 7n - 30 = 0$$

$$\Rightarrow (n-10)(n+3) = 0$$

either $n - 10 = 0 \Rightarrow n = 10$

or $n + 3 = 0 \Rightarrow n = -3$

Then Cancel $(-3) \Rightarrow \boxed{n = 10}$

Ex if $\frac{P(n,4)}{P(n-1,4)} = \frac{5}{3}$, $n > 4$ (4)

Find n?

$$\Rightarrow \frac{\frac{n!}{(n-4)!}}{\frac{(n-1)!}{(n-5)!}} = \frac{5}{3}$$

$$\Rightarrow \frac{\frac{n(n-1)(n-2)(n-3)(n-4)!}{(n-4)!}}{\frac{(n-1)(n-2)(n-3)(n-4)(n-5)!}{(n-5)!}} = \frac{5}{3}$$

$$\Rightarrow \frac{n}{(n-4)} = \frac{5}{3}$$

$3n = 5(n-4)$

$3n = 5n - 20$

$3n - 5n = -20$

$-2n = -20$

$2n = 20$

$\Rightarrow \boxed{n = 10}$

if $5P(4,r) = 6P(5,r-1)$, find r? (5)

$$\Rightarrow 5 \times \frac{4!}{(4-r)!} = 6 \times \frac{5!}{(5-r+1)!}$$

$$\Rightarrow \frac{5 \cdot 4!}{(4-r)!} = \frac{6 \cdot 5!}{(6-r)!}$$

$$\Rightarrow 6 \cdot 5!\,(4-r)! = 5!\,(6-r)!$$

$$\Rightarrow 6 \cdot \cancel{5!}\,\cancel{(4-r)}(3-r)! = \cancel{5!}(6-r)(5-r)\cancel{(4-r)}\cancel{(3-r)!}$$

$$6 = (5-r)(6-r)$$

$$6 = 30 - 11r + r^2$$

$$\Rightarrow r^2 - 11r + 24 = 0$$

$$\Rightarrow (r-8)(r-3) = 0$$

$r = 3$

$r = 8$

Then $\boxed{r = 3}$ since $n \geq r$

Combination

$$C(n,r) = \frac{n!}{r!\,(n-r)!}, \quad n \geq r$$

Ex:

$$C(52,4) = \frac{52!}{4!\,48!}$$

$$= \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot \cancel{48!}}{4!\,\cancel{48!}}$$

$$= \frac{52 \cdot 51 \cdot 50 \cdot 49}{24}$$

$$= 270725$$

Ex: if $C(n,x) = 56$ and $P(n,x) = 336$

find n, x?

$\Rightarrow$ since $P(n,x) = \frac{n!}{(n-x)!} = 336$

and $C(n,x) = \frac{n!}{x!\,(n-x)!} = 56$

so [redacted] $\frac{C(n,x)}{P(n,x)} \Rightarrow \frac{56}{336}$

$\Rightarrow$ [redacted] $\frac{\cancel{n!}}{x!\,\cancel{(n-x)!}} \cdot \frac{\cancel{(n-x)!}}{\cancel{n!}} = \frac{56}{336}$

$$\frac{1}{x!} = \frac{56}{336}$$

$$\Rightarrow \frac{1}{x!} = \frac{1}{6}$$

$$\Rightarrow \frac{1}{x!} = \frac{1}{3!}$$

$$\Rightarrow x = 3$$

Now sub x in $p(n, x)$

$$\Rightarrow \frac{n!}{(n-3)!} = 336$$

$$\Rightarrow \frac{n(n-1)(n-2)(n-3)!}{(n-3)!} = 336$$

$$\Rightarrow n(n-1)(n-2) = 336$$

So $336 = 8 \times 7 \times 6$

$$\Rightarrow n(n-1)(n-2) = 8 \times 7 \times 6$$

$$\Rightarrow \boxed{n = 8}$$

Ex if $C(n,8) = C(n,6)$ find n?

$$\Rightarrow \frac{n!}{8!(n-8)!} = \frac{n!}{6!(n-6)!}$$

$$\Rightarrow \frac{n(n-1)(n-2)(n-3)(n-4)(n-5)(n-6)(n-7)(n-8)!}{8!\,(n-8)!} =$$

$$\frac{n(n-1)(n-2)(n-3)(n-4)(n-5)(n-6)!}{6!\,(n-6)!}$$

$$\Rightarrow \frac{(n-6)(n-7)}{8.7.6!} = \frac{1}{6!}$$

$$\Rightarrow \frac{(n-6)(n-7)}{56} = 1$$

$$\Rightarrow (n-6)(n-7) = 56$$

$$\Rightarrow n^2 - 13n + 42 - 56 = 0$$

$$\Rightarrow n^2 - 13n - 14 = 0$$

$$(n-14)(n+1) = 0$$

$n = 14$

$n = 1$

$\Rightarrow \boxed{n = 14}$

Ex if $p(n,r) = 1680$

and $C(n,r) = 70$

find n, r?

$$\frac{n!}{(n-r)!} = 1680$$

$$\frac{n!}{r!(n-r)!} = 70$$

now $\frac{C}{P} = \frac{70}{1680}$

$$\Rightarrow \frac{n!}{r!(n-r)!} \cdot \frac{(n-r)!}{n!} = \frac{70}{1680}$$

$$\frac{1}{r!} = \frac{1}{24}$$

$$\Rightarrow \frac{1}{r!} = \frac{1}{4!}$$

$$\Rightarrow \boxed{r = 4}$$

sub in $p(n,r) = 1680$

$$\Rightarrow p(n,4) = 1680$$

$$\Rightarrow \frac{n!}{(n-4)!} = 1680$$

$$\frac{n(n-1)(n-2)(n-3)(n-4)!}{(n-4)!} = 1680 \quad \text{(10)}$$

$$\Rightarrow n(n-1)(n-2)(n-3) = 1680$$

$$\Rightarrow (n^2-3n)(n^2-3n+2) = 1680$$

$$\Rightarrow (n^2-3n+1)^2 = 1680$$

$$\Rightarrow (n^2-3n+1)^2 = 41^2$$

$$\Rightarrow n^2-3n+1 = 41$$

$$\Rightarrow n^2-3n+1-41 = 0$$

$$\Rightarrow n^2-3n-40 = 0$$

$$(n-8)(n+5) = 0$$

$$\boxed{n = 8}$$

laid down some axioms to interpret probability, in his book 'Foundation of Probability' published in 1933. In this Chapter, we will study about this approach called *axiomatic approach of probability*. To understand this approach we must know about few basic terms viz. random experiment, sample space, events, etc. Let us learn about these all, in what follows next.

16.2 Random Experiments

In our day to day life, we perform many activities which have a fixed result no matter any number of times they are repeated. For example given any triangle, without knowing the three angles, we can definitely say that the sum of measure of angles is 180°.

We also perform many experimental activities, where the result may not be same, when they are repeated under identical conditions. For example, when a coin is tossed it may turn up a head or a tail, but we are not sure which one of these results will actually be obtained. Such experiments are called *random experiments*.

An experiment is called random experiment if it satisfies the following two conditions:

(i) It has more than one possible outcome.

(ii) It is not possible to predict the outcome in advance.

Check whether the experiment of tossing a die is random or not?

In this chapter, we shall refer the random experiment by experiment only unless stated otherwise.

16.2.1 ***Outcomes and sample space*** A possible result of a random experiment is called its *outcome*.

Consider the experiment of rolling a die. The outcomes of this experiment are 1, 2, 3, 4, 5, or 6, if we are interested in the number of dots on the upper face of the die.

The set of outcomes {1, 2, 3, 4, 5, 6} is called the *sample space of the experiment*.

Thus, the set of all possible outcomes of a random experiment is called the *sample space* associated with the experiment. Sample space is denoted by the symbol S.

Each element of the sample space is called a *sample point*. In other words, each outcome of the random experiment is also called *sample point*.

Let us now consider some examples.

Example 1 Two coins (a one rupee coin and a two rupee coin) are tossed once. Find a sample space.

Solution Clearly the coins are distinguishable in the sense that we can speak of the first coin and the second coin. Since either coin can turn up Head (H) or Tail(T), the possible outcomes may be

Heads on both coins = (H,H) = HH
Head on first coin and Tail on the other = (H,T) = HT
Tail on first coin and Head on the other = (T,H) = TH
Tail on both coins = (T,T) = TT
Thus, the sample space is S = {HH, HT, TH, TT}

Note The outcomes of this experiment are ordered pairs of H and T. For the sake of simplicity the commas are omitted from the ordered pairs.

Example 2 Find the sample space associated with the experiment of rolling a pair of dice (one is blue and the other red) once. Also, find the number of elements of this sample space.

Solution Suppose 1 appears on blue die and 2 on the red die. We denote this outcome by an ordered pair (1,2). Similarly, if '3' appears on blue die and '5' on red, the outcome is denoted by the ordered pair (3,5).

In general each outcome can be denoted by the ordered pair (x, y), where x is the number appeared on the blue die and y is the number appeared on the red die. Therefore, this sample space is given by

S = {(x, y): x is the number on the blue die and y is the number on the red die}.

The number of elements of this sample space is $6 \times 6 = 36$ and the sample space is given below:

{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6)
(3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6)
(5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

Example 3 In each of the following experiments specify appropriate sample space

(i) A boy has a 1 rupee coin, a 2 rupee coin and a 5 rupee coin in his pocket. He takes out two coins out of his pocket, one after the other.

(ii) A person is noting down the number of accidents along a busy highway during a year.

Solution (i) Let Q denote a 1 rupee coin, H denotes a 2 rupee coin and R denotes a 5 rupee coin. The first coin he takes out of his pocket may be any one of the three coins Q, H or R. Corresponding to Q, the second draw may be H or R. So the result of two draws may be QH or QR. Similarly, corresponding to H, the second draw may be Q or R.

Therefore, the outcomes may be HQ or HR. Lastly, corresponding to R, the second draw may be H or Q.

So, the outcomes may be RH or RQ.

Thus, the sample space is S={QH, QR, HQ, HR, RH, RQ}

(ii) The number of accidents along a busy highway during the year of observation can be either 0 (for no accident) or 1 or 2, or some other positive integer.

Thus, a sample space associated with this experiment is S= {0,1,2,...}

Example 4 A coin is tossed. If it shows head, we draw a ball from a bag consisting of 3 blue and 4 white balls; if it shows tail we throw a die. Describe the sample space of this experiment.

Solution Let us denote blue balls by B_1, B_2, B_3 and the white balls by W_1, W_2, W_3, W_4. Then a sample space of the experiment is

$$S = \{ HB_1, HB_2, HB_3, HW_1, HW_2, HW_3, HW_4, T1, T2, T3, T4, T5, T6\}.$$

Here HB_i means head on the coin and ball B_i is drawn, HW_i means head on the coin and ball W_i is drawn. Similarly, Ti means tail on the coin and the number i on the die.

Example 5 Consider the experiment in which a coin is tossed repeatedly until a head comes up. Describe the sample space.

Solution In the experiment head may come up on the first toss, or the 2nd toss, or the 3rd toss and so on till head is obtained. Hence, the desired sample space is

S= {H, TH, TTH, TTTH, TTTTH,...}

EXERCISE 16.1

In each of the following Exercises 1 to 7, describe the sample space for the indicated experiment.

1. A coin is tossed three times.
2. A die is thrown two times.
3. A coin is tossed four times.
4. A coin is tossed and a die is thrown.
5. A coin is tossed and then a die is rolled only in case a head is shown on the coin.
6. 2 boys and 2 girls are in Room X, and 1 boy and 3 girls in Room Y. Specify the sample space for the experiment in which a room is selected and then a person.
7. One die of red colour, one of white colour and one of blue colour are placed in a bag. One die is selected at random and rolled, its colour and the number on its uppermost face is noted. Describe the sample space.
8. An experiment consists of recording boy–girl composition of families with 2 children.
   (i) What is the sample space if we are interested in knowing whether it is a boy or girl in the order of their births?

Description of events	Corresponding subset of 'S'
Number of tails is exactly 2	A = {TT}
Number of tails is atleast one	B = {HT, TH, TT}
Number of heads is atmost one	C = {HT, TH, TT}
Second toss is not head	D = { HT, TT}
Number of tails is atmost two	S = {HH, HT, TH, TT}
Number of tails is more than two	ϕ

The above discussion suggests that a subset of sample space is associated with an event and an event is associated with a subset of sample space. In the light of this we define an event as follows.

Definition Any subset E of a sample space S is called *an event*.

16.3.1 ***Occurrence of an event*** Consider the experiment of throwing a die. Let E denotes the event " a number less than 4 appears". If actually '1' had appeared on the die then we say that event E has occurred. As a matter of fact if outcomes are 2 or 3, we say that event E has occurred

Thus, the event E of a sample space S is said to have occurred if the outcome ω of the experiment is such that $\omega \in$ E. If the outcome ω is such that $\omega \notin$ E, we say that the event E has not occurred.

16.3.2 ***Types of events*** Events can be classified into various types on the basis of the elements they have.

1. Impossible and Sure Events The empty set ϕ and the sample space S describe events. In fact ϕ is called an *impossible event* and S, i.e., the whole sample space is called the *sure event*.

To understand these let us consider the experiment of rolling a die. The associated sample space is

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

Let E be the event " the number appears on the die is a multiple of 7". Can you write the subset associated with the event E?

Clearly no outcome satisfies the condition given in the event, i.e., no element of the sample space ensures the occurrence of the event E. Thus, we say that the empty set only correspond to the event E. In other words we can say that it is impossible to have a multiple of 7 on the upper face of the die. Thus, the event E = ϕ is an impossible event.

Now let us take up another event F "the number turns up is odd or even". Clearly

F = {1, 2, 3, 4, 5, 6,} = S, i.e., all outcomes of the experiment ensure the occurrence of the event F. Thus, the event F = S is a sure event.

2. Simple Event If an event E has only one sample point of a sample space, it is called a *simple* (or *elementary*) *event*.

In a sample space containing *n* distinct elements, there are exactly *n* simple events.

For example in the experiment of tossing two coins, a sample space is

S={HH, HT, TH, TT}

There are four simple events corresponding to this sample space. These are

E_1= {HH}, E_2={HT}, E_3= { TH} and E_4={TT}.

3. Compound Event If an event has more than one sample point, it is called a *Compound event*.

For example, in the experiment of "tossing a coin thrice" the events

E: 'Exactly one head appeared'

F: 'Atleast one head appeared'

G: 'Atmost one head appeared' etc.

are all compound events. The subsets of S associated with these events are

E={HTT,THT,TTH}

F={HTT,THT, TTH, HHT, HTH, THH, HHH}

G= {TTT, THT, HTT, TTH}

Each of the above subsets contain more than one sample point, hence they are all compound events.

16.3.3 ***Algebra of events*** In the Chapter on Sets, we have studied about different ways of combining two or more sets, viz, union, intersection, difference, complement of a set etc. Like-wise we can combine two or more events by using the analogous set notations.

Let A, B, C be events associated with an experiment whose sample space is S.

1. Complementary Event For every event A, there corresponds another event A' called the complementary event to A. It is also called the *event 'not* A'.

For example, take the experiment 'of tossing three coins'. An associated sample space is

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

Let A={HTH, HHT, THH} be the event 'only one tail appears'

Clearly for the outcome HTT, the event A has not occurred. But we may say that the event 'not A' has occurred. Thus, with every outcome which is not in A, we say that 'not A' occurs.

Thus the complementary event 'not A' to the event A is

$$A' = \{HHH, HTT, THT, TTH, TTT\}$$

or $\quad A' = \{\omega : \omega \in S \text{ and } \omega \notin A\} = S - A.$

2. The Event 'A or B' Recall that union of two sets A and B denoted by $A \cup B$ contains all those elements which are either in A or in B or in both.

When the sets A and B are two events associated with a sample space, then '$A \cup B$' is the event 'either A or B or both'. This event '$A \cup B$' is also called 'A or B'.

Therefore $\quad$ Event 'A or B' $= A \cup B$

$$= \{\omega : \omega \in A \text{ or } \omega \in B\}$$

3. The Event 'A and B' We know that intersection of two sets $A \cap B$ is the set of those elements which are common to both A and B. i.e., which belong to both 'A and B'.

If A and B are two events, then the set $A \cap B$ denotes the event 'A and B'.

Thus, $\quad A \cap B = \{\omega : \omega \in A \text{ and } \omega \in B\}$

For example, in the experiment of 'throwing a die twice' Let A be the event 'score on the first throw is six' and B is the event 'sum of two scores is atleast 11' then

$$A = \{(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}, \text{ and } B = \{(5,6), (6,5), (6,6)\}$$

so $\quad A \cap B = \{(6,5), (6,6)\}$

Note that the set $A \cap B = \{(6,5), (6,6)\}$ may represent the event 'the score on the first throw is six and the sum of the scores is atleast 11'.

4. The Event 'A but not B' We know that A–B is the set of all those elements which are in A but not in B. Therefore, the set A–B may denote the event 'A but not B'.We know that

$$A - B = A \cap B'$$

Example 6 Consider the experiment of rolling a die. Let A be the event 'getting a prime number', B be the event 'getting an odd number'. Write the sets representing the events (i) Aor B (ii) A and B (iii) A but not B (iv) 'not A'.

Solution Here $\quad S = \{1, 2, 3, 4, 5, 6\}, A = \{2, 3, 5\}$ and $B = \{1, 3, 5\}$

Obviously

(i) 'A or B' $= A \cup B = \{1, 2, 3, 5\}$

(ii) 'A and B' $= A \cap B = \{3,5\}$

(iii) 'A but not B' $= A - B = \{2\}$

(iv) 'not A' $= A' = \{1,4,6\}$

16.3.4 ***Mutually exclusive events*** In the experiment of rolling a die, a sample space is S = {1, 2, 3, 4, 5, 6}. Consider events, A 'an odd number appears' and B 'an even number appears'

Clearly the event A excludes the event B and vice versa. In other words, there is no outcome which ensures the occurrence of events A and B simultaneously. Here

$A = \{1, 3, 5\}$ and $B = \{2, 4, 6\}$

Clearly $A \cap B = \phi$, i.e., A and B are disjoint sets.

In general, two events A and B are called *mutually exclusive* events if the occurrence of any one of them excludes the occurrence of the other event, i.e., if they can not occur simultaneously. In this case the sets A and B are disjoint.

Again in the experiment of rolling a die, consider the events A 'an odd number appears' and event B 'a number less than 4 appears'

Obviously $A = \{1, 3, 5\}$ and $B = \{1, 2, 3\}$

Now $3 \in A$ as well as $3 \in B$

Therefore, A and B are not mutually exclusive events.

Remark Simple events of a sample space are always mutually exclusive.

16.3.5 ***Exhaustive events*** Consider the experiment of throwing a die. We have S = {1, 2, 3, 4, 5, 6}. Let us define the following events

A: 'a number less than 4 appears',
B: 'a number greater than 2 but less than 5 appears'
and C: 'a number greater than 4 appears'.

Then A = {1, 2, 3}, B = {3,4} and C = {5, 6}. We observe that

$$A \cup B \cup C = \{1, 2, 3\} \cup \{3, 4\} \cup \{5, 6\} = S.$$

Such events A, B and C are called exhaustive events. In general, if $E_1, E_2, ..., E_n$ are *n* events of a sample space S and if

$$E_1 \cup E_2 \cup E_3 \cup ... \cup E_n = \bigcup_{i=1}^{n} E_i = S$$

then $E_1, E_2, ..., E_n$ are called *exhaustive events*.In other words, events $E_1, E_2, ..., E_n$ are said to be exhaustive if atleast one of them necessarily occurs whenever the experiment is performed.

Further, if $E_i \cap E_j = \phi$ for $i \neq j$ i.e., events E_i and E_j are pairwise disjoint and $\bigcup_{i=1}^{n} E_i = S$, then events $E_1, E_2, ..., E_n$ are called *mutually exclusive and exhaustive events*.

We now consider some examples.

Example 7 Two dice are thrown and the sum of the numbers which come up on the dice is noted. Let us consider the following events associated with this experiment

A: 'the sum is even'.
B: 'the sum is a multiple of 3'.
C: 'the sum is less than 4'.
D: 'the sum is greater than 11'.

Which pairs of these events are mutually exclusive?

Solution There are 36 elements in the sample space S = {(x, y): x, y = 1, 2, 3, 4, 5, 6}. Then

A = {(1, 1), (1, 3), (1, 5), (2, 2), (2, 4), (2, 6), (3, 1), (3, 3), (3, 5), (4, 2), (4, 4), (4, 6), (5, 1), (5, 3), (5, 5), (6, 2), (6, 4), (6, 6)}

B = {(1, 2), (2, 1), (1, 5), (5, 1), (3, 3), (2, 4), (4, 2), (3, 6), (6, 3), (4, 5), (5, 4), (6, 6)}

C = {(1, 1), (2, 1), (1, 2)} and D = {(6, 6)}

We find that

$$A \cap B = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1), (6, 6)\} \neq \phi$$

Therefore, A and B are not mutually exclusive events.

Similarly $A \cap C \neq \phi$, $A \cap D \neq \phi$, $B \cap C \neq \phi$ and $B \cap D \neq \phi$.

Thus, the pairs of events, (A, C), (A, D), (B, C), (B, D) are not mutually exclusive events.

Also $C \cap D = \phi$ and so C and D are mutually exclusive events.

Example 8 A coin is tossed three times, consider the following events.

A: 'No head appears', B: 'Exactly one head appears' and C: 'Atleast two heads appear'.

Do they form a set of mutually exclusive and exhaustive events?

Solution The sample space of the experiment is

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

and A = {TTT}, B = {HTT, THT, TTH}, C = {HHT, HTH, THH, HHH}

Now

$A \cup B \cup C$ = {TTT, HTT, THT, TTH, HHT, HTH, THH, HHH} = S

Therefore, A, B and C are exhaustive events.

Also, $A \cap B = \phi$, $A \cap C = \phi$ and $B \cap C = \phi$

Therefore, the events are pair-wise disjoint, i.e., they are mutually exclusive.

Hence, A, B and C form a set of mutually exclusive and exhaustive events.

Therefore, $P(C) = \frac{3}{9} = \frac{1}{3}$

(iv) Clearly the event 'not blue' is 'not C'. We know that P(not C) = 1 – P(C)

Therefore $P(\text{not } C) = 1 - \frac{1}{3} = \frac{2}{3}$

(v) The event 'either red or blue' may be described by the set 'A or C'

Since, A and C are mutually exclusive events, we have

$$P(A \text{ or } C) = P(A \cup C) = P(A) + P(C) = \frac{4}{9} + \frac{1}{3} = \frac{7}{9}$$

Example 12 Two students Anil and Ashima appeared in an examination. The probability that Anil will qualify the examination is 0.05 and that Ashima will qualify the examination is 0.10. The probability that both will qualify the examination is 0.02. Find the probability that

(a) Both Anil and Ashima will not qualify the examination.

(b) Atleast one of them will not qualify the examination and

(c) Only one of them will qualify the examination.

Solution Let E and F denote the events that Anil and Ashima will qualify the examination, respectively. Given that

$$P(E) = 0.05,\ P(F) = 0.10 \text{ and } P(E \cap F) = 0.02.$$

Then

(a) The event 'both Anil and Ashima will not qualify the examination' may be expressed as $E' \cap F'$.

Since, E' is 'not E', i.e., Anil will not qualify the examination and F' is 'not F', i.e., Ashima will not qualify the examination.

Also $E' \cap F' = (E \cup F)'$ (by Demorgan's Law)

Now $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

or $P(E \cup F) = 0.05 + 0.10 - 0.02 = 0.13$

Therefore $P(E' \cap F') = P(E \cup F)' = 1 - P(E \cup F) = 1 - 0.13 = 0.87$

(b) P (atleast one of them will not qualify)

= 1 – P(both of them will qualify)

= 1 – 0.02 = 0.98

(c) The event only one of them will qualify the examination is same as the event either (Anil will qualify, and Ashima will not qualify) or (Anil will not qualify and Ashima will qualify) i.e., $E \cap F'$ or $E' \cap F$, where $E \cap F'$ and $E' \cap F$ are mutually exclusive.

Therefore, P(only one of them will qualify) $= P(E \cap F' \text{ or } E' \cap F)$

$$= P(E \cap F') + P(E' \cap F) = P(E) - P(E \cap F) + P(F) - P(E \cap F)$$

$$= 0.05 - 0.02 + 0.10 - 0.02 = 0.11$$

Example 13 A committee of two persons is selected from two men and two women. What is the probability that the committee will have (a) no man? (b) one man? (c) two men?

Solution The total number of persons = 2 + 2 = 4. Out of these four person, two can be selected in ${}^4C_2$ ways.

(a) No men in the committee of two means there will be two women in the committee. Out of two women, two can be selected in ${}^2C_2 = 1$ way.

Therefore $$P(\text{no man}) = \frac{{}^2C_2}{{}^4C_2} = \frac{1 \times 2 \times 1}{4 \times 3} = \frac{1}{6}$$

(b) One man in the committee means that there is one woman. One man out of 2 can be selected in ${}^2C_1$ ways and one woman out of 2 can be selected in ${}^2C_1$ ways. Together they can be selected in ${}^2C_1 \times {}^2C_1$ ways.

Therefore $$P(\text{One man}) = \frac{{}^2C_1 \times {}^2C_1}{{}^4C_2} = \frac{2 \times 2}{2 \times 3} = \frac{2}{3}$$

(c) Two men can be selected in ${}^2C_2$ way.

Hence $$P(\text{Two men}) = \frac{{}^2C_2}{{}^4C_2} = \frac{1}{{}^4C_2} = \frac{1}{6}$$

EXERCISE 16.3

1. Which of the following can not be valid assignment of probabilities for outcomes of sample Space $S = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}$

Conditional Probability

Uptill now in probability, we have discussed the methods of finding the probability of events. If we have two events from the same sample space, does the information about the occurrence of one of the events affect the probability of the other event? Let us try to answer this question by taking up a random experiment in which the outcomes are equally likely to occur.

Consider the experiment of tossing three fair coins. The sample space of the experiment is

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

Since the coins are fair, we can assign the probability $\frac{1}{8}$ to each sample point. Let E be the event 'at least two heads appear' and F be the event 'first coin shows tail'. Then

$$E = \{HHH, HHT, HTH, THH\}$$

and $$F = \{THH, THT, TTH, TTT\}$$

Therefore $$P(E) = P(\{HHH\}) + P(\{HHT\}) + P(\{HTH\}) + P(\{THH\})$$

$$= \frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{1}{2} \text{ (Why ?)}$$

and $$P(F) = P(\{THH\}) + P(\{THT\}) + P(\{TTH\}) + P(\{TTT\})$$

$$= \frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{1}{2}$$

Also $$E \cap F = \{THH\}$$

with $$P(E \cap F) = P(\{THH\}) = \frac{1}{8}$$

Now, suppose we are given that the first coin shows tail, i.e. F occurs, then what is the probability of occurrence of E? With the information of occurrence of F, we are sure that the cases in which first coin does not result into a tail should not be considered while finding the probability of E. This information reduces our sample space from the set S to its subset F for the event E. In other words, the additional information really amounts to telling us that the situation may be considered as being that of a new random experiment for which the sample space consists of all those outcomes only which are favourable to the occurrence of the event F.

Now, the sample point of F which is favourable to event E is THH.

Thus, Probability of E considering F as the sample space = $\frac{1}{4}$,

or Probability of E given that the event F has occurred = $\frac{1}{4}$

This probability of the event E is called the *conditional probability of E given that F has already occurred*, and is denoted by P (E|F).

Thus $$P(E|F) = \frac{1}{4}$$

Note that the elements of F which favour the event E are the common elements of E and F, i.e. the sample points of $E \cap F$.

Thus, we can also write the conditional probability of E given that F has occurred as

$$P(E|F) = \frac{\text{Number of elementary events favourable to } E \cap F}{\text{Number of elementary events which are favourable to } F}$$

$$= \frac{n(E \cap F)}{n(F)}$$

Dividing the numerator and the denominator by total number of elementary events of the sample space, we see that P(E|F) can also be written as

$$P(E|F) = \frac{\frac{n(E \cap F)}{n(S)}}{\frac{n(F)}{n(S)}} = \frac{P(E \cap F)}{P(F)} \qquad \text{... (1)}$$

Note that (1) is valid only when $P(F) \neq 0$ i.e., $F \neq \phi$ (Why?)

Thus, we can define the conditional probability as follows :

Definition 1 If E and F are two events associated with the same sample space of a random experiment, the conditional probability of the event E given that F has occurred, i.e. P (E|F) is given by

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \text{ provided } P(F) \neq 0$$

13.2.1 *Properties of conditional probability*

Let E and F be events of a sample space S of an experiment, then we have

Property 1 $P(S|F) = P(F|F) = 1$

We know that

$$P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$$

Also $$P(F|F) = \frac{P(F \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$$

Thus $$P(S|F) = P(F|F) = 1$$

Property 2 *If* A *and* B *are any two events of a sample space* S *and* F *is an event of* S *such that* $P(F) \neq 0$, *then*

$$P((A \cup B)|F) = P(A|F) + P(B|F) - P((A \cap B)|F)$$

In particular, if A *and* B *are disjoint events, then*

$$P((A\cup B)|F) = P(A|F) + P(B|F)$$

We have

$$P((A\cup B)|F) = \frac{P[(A\cup B)\cap F]}{P(F)}$$

$$= \frac{P[(A\cap F)\cup(B\cap F)]}{P(F)}$$

(by distributive law of union of sets over intersection)

$$= \frac{P(A\cap F)+P(B\cap F)-P(A\cap B\cap F)}{P(F)}$$

$$= \frac{P(A\cap F)}{P(F)}+\frac{P(B\cap F)}{P(F)}-\frac{P[(A\cap B)\cap F]}{P(F)}$$

$$= P(A|F) + P(B|F) - P((A\cap B)|F)$$

When A and B are disjoint events, then

$$P((A\cap B)|F) = 0$$

$$\Rightarrow \quad P((A\cup B)|F) = P(A|F) + P(B|F)$$

Property 3 $P(E'|F) = 1 - P(E|F)$

From Property 1, we know that $P(S|F) = 1$

$\Rightarrow \quad P(E\cup E'|F) = 1 \qquad$ since $S = E\cup E'$

$\Rightarrow \quad P(E|F) + P(E'|F) = 1 \qquad$ since E and E′ are disjoint events

Thus, $\quad P(E'|F) = 1 - P(E|F)$

Let us now take up some examples.

Example 1 If $P(A) = \frac{7}{13}$, $P(B) = \frac{9}{13}$ and $P(A\cap B) = \frac{4}{13}$, evaluate $P(A|B)$.

Solution We have $P(A|B) = \dfrac{P(A\cap B)}{P(B)} = \dfrac{\frac{4}{13}}{\frac{9}{13}} = \dfrac{4}{9}$

Example 2 A family has two children. What is the probability that both the children are boys given that at least one of them is a boy ?

Solution Let b stand for boy and g for girl. The sample space of the experiment is

$$S = \{(b, b), (g, b), (b, g), (g, g)\}$$

Let E and F denote the following events :

E : 'both the children are boys'

F : 'at least one of the child is a boy'

Then $$E = \{(b,b)\} \text{ and } F = \{(b,b), (g,b), (b,g)\}$$

Now $$E \cap F = \{(b,b)\}$$

Thus $$P(F) = \frac{3}{4} \text{ and } P(E \cap F) = \frac{1}{4}$$

Therefore $$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Example 3 Ten cards numbered 1 to 10 are placed in a box, mixed up thoroughly and then one card is drawn randomly. If it is known that the number on the drawn card is more than 3, what is the probability that it is an even number?

Solution Let A be the event 'the number on the card drawn is even' and B be the event 'the number on the card drawn is greater than 3'. We have to find P(A|B).

Now, the sample space of the experiment is S = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}

Then $$A = \{2, 4, 6, 8, 10\}, \; B = \{4, 5, 6, 7, 8, 9, 10\}$$

and $$A \cap B = \{4, 6, 8, 10\}$$

Also $$P(A) = \frac{5}{10}, \; P(B) = \frac{7}{10} \text{ and } P(A \cap B) = \frac{4}{10}$$

Then $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{4}{10}}{\frac{7}{10}} = \frac{4}{7}$$

Example 4 In a school, there are 1000 students, out of which 430 are girls. It is known that out of 430, 10% of the girls study in class XII. What is the probability that a student chosen randomly studies in Class XII given that the chosen student is a girl?

Solution Let E denote the event that a student chosen randomly studies in Class XII and F be the event that the randomly chosen student is a girl. We have to find P (E|F).

Now $\quad P(F) = \frac{430}{1000} = 0.43$ and $P(E \quad F) = \frac{43}{1000} \quad 0.043$ (Why?)

Then $\quad P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{0.043}{0.43} = 0.1$

Example 5 A die is thrown three times. Events A and B are defined as below:

A : 4 on the third throw

B : 6 on the first and 5 on the second throw

Find the probability of A given that B has already occurred.

Solution The sample space has 216 outcomes.

Now
$$A = \begin{Bmatrix} (1,1,4) & (1,2,4) & ... & (1,6,4) & (2,1,4) & (2,2,4) & ... & (2,6,4) \\ (3,1,4) & (3,2,4) & ... & (3,6,4) & (4,1,4) & (4,2,4) & ... & (4,6,4) \\ (5,1,4) & (5,2,4) & ... & (5,6,4) & (6,1,4) & (6,2,4) & ... & (6,6,4) \end{Bmatrix}$$

$$B = \{(6,5,1), (6,5,2), (6,5,3), (6,5,4), (6,5,5), (6,5,6)\}$$

and $\quad A \cap B = \{(6,5,4)\}$.

Now $\quad P(B) = \frac{6}{216}$ and $P(A \cap B) = \frac{1}{216}$

Then
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{216}}{\frac{6}{216}} = \frac{1}{6}$$

Example 6 A die is thrown twice and the sum of the numbers appearing is observed to be 6. What is the conditional probability that the number 4 has appeared at least once?

Solution Let E be the event that 'number 4 appears at least once' and F be the event that 'the sum of the numbers appearing is 6'.

Then, $\quad E = \{(4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (1,4), (2,4), (3,4), (5,4), (6,4)\}$

and $\quad F = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$

We have $\quad P(E) = \frac{11}{36}$ and $P(F) = \frac{5}{36}$

Also $\quad E \cap F = \{(2,4), (4,2)\}$

Therefore $\quad P(E \cap F) = \dfrac{2}{36}$

Hence, the required probability

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{2}{36}}{\frac{5}{36}} = \frac{2}{5}$$

For the conditional probability discussed above, we have considered the elementary events of the experiment to be equally likely and the corresponding definition of the probability of an event was used. However, the same definition can also be used in the general case where the elementary events of the sample space are not equally likely, the probabilities $P(E \cap F)$ and $P(F)$ being calculated accordingly. Let us take up the following example.

Example 7 Consider the experiment of tossing a coin. If the coin shows head, toss it again but if it shows tail, then throw a die. Find the conditional probability of the event that 'the die shows a number greater than 4' given that 'there is at least one tail'.

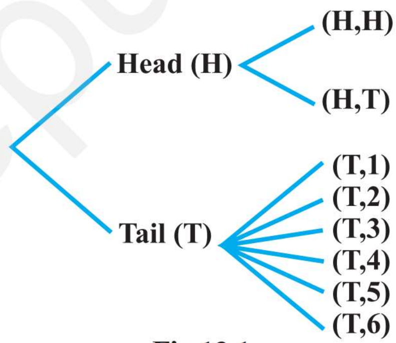


Fig 13.1

Solution The outcomes of the experiment can be represented in following diagrammatic manner called the 'tree diagram'.

The sample space of the experiment may be described as

$$S = \{(H,H), (H,T), (T,1), (T,2), (T,3), (T,4), (T,5), (T,6)\}$$

where (H, H) denotes that both the tosses result into head and (T, i) denote the first toss result into a tail and the number i appeared on the die for $i = 1,2,3,4,5,6$. Thus, the probabilities assigned to the 8 elementary events

(H, H), (H, T), (T, 1), (T, 2), (T, 3) (T, 4), (T, 5), (T, 6)

are $\frac{1}{4}, \frac{1}{4}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}$ respectively which is clear from the Fig 13.2.

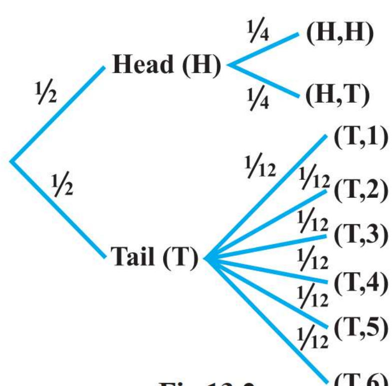


Fig 13.2

Let F be the event that 'there is at least one tail' and E be the event 'the die shows a number greater than 4'. Then

$$F = \{(H,T), (T,1), (T,2), (T,3), (T,4), (T,5), (T,6)\}$$

$$E = \{(T,5), (T,6)\} \text{ and } E \cap F = \{(T,5), (T,6)\}$$

Now
$$P(F) = P(\{(H,T)\}) + P(\{(T,1)\}) + P(\{(T,2)\}) + P(\{(T,3)\}) + P(\{(T,4)\}) + P(\{(T,5)\}) + P(\{(T,6)\})$$

$$= \frac{1}{4} \quad \frac{1}{12} \quad \frac{1}{12} \quad \frac{1}{12} \quad \frac{1}{12} \quad \frac{1}{12} \quad \frac{1}{12} \quad \frac{3}{4}$$

and
$$P(E \cap F) = P(\{(T,5)\}) + P(\{(T,6)\}) = \frac{1}{12} \quad \frac{1}{12} \quad \frac{1}{6}$$

Hence
$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{1}{6}}{\frac{3}{4}} = \frac{2}{9}$$

EXERCISE 13.1

1. Given that E and F are events such that P(E) = 0.6, P(F) = 0.3 and $P(E \cap F) = 0.2$, find P(E|F) and P(F|E)
2. Compute P(A|B), if P(B) = 0.5 and $P(A \cap B) = 0.32$
3. If P(A) = 0.8, P(B) = 0.5 and P(B|A) = 0.4, find
   (i) $P(A \cap B)$ (ii) P(A|B) (iii) $P(A \cup B)$
4. Evaluate $P(A \cup B)$, if $2P(A) = P(B) = \frac{5}{13}$ and $P(A|B) = \frac{2}{5}$
5. If $P(A) = \frac{6}{11}$, $P(B) = \frac{5}{11}$ and $P(A \cup B)$ $\frac{7}{11}$, find
   (i) $P(A \cap B)$ (ii) P(A|B) (iii) P(B|A)

Determine P(E|F) in Exercises 6 to 9.

6. A coin is tossed three times, where
   (i) E : head on third toss , F : heads on first two tosses
   (ii) E : at least two heads , F : at most two heads
   (iii) E : at most two tails , F : at least one tail

Random Variables and Probability Distributions

Random Variables

Suppose that to each point of a sample space we assign a number. We then have a *function* defined on the sample space. This function is called a *random variable* (or *stochastic variable*) or more precisely a *random function* (*stochastic function*). It is usually denoted by a capital letter such as X or Y. In general, a random variable has some specified physical, geometrical, or other significance.

EXAMPLE 2.1 Suppose that a coin is tossed twice so that the sample space is $S = \{HH, HT, TH, TT\}$. Let X represent the number of heads that can come up. With each sample point we can associate a number for X as shown in Table 2-1. Thus, for example, in the case of HH (i.e., 2 heads), $X = 2$ while for TH (1 head), $X = 1$. It follows that X is a random variable.

Table 2-1

Sample Point	*HH*	*HT*	*TH*	*TT*
X	2	1	1	0

It should be noted that many other random variables could also be defined on this sample space, for example, the square of the number of heads or the number of heads minus the number of tails.

A random variable that takes on a finite or countably infinite number of values (see page 4) is called a *discrete random variable* while one which takes on a noncountably infinite number of values is called a *nondiscrete random variable*.

Discrete Probability Distributions

Let X be a discrete random variable, and suppose that the possible values that it can assume are given by $x_1, x_2, x_3, \dots$, arranged in some order. Suppose also that these values are assumed with probabilities given by

$$P(X = x_k) = f(x_k) \qquad k = 1, 2, \dots \tag{1}$$

It is convenient to introduce the *probability function*, also referred to as *probability distribution*, given by

$$P(X = x) = f(x) \tag{2}$$

For $x = x_k$, this reduces to (1) while for other values of x, $f(x) = 0$.

In general, $f(x)$ is a probability function if

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

where the sum in 2 is taken over all possible values of x.

EXAMPLE 2.2 Find the probability function corresponding to the random variable X of Example 2.1. Assuming that the coin is fair, we have

$$P(HH) = \frac{1}{4} \qquad P(HT) = \frac{1}{4} \qquad P(TH) = \frac{1}{4} \qquad P(TT) = \frac{1}{4}$$

Then

$$P(X = 0) = P(TT) = \frac{1}{4}$$

$$P(X = 1) = P(HT \cup TH) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(X = 2) = P(HH) = \frac{1}{4}$$

The probability function is thus given by Table 2-2.

Table 2-2

x	0	1	2
$f(x)$	1/4	1/2	1/4

Distribution Functions for Random Variables

The *cumulative distribution function*, or briefly the *distribution function*, for a random variable X is defined by

$$F(x) = P(X \le x) \tag{3}$$

where x is any real number, i.e., $-\infty < x < \infty$.

The distribution function $F(x)$ has the following properties:

1. $F(x)$ is nondecreasing [i.e., $F(x) \le F(y)$ if $x \le y$].
2. $\lim_{x\to-\infty} F(x) = 0$; $\lim_{x\to\infty} F(x) = 1$.
3. $F(x)$ is continuous from the right [i.e., $\lim_{h\to 0^+} F(x + h) = F(x)$ for all x].

Distribution Functions for Discrete Random Variables

The distribution function for a discrete random variable X can be obtained from its probability function by noting that, for all x in $(-\infty, \infty)$,

$$F(x) = P(X \le x) = \sum_{u \le x} f(u) \tag{4}$$

where the sum is taken over all values u taken on by X for which $u \le x$.

If X takes on only a finite number of values $x_1, x_2, \ldots, x_n$, then the distribution function is given by

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \le x < x_2 \\ f(x_1) + f(x_2) & x_2 \le x < x_3 \\ \vdots & \vdots \\ f(x_1) + \cdots + f(x_n) & x_n \le x < \infty \end{cases} \tag{5}$$

EXAMPLE 2.3 (a) Find the distribution function for the random variable X of Example 2.2. (b) Obtain its graph.

(a) The distribution function is

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ \frac{1}{4} & 0 \le x < 1 \\ \frac{3}{4} & 1 \le x < 2 \\ 1 & 2 \le x < \infty \end{cases}$$

(b) The graph of $F(x)$ is shown in Fig. 2-1.

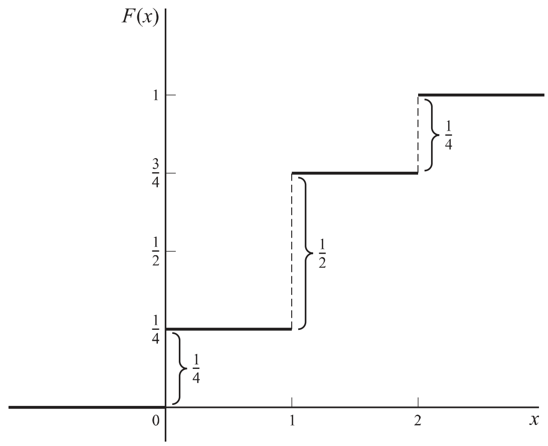


Fig. 2-1

The following things about the above distribution function, which are true in general, should be noted.

1. The magnitudes of the jumps at 0, 1, 2 are $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ which are precisely the probabilities in Table 2-2. This fact enables one to obtain the probability function from the distribution function.
2. Because of the appearance of the graph of Fig. 2-1, it is often called a *staircase function* or *step function*. The value of the function at an integer is obtained from the higher step; thus the value at 1 is $\frac{3}{4}$ and not $\frac{1}{4}$. This is expressed mathematically by stating that the distribution function is *continuous from the right* at 0, 1, 2.
3. As we proceed from left to right (i.e. going *upstairs*), the distribution function either remains the same or increases, taking on values from 0 to 1. Because of this, it is said to be a *monotonically increasing function*.

It is clear from the above remarks and the properties of distribution functions that the probability function of a discrete random variable can be obtained from the distribution function by noting that

$$f(x) = F(x) - \lim_{u \to x^-} F(u). \tag{6}$$

Continuous Random Variables

A nondiscrete random variable X is said to be *absolutely continuous*, or simply *continuous*, if its distribution function may be represented as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du \qquad (-\infty < x < \infty) \tag{7}$$

where the function $f(x)$ has the properties

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

It follows from the above that if X is a continuous random variable, then the probability that X takes on any one particular value is zero, whereas the *interval probability* that X lies *between two different values*, say, a and b, is given by

$$P(a < X < b) = \int_{a}^{b} f(x)\,dx \tag{8}$$

EXAMPLE 2.4 If an individual is selected at random from a large group of adult males, the probability that his height X is precisely 68 inches (i.e., 68.000 . . . inches) would be zero. However, there is a probability greater than zero than X is between 67.000 . . . inches and 68.500 . . . inches, for example.

A function $f(x)$ that satisfies the above requirements is called a *probability function* or *probability distribution* for a continuous random variable, but it is more often called a *probability density function* or simply *density function*. Any function $f(x)$ satisfying Properties 1 and 2 above will automatically be a density function, and required probabilities can then be obtained from (8).

EXAMPLE 2.5 (a) Find the constant c such that the function

$$f(x) = \begin{cases} cx^2 & 0 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

is a density function, and (b) compute $P(1 < X < 2)$.

(a) Since $f(x)$ satisfies Property 1 if $c \geq 0$, it must satisfy Property 2 in order to be a density function. Now

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^3 cx^2\,dx = \left.\frac{cx^3}{3}\right|_0^3 = 9c$$

and since this must equal 1, we have $c = 1/9$.

(b)
$$P(1 < X < 2) = \int_1^2 \frac{1}{9}x^2\,dx = \left.\frac{x^3}{27}\right|_1^2 = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$$

In case $f(x)$ is continuous, which we shall assume unless otherwise stated, the probability that X is equal to any particular value is zero. In such case we can replace either or both of the signs $<$ in (8) by $\leq$. Thus, in Example 2.5,

$$P(1 \leq X \leq 2) = P(1 \leq X < 2) = P(1 < X \leq 2) = P(1 < X < 2) = \frac{7}{27}$$

EXAMPLE 2.6 (a) Find the distribution function for the random variable of Example 2.5. (b) Use the result of (a) to find $P(1 < x \leq 2)$.

(a) We have

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du$$

If $x < 0$, then $F(x) = 0$. If $0 \leq x < 3$, then

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

If $x \geq 3$, then

$$F(x) = \int_0^3 f(u)\,du + \int_3^x f(u)\,du = \int_0^3 \frac{1}{9}u^2\,du + \int_3^x 0\,du = 1$$

Thus the required distribution function is

$$F(x) = \begin{cases} 0 & x < 0 \\ x^3/27 & 0 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

Note that $F(x)$ increases monotonically from 0 to 1 as is required for a distribution function. It should also be noted that $F(x)$ in this case is continuous.

(b) We have

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

as in Example 2.5.

The probability that X is between x and $x + \Delta x$ is given by

$$P(x \leq X \leq x + \Delta x) = \int_x^{x+\Delta x} f(u)\,du \qquad (9)$$

so that if Δx is small, we have approximately

$$P(x \leq X \leq x + \Delta x) = f(x)\Delta x \qquad (10)$$

We also see from (7) on differentiating both sides that

$$\frac{dF(x)}{dx} = f(x) \qquad (11)$$

at all points where $f(x)$ is continuous; i.e., the derivative of the distribution function is the density function.

It should be pointed out that random variables exist that are neither discrete nor continuous. It can be shown that the random variable X with the following distribution function is an example.

$$F(x) = \begin{cases} 0 & x < 1 \\ \dfrac{x}{2} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

In order to obtain (11), we used the basic property

$$\frac{d}{dx}\int_a^x f(u)\,du = f(x) \qquad (12)$$

which is one version of the Fundamental Theorem of Calculus.

Graphical Interpretations

If $f(x)$ is the density function for a random variable X, then we can represent $y = f(x)$ graphically by a curve as in Fig. 2-2. Since $f(x) \geq 0$, the curve cannot fall below the x axis. The entire area bounded by the curve and the x axis must be 1 because of Property 2 on page 36. Geometrically the probability that X is between a and b, i.e., $P(a < X < b)$, is then represented by the area shown shaded, in Fig. 2-2.

The distribution function $F(x) = P(X \leq x)$ is a monotonically increasing function which increases from 0 to 1 and is represented by a curve as in Fig. 2-3.

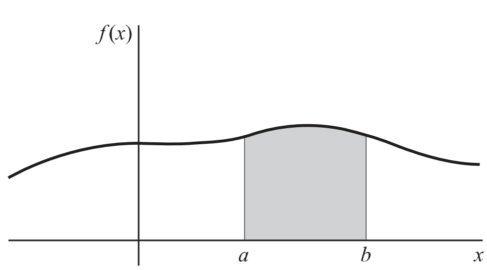


Fig. 2-2

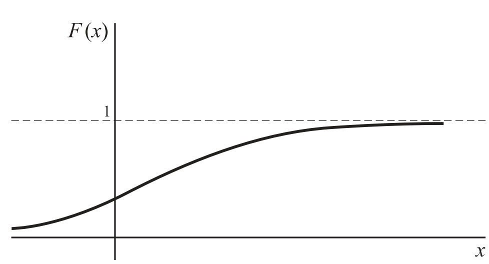


Fig. 2-3

Mathematical Expectation

4.1 Mean of a Random Variable

In Chapter 1, we discussed the sample mean, which is the arithmetic mean of the data. Now consider the following. If two coins are tossed 16 times and X is the number of heads that occur per toss, then the values of X are 0, 1, and 2. Suppose that the experiment yields no heads, one head, and two heads a total of 4, 7, and 5 times, respectively. The average number of heads per toss of the two coins is then

$$\frac{(0)(4) + (1)(7) + (2)(5)}{16} = 1.06.$$

This is an average value of the data and yet it is not a possible outcome of $\{0, 1, 2\}$. Hence, an average is not necessarily a possible outcome for the experiment. For instance, a salesman's average monthly income is not likely to be equal to any of his monthly paychecks.

Let us now restructure our computation for the average number of heads so as to have the following equivalent form:

$$(0)\left(\frac{4}{16}\right) + (1)\left(\frac{7}{16}\right) + (2)\left(\frac{5}{16}\right) = 1.06.$$

The numbers 4/16, 7/16, and 5/16 are the fractions of the total tosses resulting in 0, 1, and 2 heads, respectively. These fractions are also the relative frequencies for the different values of X in our experiment. In fact, then, we can calculate the mean, or average, of a set of data by knowing the distinct values that occur and their relative frequencies, without any knowledge of the total number of observations in our set of data. Therefore, if 4/16, or 1/4, of the tosses result in no heads, 7/16 of the tosses result in one head, and 5/16 of the tosses result in two heads, the mean number of heads per toss would be 1.06 no matter whether the total number of tosses were 16, 1000, or even 10,000.

This method of relative frequencies is used to calculate the average number of heads per toss of two coins that we might expect in the long run. We shall refer to this average value as the **mean of the random variable** X or the **mean of the probability distribution of** X and write it as μ_x or simply as μ when it is clear to which random variable we refer. It is also common among statisticians to refer to this mean as the mathematical expectation, or the expected value of the random variable X, and denote it as $E(X)$.

Assuming that 1 fair coin was tossed twice, we find that the sample space for our experiment is

$$S = \{HH, HT, TH, TT\}.$$

Since the 4 sample points are all equally likely, it follows that

$$P(X = 0) = P(TT) = \frac{1}{4}, \quad P(X = 1) = P(TH) + P(HT) = \frac{1}{2},$$

and

$$P(X = 2) = P(HH) = \frac{1}{4},$$

where a typical element, say TH, indicates that the first toss resulted in a tail followed by a head on the second toss. Now, these probabilities are just the relative frequencies for the given events in the long run. Therefore,

$$\mu = E(X) = (0)\left(\frac{1}{4}\right) + (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{4}\right) = 1.$$

This result means that a person who tosses 2 coins over and over again will, on the average, get 1 head per toss.

The method described above for calculating the expected number of heads per toss of 2 coins suggests that the mean, or expected value, of any discrete random variable may be obtained by multiplying each of the values $x_1, x_2, \ldots, x_n$ of the random variable X by its corresponding probability $f(x_1), f(x_2), \ldots, f(x_n)$ and summing the products. This is true, however, only if the random variable is discrete. In the case of continuous random variables, the definition of an expected value is essentially the same with summations replaced by integrations.

Definition 4.1: Let X be a random variable with probability distribution $f(x)$. The **mean**, or **expected value**, of X is

$$\mu = E(X) = \sum_x x f(x)$$

if X is discrete, and

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

if X is continuous.

The reader should note that the way to calculate the expected value, or mean, shown here is different from the way to calculate the sample mean described in Chapter 1, where the sample mean is obtained by using data. In mathematical expectation, the expected value is calculated by using the probability distribution.

However, the mean is usually understood as a "center" value of the underlying distribution if we use the expected value, as in Definition 4.1.

Example 4.1: A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

***Solution*:** Let X represent the number of good components in the sample. The probability distribution of X is

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \qquad x = 0, 1, 2, 3.$$

Simple calculations yield $f(0) = 1/35$, $f(1) = 12/35$, $f(2) = 18/35$, and $f(3) = 4/35$. Therefore,

$$\mu = E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \frac{12}{7} = 1.7.$$

Thus, if a sample of size 3 is selected at random over and over again from a lot of 4 good components and 3 defective components, it will contain, on average, 1.7 good components.

Example 4.2: A salesperson for a medical device company has two appointments on a given day. At the first appointment, he believes that he has a 70% chance to make the deal, from which he can earn \$1000 commission if successful. On the other hand, he thinks he only has a 40% chance to make the deal at the second appointment, from which, if successful, he can make \$1500. What is his expected commission based on his own probability belief? Assume that the appointment results are independent of each other.

***Solution*:** First, we know that the salesperson, for the two appointments, can have 4 possible commission totals: \$0, \$1000, \$1500, and \$2500. We then need to calculate their associated probabilities. By independence, we obtain

$$f(\$0) = (1 - 0.7)(1 - 0.4) = 0.18, \quad f(\$2500) = (0.7)(0.4) = 0.28,$$
$$f(\$1000) = (0.7)(1 - 0.4) = 0.42, \text{ and } f(\$1500) = (1 - 0.7)(0.4) = 0.12.$$

Therefore, the expected commission for the salesperson is

$$E(X) = (\$0)(0.18) + (\$1000)(0.42) + (\$1500)(0.12) + (\$2500)(0.28) = \$1300.$$

Examples 4.1 and 4.2 are designed to allow the reader to gain some insight into what we mean by the expected value of a random variable. In both cases the random variables are discrete. We follow with an example involving a continuous random variable, where an engineer is interested in the *mean life* of a certain type of electronic device. This is an illustration of a *time to failure* problem that occurs often in practice. The expected value of the life of a device is an important parameter for its evaluation.

Example 4.3: Let X be the random variable that denotes the life in hours of a certain electronic device. The probability density function is

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

Find the expected life of this type of device.

Solution: Using Definition 4.1, we have

$$\mu = E(X) = \int_{100}^{\infty} x \frac{20,000}{x^3}\, dx = \int_{100}^{\infty} \frac{20,000}{x^2}\, dx = 200.$$

Therefore, we can expect this type of device to last, *on average*, 200 hours.

Now let us consider a new random variable $g(X)$, which depends on X; that is, each value of $g(X)$ is determined by the value of X. For instance, $g(X)$ might be X^2 or $3X - 1$, and whenever X assumes the value 2, $g(X)$ assumes the value $g(2)$. In particular, if X is a discrete random variable with probability distribution $f(x)$, for $x = -1, 0, 1, 2$, and $g(X) = X^2$, then

$$\begin{aligned} P[g(X) = 0] &= P(X = 0) = f(0), \\ P[g(X) = 1] &= P(X = -1) + P(X = 1) = f(-1) + f(1), \\ P[g(X) = 4] &= P(X = 2) = f(2), \end{aligned}$$

and so the probability distribution of $g(X)$ may be written

$g(x)$	0	1	4
$P[g(X) = g(x)]$	$f(0)$	$f(-1) + f(1)$	$f(2)$

By the definition of the expected value of a random variable, we obtain

$$\begin{aligned} \mu_{g(X)} &= E[g(x)] = 0f(0) + 1[f(-1) + f(1)] + 4f(2) \\ &= (-1)^2 f(-1) + (0)^2 f(0) + (1)^2 f(1) + (2)^2 f(2) = \sum_x g(x)f(x). \end{aligned}$$

This result is generalized in Theorem 4.1 for both discrete and continuous random variables.

Theorem 4.1: Let X be a random variable with probability distribution $f(x)$. The expected value of the random variable $g(X)$ is

$$\mu_{g(X)} = E[g(X)] = \sum_x g(x)f(x)$$

if X is discrete, and

$$\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\, dx$$

if X is continuous.

Example 4.4: Suppose that the number of cars X that pass through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

x	4	5	6	7	8	9
$P(X=x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(X) = 2X - 1$ represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

Solution: By Theorem 4.1, the attendant can expect to receive

$$\begin{aligned} E[g(X)] = E(2X-1) &= \sum_{x=4}^{9}(2x-1)f(x) \\ &= (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right) \\ &\quad + (15)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right) = \$12.67. \end{aligned}$$

Example 4.5: Let X be a random variable with density function

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

Find the expected value of $g(X) = 4X + 3$.

Solution: By Theorem 4.1, we have

$$E(4X+3) = \int_{-1}^{2} \frac{(4x+3)x^2}{3}\, dx = \frac{1}{3}\int_{-1}^{2}(4x^3+3x^2)\, dx = 8.$$

We shall now extend our concept of mathematical expectation to the case of two random variables X and Y with joint probability distribution $f(x, y)$.

Definition 4.2: Let X and Y be random variables with joint probability distribution $f(x, y)$. The mean, or expected value, of the random variable $g(X, Y)$ is

$$\mu_{g(X,Y)} = E[g(X,Y)] = \sum_x \sum_y g(x,y)f(x,y)$$

if X and Y are discrete, and

$$\mu_{g(X,Y)} = E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\, dx\, dy$$

if X and Y are continuous.

Generalization of Definition 4.2 for the calculation of mathematical expectations of functions of several random variables is straightforward.

Example 4.6: Let X and Y be the random variables with joint probability distribution indicated in Table 3.1 on page 96. Find the expected value of $g(X,Y) = XY$. The table is reprinted here for convenience.

	$f(x,y)$	x = 0	1	2	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

***Solution*:** By Definition 4.2, we write

$$\begin{aligned} E(XY) &= \sum_{x=0}^{2}\sum_{y=0}^{2} xyf(x,y) \\ &= (0)(0)f(0,0) + (0)(1)f(0,1) \\ &\quad + (1)(0)f(1,0) + (1)(1)f(1,1) + (2)(0)f(2,0) \\ &= f(1,1) = \frac{3}{14}. \end{aligned}$$

Example 4.7: Find $E(Y/X)$ for the density function

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

***Solution*:** We have

$$E\left(\frac{Y}{X}\right) = \int_0^1 \int_0^2 \frac{y(1+3y^2)}{4}\,dxdy = \int_0^1 \frac{y+3y^3}{2}\,dy = \frac{5}{8}.$$

Note that if $g(X,Y) = X$ in Definition 4.2, we have

$$E(X) = \begin{cases} \sum\limits_x \sum\limits_y xf(x,y) = \sum\limits_x xg(x) & \text{(discrete case),} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dy\,dx = \int_{-\infty}^{\infty} xg(x)\,dx & \text{(continuous case),} \end{cases}$$

where $g(x)$ is the marginal distribution of X. Therefore, in calculating $E(X)$ over a two-dimensional space, one may use either the joint probability distribution of X and Y or the marginal distribution of X. Similarly, we define

$$E(Y) = \begin{cases} \sum\limits_y \sum\limits_x yf(x,y) = \sum\limits_y yh(y) & \text{(discrete case),} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\,dxdy = \int_{-\infty}^{\infty} yh(y)\,dy & \text{(continuous case),} \end{cases}$$

where $h(y)$ is the marginal distribution of the random variable Y.

4.2 Variance and Covariance of Random Variables

The mean, or expected value, of a random variable X is of special importance in statistics because it describes where the probability distribution is centered. By itself, however, the mean does not give an adequate description of the shape of the distribution. We also need to characterize the variability in the distribution. In Figure 4.1, we have the histograms of two discrete probability distributions that have the same mean, $\mu = 2$, but differ considerably in variability, or the dispersion of their observations about the mean.

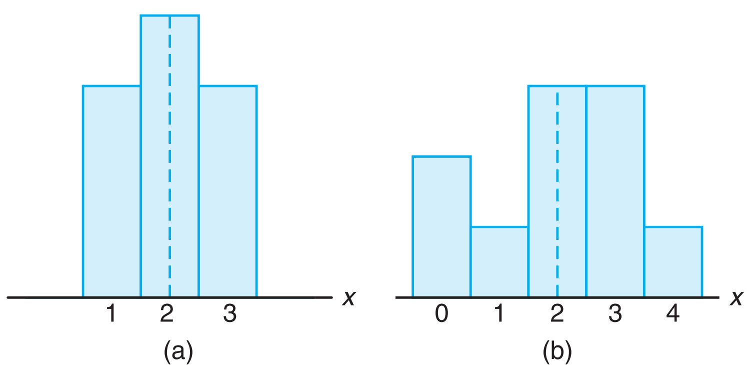


Figure 4.1: Distributions with equal means and unequal dispersions.

The most important measure of variability of a random variable X is obtained by applying Theorem 4.1 with $g(X) = (X - \mu)^2$. The quantity is referred to as the **variance of the random variable** X or the **variance of the probability**

distribution of X and is denoted by $\text{Var}(X)$ or the symbol σ_X^2, or simply by σ^2 when it is clear to which random variable we refer.

Definition 4.3: Let X be a random variable with probability distribution $f(x)$ and mean μ. The variance of X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x), \qquad \text{if } X \text{ is discrete, and}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx, \qquad \text{if } X \text{ is continuous.}$$

The positive square root of the variance, σ, is called the **standard deviation** of X.

The quantity $x-\mu$ in Definition 4.3 is called the **deviation of an observation** from its mean. Since the deviations are squared and then averaged, σ^2 will be much smaller for a set of x values that are close to μ than it will be for a set of values that vary considerably from μ.

Example 4.8: Let the random variable X represent the number of automobiles that are used for official business purposes on any given workday. The probability distribution for company A [Figure 4.1(a)] is

x	1	2	3
$f(x)$	0.3	0.4	0.3

and that for company B [Figure 4.1(b)] is

x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

Show that the variance of the probability distribution for company B is greater than that for company A.

Solution: For company A, we find that

$$\mu_A = E(X) = (1)(0.3) + (2)(0.4) + (3)(0.3) = 2.0,$$

and then

$$\sigma_A^2 = \sum_{x=1}^{3} (x-2)^2 = (1-2)^2(0.3) + (2-2)^2(0.4) + (3-2)^2(0.3) = 0.6.$$

For company B, we have

$$\mu_B = E(X) = (0)(0.2) + (1)(0.1) + (2)(0.3) + (3)(0.3) + (4)(0.1) = 2.0,$$

and then

$$\begin{aligned}\sigma_B^2 &= \sum_{x=0}^{4} (x-2)^2 f(x)\\ &= (0-2)^2(0.2) + (1-2)^2(0.1) + (2-2)^2(0.3)\\ &\quad + (3-2)^2(0.3) + (4-2)^2(0.1) = 1.6.\end{aligned}$$

Clearly, the variance of the number of automobiles that are used for official business purposes is greater for company B than for company A.

An alternative and preferred formula for finding σ^2, which often simplifies the calculations, is stated in the following theorem.

Theorem 4.2: The variance of a random variable X is

$$\sigma^2 = E(X^2) - \mu^2.$$

***Proof*:** For the discrete case, we can write

$$\begin{aligned}\sigma^2 &= \sum_x (x-\mu)^2 f(x) = \sum_x (x^2 - 2\mu x + \mu^2) f(x) \\ &= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x).\end{aligned}$$

Since $\mu = \sum_x x f(x)$ by definition, and $\sum_x f(x) = 1$ for any discrete probability distribution, it follows that

$$\sigma^2 = \sum_x x^2 f(x) - \mu^2 = E(X^2) - \mu^2.$$

For the continuous case the proof is step by step the same, with summations replaced by integrations.

Example 4.9: Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X.

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

Using Theorem 4.2, calculate σ^2.

***Solution*:** First, we compute

$$\mu = (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61.$$

Now,

$$E(X^2) = (0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87.$$

Therefore,

$$\sigma^2 = 0.87 - (0.61)^2 = 0.4979.$$

Example 4.10: The weekly demand for a drinking-water product, in thousands of liters, from a local chain of efficiency stores is a continuous random variable X having the probability density

$$f(x) = \begin{cases} 2(x-1), & 1 < x < 2, \\ 0, & \text{elsewhere}. \end{cases}$$

Find the mean and variance of X.

***Solution*:** Calculating $E(X)$ and $E(X^2$, we have

$$\mu = E(X) = 2\int_1^2 x(x-1)\ dx = \frac{5}{3}$$

and

$$E(X^2) = 2\int_1^2 x^2(x-1)\ dx = \frac{17}{6}.$$

Therefore,

$$\sigma^2 = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}.$$

At this point, the variance or standard deviation has meaning only when we compare two or more distributions that have the same units of measurement. Therefore, we could compare the variances of the distributions of contents, measured in liters, of bottles of orange juice from two companies, and the larger value would indicate the company whose product was more variable or less uniform. It would not be meaningful to compare the variance of a distribution of heights to the variance of a distribution of aptitude scores. In Section 4.4, we show how the standard deviation can be used to describe a single distribution of observations.

We shall now extend our concept of the variance of a random variable X to include random variables related to X. For the random variable $g(X)$, the variance is denoted by $\sigma^2_{g(X)}$ and is calculated by means of the following theorem.

Theorem 4.3: Let X be a random variable with probability distribution $f(x)$. The variance of the random variable $g(X)$ is

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \sum_x [g(x) - \mu_{g(X)}]^2 f(x)$$

if X is discrete, and

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x)\ dx$$

if X is continuous.

***Proof*:** Since $g(X)$ is itself a random variable with mean $\mu_{g(X)}$ as defined in Theorem 4.1, it follows from Definition 4.3 that

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]\}.$$

Now, applying Theorem 4.1 again to the random variable $[g(X)-\mu_{g(X)}]^2$ completes the proof.

Example 4.11: Calculate the variance of $g(X) = 2X + 3$, where X is a random variable with probability distribution

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

***Solution*:** First, we find the mean of the random variable $2X+3$. According to Theorem 4.1,

$$\mu_{2X+3} = E(2X+3) = \sum_{x=0}^{3}(2x+3)f(x) = 6.$$

Now, using Theorem 4.3, we have

$$\sigma^2_{2X+3} = E\{[(2X+3) - \mu_{2x+3}]^2\} = E[(2X+3-6)^2]$$
$$= E(4X^2 - 12X + 9) = \sum_{x=0}^{3}(4x^2 - 12x + 9)f(x) = 4.$$

Example 4.12: Let X be a random variable having the density function given in Example 4.5 on page 115. Find the variance of the random variable $g(X) = 4X + 3$.

***Solution*:** In Example 4.5, we found that $\mu_{4X+3} = 8$. Now, using Theorem 4.3,

$$\sigma^2_{4X+3} = E\{[(4X+3) - 8]^2\} = E[(4X-5)^2]$$
$$= \int_{-1}^{2}(4x-5)^2\frac{x^2}{3}\,dx = \frac{1}{3}\int_{-1}^{2}(16x^4 - 40x^3 + 25x^2)\,dx = \frac{51}{5}.$$

If $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$, where $\mu_X = E(X)$ and $\mu_Y = E(Y)$, Definition 4.2 yields an expected value called the **covariance** of X and Y, which we denote by σ_{XY} or $\mathrm{Cov}(X,Y)$.

Basic Concepts

Introduce several basic vocabulary words used in studying statistics: *statistic, population*, *variable*.

- Statistic: is the science of collecting studies to collect, organize, summarize, analyze, and draw conclusions from data.

  - A variable is a characteristic or attribute that can assume different values.
  - Data: are the values that a variable can assume.
  - Random Variable: variables whose determined by chance.
  - Data set: Collection of data values.

Branches of statistics

There are two branches:

1- **Descriptive Statistic:** consists of the collection, organization, summarization, and presentation of data. For example the average age of the student is 14 years.

2- **Inferential statistics:** consists of generalizing from samples to populations, performing estimations and hypothesis testing, determining predictions. For example the relation between smoking and lung cancer.

- Population

  A population: consists of all subjects (human or otherwise) that are being studied.

- Sample

  A sample: is a group of subjects selected from a population.

- Discrete variable

  Discrete variable: Assume values that can be counted.
  Examples: number of students present or students’ grade level

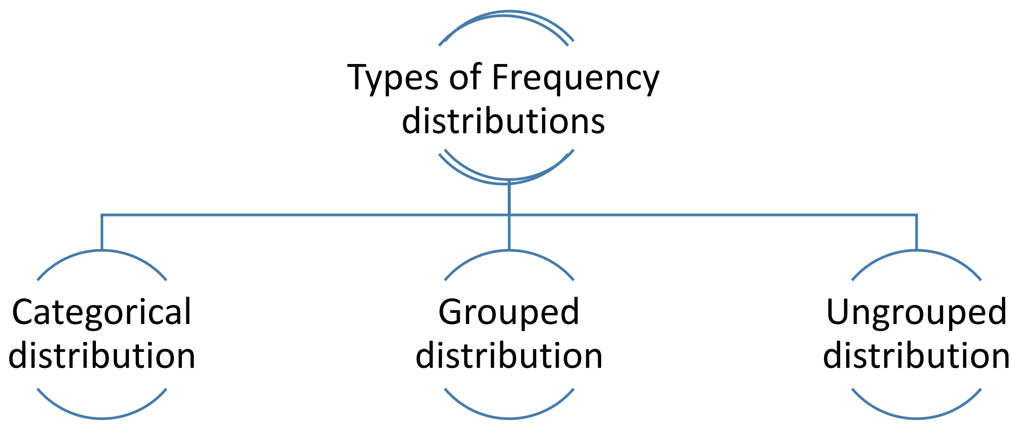
- Continuous variable

  Continuous variable: can assume all values between any two specific values. They are obtained by measuring.
  Examples: height of students in class, weight of students in class, time it takes to get to school, or distance traveled between classes.

- Computer in Statistics:

Data Organization

➢ Frequency distribution table



When data are collected in original form, they are called raw data. For example: row data

2	5	8	7	2	2
6	8	5	2	5	7
4	5	6	2	8	6

A frequency distribution is the organization of raw data in table form, using classes and frequencies. The researches organized the raw data into

Score	*f*
8	3
7	2
6	3
5	4
4	1
2	5

Categorical Frequency Distribution

Categorical Frequency Distribution: can be used for data that can be placed in specific categories, such as nominal- or ordinal-level data.

Example: Twenty-five army indicates were given a blood test to determine their blood type.

Raw Data: A,B,B,AB,O O,O,B,AB,B B,B,O,A,O A,O,O,O,AB AB,A,O,B,A

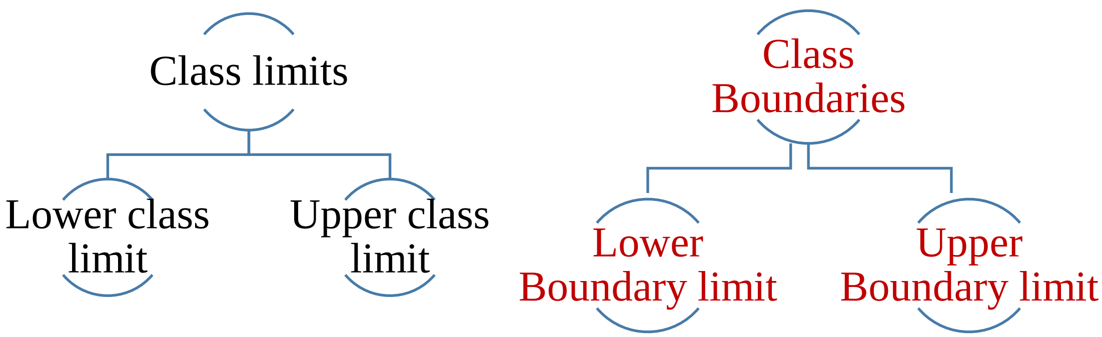
Percent= $\frac{f}{n}*100$

Class	Tally	Frequency (f)	Percent
A	~~IIII~~	5	20
B	~~IIII~~ II	7	28
O	~~IIII~~ IIII	9	36
AB	IIII	4	16
		n=25	100

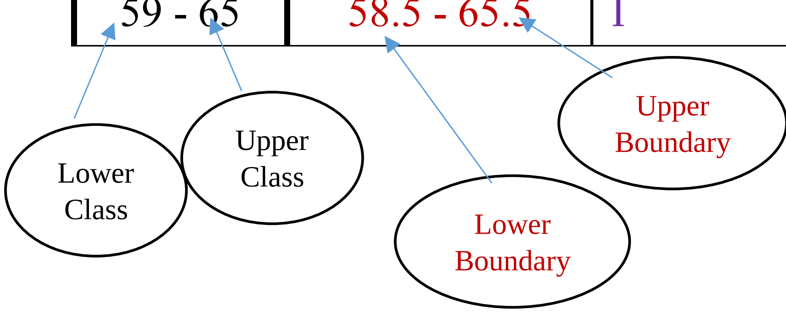
Grouped Frequency Distribution

- **Grouped frequency distributions** can be used when the range of values in the data set is very large. The data must be grouped into classes that are more than one unit in width. For example the life of boat batteries in hours.
- The smallest and largest possible data values in a class are the ***lower*** and ***upper class limits***. ***Class boundaries*** separate the classes.
- To find a class boundary, average the upper class limit of one class and the lower class limit of the next class.

- The **class width** can be calculated by subtracting
    - ☐ successive lower class limits (or boundaries)
    - ☐ successive upper class limits (or boundaries)
    - ☐ upper and lower class boundaries
- The ***class midpoint X_m*** can be calculated by averaging
    - ☐ upper and lower class limits (or boundaries)



Class limits	Class Boundaries	Tally	Frequency (f)
24 - 30	23.5 - 30.5	III	3
31 - 37	30.5 - 37.5	I	1
38 - 44	37.5 - 44.5	~~IIII~~	5
45 - 51	44.5 - 51.5	~~IIII~~ IIII	9
52 - 58	51.5 - 58.5	~~IIII~~ I	6
59 - 65	58.5 - 65.5	I	1



- In the life of boat batteries example, the values 24 and 30 of the first class are the **class limits.**
- The **lower class** limit is 24 and the **upper class** limit is 30.
- **The *Class boundaries*** are used to separate the classes. So that there are no gaps in the frequency distribution

➢ Lower boundary= lower limit - 0.5

➢ Upper boundary= upper limit + 0.5

- Class limits should have the same decimal place value as the data, but the class boundaries should have one additional place value and end in a 5.

For example: Class limit 7.8 – 8.8

Class boundary 7.75 – 8.85

➢ Lower boundary= lower limit - 0.05

=7.8- 0.05 =7.75

➢ Upper boundary= upper limit + 0.05

– =8.8+0.05=8.85

Class width = lower of second class limit- lower of first class limit

Or

Class width = Upper of second class limit- Upper of first class limit

Class width: 31 – 24 =7

The class midpoint X_m is found by adding the lower and upper class limit (or boundary) and dividing by 2.

$$X_m = \frac{lower\ limit + upper\ limit}{2}$$

Or

$$X_m = \frac{lower\ boundary + upper\ boundary}{2}$$

For Example: $\frac{24+30}{2} = 27$, $\frac{23.5 + 30.5}{2} = 27$

- Find the boundaries for the following class limits:
    - 44 - 37
    - 10.3 - 11.5
    - 22.2 – 23.0
    - 547.04 - 553.20
- Find the class width for the following class limits:
    - 37 – 44
    - 45 – 52
    - 625 – 654
    - 655 - 684
- Find the class width for the following class boundaries:
    - 10.5 – 11.5
    - 22.15 – 27.15

Rules for Classes in Grouped Frequency Distributions

1. There should be 5-20 classes.
2. The class width should be an odd number.
3. The classes must be mutually exclusive.

Age
10 – 20
20 – 30
30 – 40
40 – 50

Better wat to construct a frequency distribution

Age
10 – 20
21 – 31
32 – 42
43 – 53

4. The classes must be continuous.
5. The classes must be exhaustive.
6. The classes must be equal in width (except in open-ended distributions).

Procedure for Constructing a Grouped Frequency Distribution

- STEP 1 Determine the classes.
  - ✓ Find the highest and lowest value
  - ✓ Find the range
  - ✓ Select the number of classes desired.
  - ✓ Find the width by divided the range by the number of classes and rounding up.
  - ✓ Select a starting point (usually the lowest value), add the width to get the lower limits.
  - ✓ Find the upper class limits.
  - ✓ Find the boundaries.
- STEP 2 Tally the data.
- STEP 3 Find the frequencies.
- STEP 4 Find the cumulative frequencies by keeping a running total of the frequencies.

Constructing a Grouped Frequency Distribution

Example

The following data represent the record high temperatures for each of the 50 states. Construct a grouped frequency distribution for the data using 7 classes.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

STEP 1 Determine the classes. Find the class width by dividing the range by the number of classes 7.

Range = High – Low

= 134 – 100 = 34

Width = $\frac{\text{Range}}{7} = \frac{34}{7} = 5$

Note: Rounding Rule: Always round up if a remainder.

STEP 2 Tally the data.

STEP 3 Find the frequencies.

Class Limits	Class Boundaries	Frequency	Cumulative Frequency
100 - 104	99.5 - 104.5	2	
105 - 109	104.5 - 109.5	8	
110 - 114	109.5 - 114.5	18	
115 - 119	114.5 - 119.5	13	
120 - 124	119.5 - 124.5	7	
125 - 129	124.5 - 129.5	1	
130 - 134	129.5 - 134.5	1	

STEP 4 Find the cumulative frequencies by keeping a running total of the frequencies.

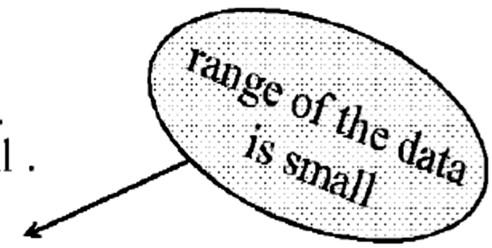
Class Limits	Class Boundaries	Frequency	Cumulative Frequency
100 - 104	99.5 - 104.5	2	2
105 - 109	104.5 - 109.5	8	10
110 - 114	109.5 - 114.5	18	28
115 - 119	114.5 - 119.5	13	41
120 - 124	119.5 - 124.5	7	48
125 - 129	124.5 - 129.5	1	49
130 - 134	129.5 - 134.5	1	50

Ungrouped Frequency distribution:

Example : The data shown here represent the number of miles per gallon that 30 selected four-wheel- drive sports utility vehicles obtained in city driving.

12	17	12	14	16	18
16	18	12	16	17	15
15	16	12	15	16	16
12	14	15	12	15	15
19	13	16	18	16	14

STEP 1 Determine the classes.
The rang of the data set is small .



Range = High – Low

$= 19 - 12 = 7$

So the class consisting of the single data value can be used. They are 12,13,14,15,16,17,18,19.

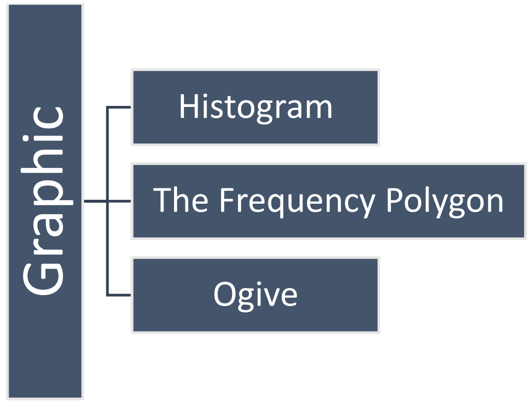
- This type of distribution is called ungrouped frequency distribution

STEP 2 Tally the data.
STEP 3 Find the frequencies.

Class Limits	Class Boundaries	Frequency	Cumulative Frequency
12	11.5-12.5	6	0
13	12.5-13.5	1	6
14	13.5-14.5	3	7
15	14.5-15.5	6	10
16	15.5-16.5	8	16
17	16.5-17.5	2	24
18	17.5-18.5	3	26
19	18.5-19.5	1	29
			30

Graphic

The three most commonly used graphs in research are:



Purpose of graphs in statistics is to convey the **data** to the viewers **in pictorial form**

- **Easier** for most people to understand the **meaning of data** in form of graphs
- They can also be used to discover a **trend or pattern** in a situation over a period of time
- Useful in **getting** the audience's **attention** in a publication or a speaking presentation

➢ Histogram

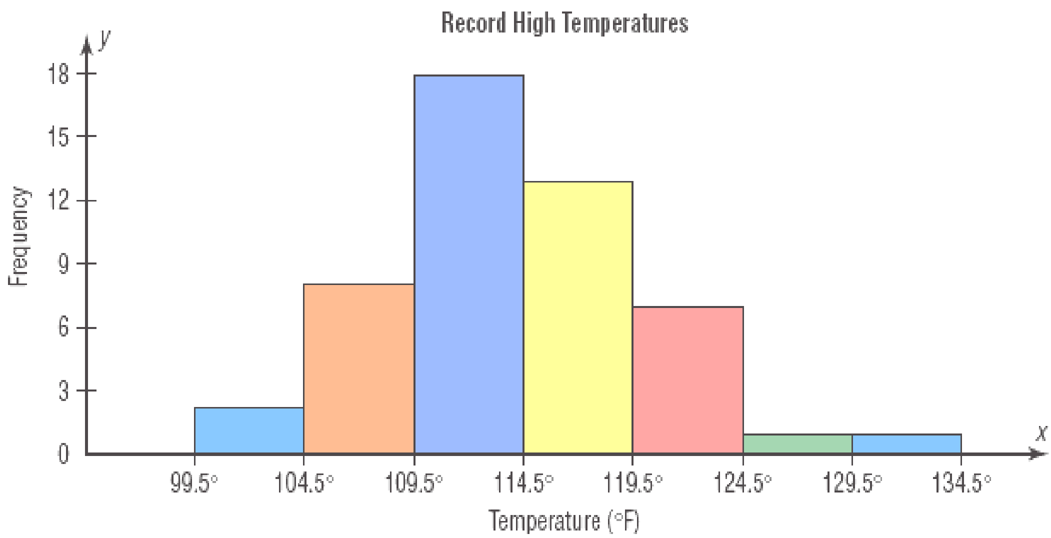
The histogram is a graph that displays the data by using contiguous vertical bars (unless the frequency of a class is 0) of various heights to represent the frequencies of the classes.

❑ **The <u>class boundaries</u> are represented on the horizontal axis**

Example 2-4: Construct a histogram to represent the data for the record high temperatures for each of the 50 states (see Example 2–2 for the data).

Class Limits	Class Boundaries	Frequency
100 - 104	99.5 - 104.5	2
105 - 109	104.5 - 109.5	8
110 - 114	109.5 - 114.5	18
115 - 119	114.5 - 119.5	13
120 - 124	119.5 - 124.5	7
125 - 129	124.5 - 129.5	1
130 - 134	129.5 - 134.5	1

Histograms use class boundaries and frequencies of the classes



➢ Frequency Polygons

- The ***frequency polygon*** is a graph that displays the data by using lines that connect points plotted for the frequencies at the class midpoints. The frequencies are represented by the heights of the points.

- The ~~class midpoints~~ are represented on the horizontal axis.

Construct a frequency polygon to represent the data for the record high temperatures for each of the 50 states.

Step 1: find the midpoints of each class (Recall that midpoints are found by adding the upper and lower boundaries and dividing by 2).

$\frac{99.5+104.5}{2}=102$ $\frac{104.5+109.5}{2}=107$

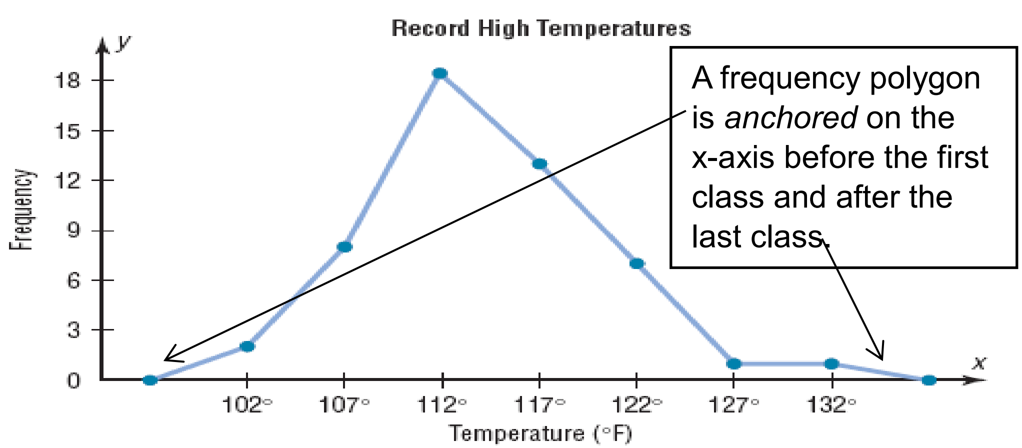
Step 2 Draw the x and y axes. Label the x axis with the midpoint of each class, and then use a suitable scale on the y axis for the frequencies.

Step 3 Using the midpoints for the x values and the frequencies as the y values, plot the points.

Step 4 Connect adjacent points with line segments. Draw a line back to the x axis at the beginning and end of the graph, at the same distance that the previous and next midpoints would be located, as shown in Figure 2–3.

Frequency polygons use class midpoints and frequencies of the classes.

Class Limits	Class Midpoints	Frequency
100 - 104	102	2
105 - 109	107	8
110 - 114	112	18
115 - 119	117	13
120 - 124	122	7
125 - 129	127	1
130 - 134	132	1



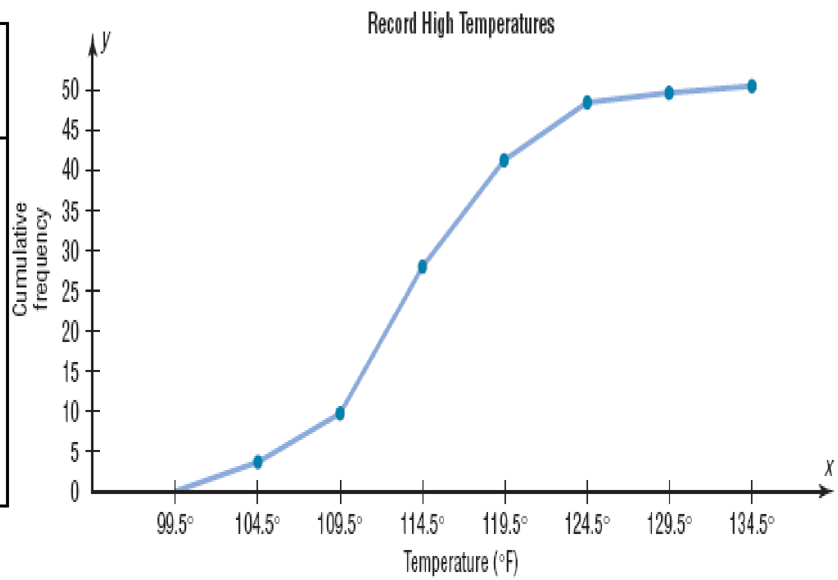
- Ogive
  - The ***ogive*** is a graph that represents the cumulative frequencies for the classes in a frequency distribution.
  - The ~~upper class boundaries~~ are represented on the horizontal axis.

Construct an ogive to represent the data for the record high temperatures for each of the 50 states.

Ogives use upper class boundaries and cumulative frequencies of the classes.

Class Limits	Class Boundaries	Frequency	Cumulative Frequency
100 - 104	99.5 - 104.5	2	2
105 - 109	104.5 - 109.5	8	10
110 - 114	109.5 - 114.5	18	28
115 - 119	114.5 - 119.5	13	41
120 - 124	119.5 - 124.5	7	48
125 - 129	124.5 - 129.5	1	49
130 - 134	129.5 - 134.5	1	50

Class Boundaries	Cumulative Frequency
Less than 104.5	2
Less than 109.5	10
Less than 114.5	28
Less than 119.5	41
Less than 124.5	48
Less than 129.5	49
Less than 134.5	50



➢ Pareto charts

When the variable displayed on the horizontal axis is qualitative or categorical, a *Pareto chart* can also be used to represent the data.
A **Pareto chart** is used to represent a frequency distribution for a categorical variable, and the frequencies are displayed by the heights of vertical bars, which are arranged in order from highest to lowest.

The data shown here consist of the number of homeless people for a sample of selected cities. Construct and analyze a Pareto chart for the data.

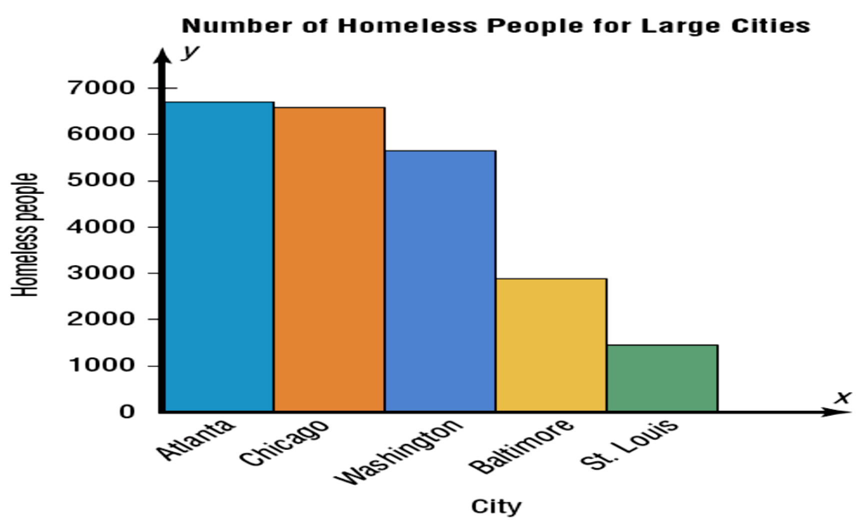
City	**Number**
Atlanta	6832
Baltimore	2904
Chicago	6680
St. Louis	1485
Washington	5518

Solution

Step 1 Arrange the data from the largest to smallest according to frequency.

City	**Number**
Atlanta	6832
Chicago	6680
Washington	5518
Baltimore	2904
St. Louis	1485

Step 2 Draw and label the *x* and *y* axes.
Step 3 Draw the bars corresponding to the frequencies. The graph shows that the number of homeless people is about the same for Atlanta and Chicago and a lot less for Baltimore and St. Louis.



➢ The Pie Graph

Pie graphs are used extensively in statistics. The purpose of the pie graph is to show the relationship of the parts to the whole by visually comparing the sizes of the sections. Percentages or proportions can be used. The variable is nominal or categorical.
A **pie graph** is a circle that is divided into sections or wedges according to the percentage of frequencies in each category of the distribution.

This frequency distribution shows the number of pounds of each snack food eaten during the Super Bowl. Construct a pie graph for the data.

Snack	**Pounds (frequency)**
Potato chips	11.2 million
Tortilla chips	8.2 million
Pretzels	4.3 million
Popcorn	3.8 million
Snack nuts	2.5 million
	Total n = 30.0 million

Solution

Step 1 Since there are 360_ in a circle, the frequency for each class must be converted into a proportional part of the circle. This conversion is done by using the formula Degrees = $\frac{f}{n} \bullet 360°$

where f _ frequency for each class and n _ sum of the frequencies. Hence, the following conversions are obtained. The degrees should sum to 360_.*

Potato chips	$\frac{11.2}{30} \cdot 360° =$	134°
Tortilla chips	$\frac{8.2}{30} \cdot 360° =$	98°
Pretzels	$\frac{4.3}{30} \cdot 360° =$	52°
Popcorn	$\frac{3.8}{30} \cdot 360° =$	46°
Snack nuts	$\frac{2.5}{30} \cdot 360° =$	30°
Total		360°

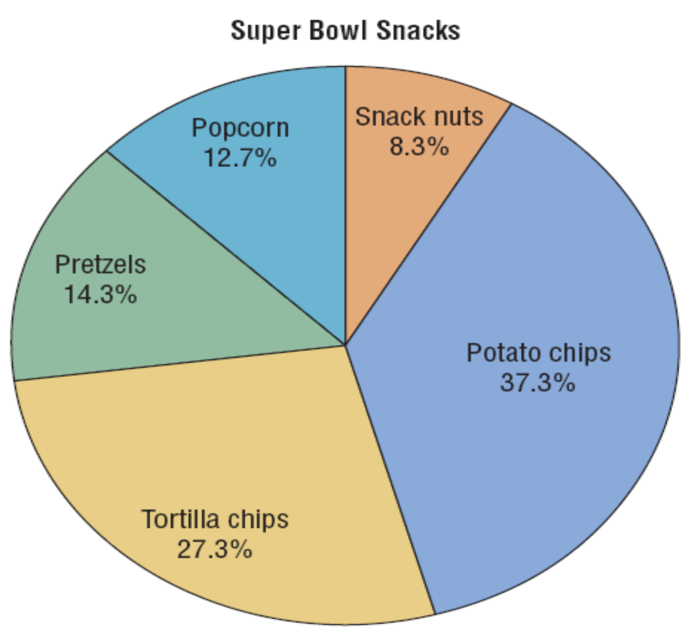
Step 2 Each frequency must also be converted to a percentage. Recall from Example 2–1 that this conversion is done by using the formula

$$\% = \frac{f}{n} \cdot 100$$

Hence, the following percentages are obtained. The percentages should sum to 100%.†

$$\begin{array}{ll} \text{Potato chips} & \frac{11.2}{30} \cdot 100 = 37.3\% \\ \text{Tortilla chips} & \frac{8.2}{30} \cdot 100 = 27.3\% \\ \text{Pretzels} & \frac{4.3}{30} \cdot 100 = 14.3\% \\ \text{Popcorn} & \frac{3.8}{30} \cdot 100 = 12.7\% \\ \text{Snack nuts} & \frac{2.5}{30} \cdot 100 = 8.3\% \\ \hline \text{Total} & 99.9\% \end{array}$$

Step 3 Next, using a protractor and a compass, draw the graph using the appropriate degree measures found in step 1, and label each section with the name and percentages, as shown in Figure 2–14.



Data Description

3-1 Measures of Central Tendency

3-2 Measures of Variation

Measures of Central Tendency: Summarize data, using measures of central tendency, such as the mean, median, mode, and midrange.

Measures of Variation: Describe data, using measures of variation, such as the range, variance, and standard deviation.

Measures of Central Tendency	
A statistic is a characteristic or measure obtained by using the data values from a sample.	**A parameter is a characteristic or measure obtained by using all the data values for a specific population.**

Mean

- The mean is the quotient of the sum of the values and the total number of values.
- The symbol $\overline{X}$ is used for sample mean.

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum X}{n}$$

- For a population, the Greek letter μ (mu) is used for the mean.

$$\mu = \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N} = \frac{\sum X}{N}$$

- The mean is the sum of the values, divided by the total number of values.

$\bar{x}$	μ
The symbol for the sample mean:	The symbol for the population mean:
$\bar{X} = \frac{\sum X}{n} = \frac{X_1 + X_2 + X_3 + \dots\dots\dots + X_n}{n}$	$\mu = \frac{\sum X}{N} = \frac{X_1 + X_2 + X_3 + \dots\dots\dots + X_n}{N}$
Where n: no. of val. In sample.	Where N: no. of val. In population.

Examples

Days off per Year

The data represent the number of days off per year for a sample of individuals selected from nine different countries. Find the mean.

20, 26, 40, 36, 23, 42, 35, 24, 30

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum X}{n}$$

$$\bar{X} = \frac{20 + 26 + 40 + 36 + 23 + 42 + 35 + 24 + 30}{9} = \frac{276}{9} = 30.7$$

The mean number of days off is 30.7 years.

Police Incidents

The number of calls that a local police department responded to for a sample of 9 months is shown. Find the mean.

475, 447, 440, 761, 993, 1052, 783, 671, 621

$$\overline{X} = \frac{\Sigma x}{n} = \frac{475 + 447 + 440 + 761 + 993 + 1052 + 783 + 671 + 621}{9}$$

$$= \frac{6243}{9} \approx 693.7$$

Finding the Mean for Grouped Data

Step 1 Make a table as shown

A Class	B Frequency f	C Midpoint X_m	D $f \cdot X_m$

Step 2 Find the midpoints of each class and place them in column C.

Step 3 multiply the frequency by the midpoint for each class, and place the product in column D.

Step 4 Find the sum of column D.

Step 5 Divide the sum obtained in column D by the sum of frequencies obtained in column B.

The formula for the mean is

$$\overline{X} = \frac{\Sigma f \cdot X_m}{n}$$

Miles Run per Week

Using the frequency distribution for Example below, find the mean. The data represent the number of miles run during one week for a sample of 20 runners.

Solution

The procedure for finding the mean for grouped data is given here.

Step 1 Make a table as shown.

A Class	B Frequency f	C Midpoint X_m	D $f \cdot X_m$
5.5–10.5	1		
10.5–15.5	2		
15.5–20.5	3		
20.5–25.5	5		
25.5–30.5	4		
30.5–35.5	3		
35.5–40.5	2		
	$n = 20$		

Step 2 Find the midpoints of each class and enter them in column C

$$X_m = \frac{5.5 + 10.5}{2} = 8 \qquad \frac{10.5 + 15.5}{2} = 13 \qquad \text{etc.}$$

Step 3 For each class, multiply the frequency by the midpoint, as shown, and place the product in column D.
1 * 8 =8   2 * 13 = 26   etc. The completed table is shown here.

A Class	B Frequency f	C Midpoint X_m	D $f \cdot X_m$
5.5–10.5	1	8	8
10.5–15.5	2	13	26
15.5–20.5	3	18	54
20.5–25.5	5	23	115
25.5–30.5	4	28	112
30.5–35.5	3	33	99
35.5–40.5	2	38	76
	$n = 20$		$\Sigma f \cdot X_m = 490$

Step 4 Find the sum of column D.
Step 5 Divide the sum by n to get the mean.

$$\overline{X} = \frac{\Sigma f \cdot X_m}{n} = \frac{490}{20} = 24.5 \text{ miles}$$

MEDIAN

The **median** is the midpoint of the data array. The symbol for the median is MD.

Finding the median

Step 1 Arrange the data values in ascending order.

Step 2 determine the number of values in the data set.

Step 3 a. If *n* is odd, select the middle data value as the median.

b. If *n* is even, find the mean of the two middle values. That is, add them and divide the sum by 2.

Examples

Police Officers Killed

The number of police officers killed in the line of duty over the last 11 years is shown. Find the median.

177 153 122 141 189 155 162 165 149 157 240

Sort in ascending order

122, 141, 149, 153, 155, 157, 162, 165, 177, 189, 240

Select the middle value.

MD = 157

The median is 157 rooms.

Tornadoes in the U.S.

The number of tornadoes that have occurred in the United States over an 8-year period follows. Find the median.

684, 764, 656, 702, 856, 1133, 1132, 1303

Find the average of the two middle values.

656, 684, 702, 764, 856, 1132, 1133, 1303

$$MD = \frac{764 + 856}{2} = \frac{1620}{2} = 810$$

The median number of tornadoes is 810.

The Mode

- The **mode** is the value that occurs most often in a data set.
- It is sometimes said to be the most typical case.
- There may be no mode, one mode (unimodal), two modes (bimodal), or many modes (multimodal).

Example

NFL Signing Bonuses

Find the mode of the signing bonuses of eight NFL players for a specific year. The bonuses in millions of dollars are

18.0, 14.0, 34.5, 10, 11.3, 10, 12.4, 10

You may find it easier to sort first.

10, 10, 10, 11.3, 12.4, 14.0, 18.0, 34.5

Select the value that occurs the most.

The mode is 10 million dollars.

Licensed Nuclear Reactors

The data show the number of licensed nuclear reactors in the United States for a recent 15-year period. Find the mode.

104 104 104 104 104 107 109 109 109 110
109 111 112 111 109

104 and 109 both occur the most. The data set is said to be bimodal.

The modes are 104 and 109.

Miles Run per Week

Find the modal class for the frequency distribution of miles that 20 runners ran in one week.

Class	Frequency
5.5 – 10.5	1
10.5 – 15.5	2
15.5 – 20.5	3
20.5 – 25.5	5
25.5 – 30.5	4
30.5 – 35.5	3
35.5 – 40.5	2

The modal class is 20.5 – 25.5.

The mode, the midpoint of the modal class, is 23 miles per week.

Area Boat Registrations

The data show the number of boats registered for six counties in southwestern Pennsylvania. Find the mode.

Westmoreland	11,008
Butler	9,002
Washington	6,843
Beaver	6,367
Fayette	4,208
Armstrong	3,782

Since the category with the highest frequency is Westmoreland, the most typical case is Westmoreland. Hence, the mode is Westmoreland.

Chapter 15

STATISTICS

15.1 Overview

In earlier classes, you have studied measures of central tendency such as mean, mode, median of ungrouped and grouped data. In addition to these measures, we often need to calculate a second type of measure called a **measure of dispersion** which measures the **variation** in the observations about the **middle value**– mean or median etc.

This chapter is concerned with some important measures of dispersion such as mean deviation, variance, standard deviation etc., and finally analysis of frequency distributions.

15.1.1 *Measures of dispersion*

(a) **Range**The measure of dispersion which is easiest to understand and easiest to calculate is the **range**. Range is defined as:

Range = Largest observation – Smallest observation

(b) **Mean Deviation**

(i) Mean deviation for ungrouped data:

For n observation $x_1, x_2, ..., x_n$, the **mean deviation about their mean** $\overline{x}$ is given by

$$\text{M.D}\ (\overline{x}) = \frac{|x_i - \overline{x}|}{n} \qquad (1)$$

Mean deviation about their median M is given by

$$\text{M.D (M)} = \frac{|x_i - \text{M}|}{n} \qquad (2)$$

(ii) Mean deviation for discrete frequency distribution

Let the given data consist of discrete observations $x_1, x_2, ... , x_n$ occurring with frequencies $f_1, f_2, ... , f_n$, respectively. In this case

$$\text{M.D}(\bar{x}) = \frac{f_i\,|x_i - \bar{x}|}{f_i} = \frac{f_i\,|x_i - \bar{x}|}{\text{N}} \qquad (3)$$

$$\text{M.D (M)} = \frac{f_i\,|x_i - \text{M}|}{\text{N}} \qquad (4)$$

where N = f_i .

(iii) Mean deviation for continuous frequency distribution (Grouped data).

$$\text{M.D}(\bar{x}) = \frac{f_i\,|x_i - \bar{x}|}{\text{N}} \qquad (5)$$

$$\text{M.D (M)} = \frac{f_i\,|x_i - \text{M}|}{\text{N}} \qquad (6)$$

where x_i are the midpoints of the classes, $\bar{x}$ and M are, respectively, the mean and median of the distribution.

(c) **Variance :** Let $x_1, x_2, ..., x_n$ be n observations with $\bar{x}$ as the mean. The variance, denoted by σ^2, is given by

$$\sigma^2 = \frac{1}{n} \quad (x_i - \bar{x})^2 \qquad (7)$$

(d) **Standard Deviation:** If σ^2 is the variance, then σ, is called the standard deviation, is given by

$$\sigma = \sqrt{\frac{1}{n} \quad (x_i - \bar{x})^2} \qquad (8)$$

(e) **Standard deviation for a discrete frequency distribution** is given by

$$\sigma = \sqrt{\frac{1}{\text{N}} \quad f_i (x_i - \bar{x})^2} \qquad (9)$$

where f_i's are the frequencies of x_i' s and N = $\overset{n}{\underset{i=1}{}} f_i$.

(f) **Standard deviation of a continuous frequency distribution (grouped data)** is given by

$$\sigma = \sqrt{\frac{1}{N} \quad f_i (x_i - \overline{x})^2} \qquad (10)$$

where x_i are the midpoints of the classes and f_i their respective frequencies. Formula (10) is same as

$$\sigma = \frac{1}{N} \sqrt{N \quad f_i x_i^2 - \left(\quad f_i x_i \right)^2} \qquad (11)$$

(g) Another formula for standard deviation :

$$\sigma_x = \frac{h}{N} \sqrt{N \quad f_i y_i^2 - \left(\quad f_i y_i \right)^2} \qquad (12)$$

where h is the width of class intervals and $y_i = \frac{x_i - A}{h}$ and A is the assumed mean.

15.1.2 ***Coefficient of variation*** It is sometimes useful to describe **variability** by expressing the standard deviation as a proportion of mean, usually a percentage. The formula for it as a percentage is

$$\text{Coefficient of variation} = \frac{\text{Standard deviation}}{\text{Mean}} \times 100$$

15.2 Solved Examples

Short Answer Type

Example 1 Find the mean deviation about the mean of the following data:

Size (x):	1	3	5	7	9	11	13	15
Frequency (f):	3	3	4	14	7	4	3	4

Solution Mean $= \overline{x} = \frac{f_i x_i}{f_i} = \frac{3+9+20+98+63+44+39+60}{42} = \frac{336}{42} = 8$

$$\text{M.D.} (\overline{x}) = \frac{f_i \mid x_i - \overline{x} \mid}{f_i} = \frac{3(7)+3(5)+4(3)+14(1)+7(1)+4(3)+3(5)+4(7)}{42}$$

$$= \frac{21+15+12+14+7+12+15+28}{42} = \frac{62}{21} = 2.95$$

Example 2 Find the variance and standard deviation for the following data:
57, 64, 43, 67, 49, 59, 44, 47, 61, 59

Solution Mean ($\overline{x}$) $= \frac{57+64+43+67+49+59+61+59+44+47}{10} = \frac{550}{10} = 55$

$$\text{Variance } (\sigma^2) = \frac{(x_i - \overline{x})^2}{n}$$

$$= \frac{2^2+9^2+12^2+12^2+6^2+4^2+6^2+4^2+11^2+8^2}{10}$$

$$= \frac{662}{10} = 66.2$$

Standard deviation (σ) $= \sqrt{\sigma^2} = \sqrt{66.2} = 8.13$

Example 3 Show that the two formulae for the standard deviation of ungrouped data.

$$\sigma = \sqrt{\frac{(x_i - \overline{x})^2}{n}} \quad \text{and} \quad \sigma' = \sqrt{\frac{x_i^2}{n} - \overline{x}^2}$$

are equivalent.

Solution We have $(x_i - \overline{x})^2 = (x_i^2 - 2\overline{x}\, x_i + \overline{x}^2)$

$$= x_i^2 + \quad -2\overline{x}\, x_i + \quad \overline{x}^2$$

$$= x_i^2 - 2\overline{x} \quad x_i + (\overline{x})^2 \quad 1$$

$$= x_i^2 - 2\overline{x}\,(n\overline{x}) + n\,\overline{x}^2$$

$$= x_i^2 - n\overline{x}^2$$

Dividing both sides by n and taking their square root, we get $\sigma = \sigma'$.

Example 4 Calculate **variance** of the following data :

Class interval	Frequency
4 - 8	3
8 - 12	6
12 - 16	4
16 - 20	7

$$\text{Mean } (\overline{x}) = \frac{f_i x_i}{f_i} = \frac{3 \times 6 + 6 \times 10 + 4 \times 14 + 7 \times 18}{20} = 13$$

Solution Variance (σ^2) = $$\frac{f_i (x_i - \overline{x})^2}{f_i} = \frac{3(-7)^2 + 6(-3)^2 + 4(1)^2 + 7(5)^2}{20}$$

$$= \frac{147 + 54 + 4 + 175}{20} = 19$$

Long Answer Type

Example 5 Calculate mean, variation and standard deviation of the following frequency distribution:

Classes	Frequency
1 - 10	11
10 - 20	29
20 - 30	18
30 - 40	4
40 - 50	5
50 - 60	3

0.1 Bernoulli Distribution

This distribution describes a natural phenomenon or a mechanical process in which you expect a particular event to appear or not.

If the outcome of the random experiment is either a success with a fixed probability p or a failure with a probability $q = 1 - p$ and that the random variable X takes either the value 1 in the case of success or a value of zero in the case of failure, then the distribution of X is the Bernoulli distribution.

That is, $X = \begin{cases} 1 & \text{when the event appears} \\ 0 & \text{when the event does not appear} \end{cases}$

and

$$f(x) = \begin{cases} p & x = 1 \\ 1 - p = q & x = 0 \end{cases}$$

This function can be written in another form (Bernoulli distribution).

$$f(x) = P(X = x) = \begin{cases} p^x q^{1-x} & x = 0, 1, \quad 0 < p < 1, \quad q = 1 - p \\ 0 & o.w. \end{cases}$$

Where $X \sim Ber(p)$; p is the parameter of the distribution.

Exercise 1 *:*

1. *Prove that the Bernoulli distribution is a probability mass function.*
2. *Find the*
   - *Average.*
   - *Variance.*
   - *Moment generating function.*
   - *Cumulative distribution function.*

Solution:

1. To prove that the Bernoulli distribution is a probability mass function.

   Since $0 < p, q < 1$ then $0 < p^x q^{1-x} < 1$, which yields $0 < f(x) < 1$ when $x = 0, 1$ and since $f(x) = 0$ for otherwise ($x \neq 0, 1$). Then $0 \leq f(x) < 1$.

   Now we want to prove $\sum_X f(x) = 1$.

   $$\sum_X f(x) = \sum_{X=0}^{1} p^X q^{1-X} = p^0 q^1 + p^1 q^0 = q + p = 1 - p + p = 1$$

   Hence, f is a p.m.f of a r.v. X.

2. To find the Average, Variance, Moment generating function for this distribution and Cumulative distribution function.

   - $\mu = E(X) = \sum_X X f(x) = \sum_{X=0}^{1} X \cdot p^X q^{1-X}$

     $= 0 \cdot p^0 q^1 + 1 \cdot p^1 q^0 = p$

     $\boxed{\mu = E(X) = p}$

   - $\sigma^2 X = E(X^2) - [E(X)]^2$.

     $$E(X^2) = \sum_{X=0}^{1} X^2 \cdot p^X q^{1-X} = 0 \cdot p^0 q^1 + 1 \cdot p^1 q^0 = p$$

Note that the moments for any r is p since

$$\mu_r = E(X^r) = \sum_{X=0}^{1} X^r \cdot p^X q^{1-X} = 0 \cdot p^0 q^1 + 1 \cdot p^1 q^0 = p$$

$$\sigma^2 X = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p) = pq.$$

$$\boxed{\sigma^2 X = pq.}$$

- $$\begin{aligned} M_X(t) = E(e^{tX}) &= \sum_X e^{tX} \cdot f(x) \\ &= \sum_{X=0}^{1} e^{tX} \cdot p^X q^{1-X} \\ &= e^0 \cdot p^0 q^1 + e^t \cdot p^1 q^0 = pe^t + q. \end{aligned}$$

$$\boxed{M_X(t) = pe^t + q}$$

- C.D.F of X is

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ q & 0 \leq x < 1 \\ \\ 1 & x \geq 1 \end{cases}$$

Example 1 *One dice was thrown. Let the random variable X be the number 6 shown by the dice face. Find:*

1. *p.m.f of X.*
2. *c.d.f of X.*
3. *$M_X(t)$.*
4. *$E(X)$.*
5. *$V(X)$.*
6. *$P(-3 \leq X < 0)$.*
7. *$P(-2 \leq X < 1)$.*
8. *$P(X \geq 3)$.*

9. $P(0 \leq X < 2)$.

Solution:

If the number 6 appears, then $X = 1$ (success).

If the number $1, 2, 3, 4$, or 5 appear, then $X = 0$ (failure).

Hence, $X \sim Ber(p)$ with $p = \frac{1}{6}$ and $q = 1 - p = \frac{5}{6}$

$$f(x) = P(X = x) = \begin{cases} (\frac{1}{6})^x(\frac{5}{6})^{1-x} & x = 0, 1 \\ 0 & o.w. \end{cases}$$

- C.D.F of X is

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ \frac{5}{6} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

- $M_X(t) = pe^t + q = \frac{1}{6}e^t + \frac{5}{6} = \frac{5+e^t}{6}$.

- $E(X) = p = \frac{1}{6}$.

- $V(X) = \sigma^2 X = E(X^2) - [E(X)]^2 = p \cdot q = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$.

- $P(-3 \leq X < 0) = 0$.

- $P(0 \leq X < 2) = p(X = 0) + p(X = 1) = \frac{5}{6} + \frac{1}{6} = 1$.

- $P(-2 \leq X < 1) = p(X = -2) + p(X = -1) + p(X = 0)$
  $= 0 + 0 + \frac{5}{6} = \frac{5}{6}$.

- $P(X \geq 3) = 0$.

The Binomial Distribution

A. It would be very tedious if, every time we had a slightly different problem, we had to determine the probability distributions from scratch. Luckily, there are enough similarities between certain types, or families, of experiments, to make it possible to develop formulas representing their general characteristics.

For example, many experiments share the common element that their outcomes can be classified into one of two events, e.g. a coin can come up heads or tails; a child can be male or female; a person can die or not die; a person can be employed or unemployed. These outcomes are often labeled as "success" or "failure." Note that there is no connotation of "goodness" here - for example, when looking at births, the statistician might label the birth of a boy as a "success" and the birth of a girl as a "failure," but the parents wouldn't necessarily see things that way. The usual notation is

p = probability of success,
q = probability of failure = $1 - p$.

Note that $p + q = 1$. In statistical terms, **A Bernoulli trial is each repetition of an experiment involving only 2 outcomes**.

We are often interested in the result of *independent, repeated bernoulli trials*, i.e. the number of successes in repeated trials.

1. *independent* - the result of one trial does not affect the result of another trial.

2. *repeated* - conditions are the same for each trial, i.e. p and q remain constant across trials. Hayes refers to this as a stationary process. If p and q can change from trial to trial, the process is nonstationary. The term identically distributed is also often used.

B. A binomial distribution gives us the probabilities associated with independent, repeated Bernoulli trials. **In a binomial distribution the probabilities of interest are those of receiving a certain number of successes, r, in n independent trials each having only two possible outcomes and the same probability, p, of success.** So, for example, using a binomial distribution, we can determine the probability of getting 4 heads in 10 coin tosses.

How does the binomial distribution do this? Basically, a two part process is involved. First, we have to determine the probability of one possible way the event can occur, and then determine the number of different ways the event can occur. That is,

P(Event) = (Number of ways event can occur) * P(One occurrence).

Suppose, for example, we want to find the probability of getting 4 heads in 10 tosses. In this case, we'll call getting a heads a "success." Also, in this case, $n = 10$, the number of successes is $r = 4$, and the number of failures (tails) is $n – r = 10 – 4 = 6$. One way this can occur is if the first 4 tosses are heads and the last 6 are tails, i.e.

S S S S F F F F F F

The likelihood of this occurring is

P(S) * P(S) * P(S) * P(S) * P(F) * P(F) * P(F) * P(F) * P(F) * P(F)

More generally, if p = probability of success and q = 1 – p = probability of failure, the probability of a specific sequence of outcomes where there are r successes and n-r failures is

$$p^r q^{n-r}$$

So, in this particular case, p = q = .5, r = 4, n-r = 6, so the probability of 4 straight heads followed by 6 straight tails is $.5^4 .5^6 = 0.0009765625$ (or 1 out of 1024).

Of course, this is just one of many ways that you can get 4 heads; further, because the repeated trials are all independent and identically distributed, each way of getting 4 heads is equally likely, e.g. the sequence S S S S F F F F F F is just as likely as the sequence S F S F F S F F S F. So, we also need to know how many different combinations produce 4 heads.

Well, we could just write them all out…but life will be much simpler if we take advantage of two counting rules:

1. The number of different ways that N distinct things may be arranged in order is

N! = (1)(2)(3)...(N-1)(N), (where 0! = 1).

An arrangement in order is called a permutation, so that the total number of permutations of N objects is N!. The symbol N! is called N factorial.

EXAMPLE. Rank candidates A, B, and C in order. The possible permutations are: ABC ACB BAC BCA CBA CAB. Hence, there are 6 possible orderings. Note that 3! = (1)(2)(3) = 6.

NOTE: Appendix E, Table 6, p. 19 contains a Table of the factorials for the integers 1 through 50. For example, $12! = 4.79002 * 10^8$. (Or see Hayes Table 8, p. 947). Your calculator may have a factorial function labeled something like x!

2. The total number of ways of selecting r distinct combinations of N objects, irrespective of order, is

$$\frac{N!}{r!(N-r)!} = \binom{N}{r} = \binom{N}{N-r}$$

We refer to this as "N choose r." Sometimes the number of combinations is known as a *binomial coefficient*, and sometimes the notation $_NC_r$ is used. So, in the present example,

$$\binom{N}{r}=\binom{10}{4}=\frac{N!}{r!(N-r)!}=\frac{10!}{4!(10-4)!}=\frac{10*9*8*7}{4*3*2*1}=\frac{5040}{24}=210$$

Note that, for 10!, I stopped once I got to 7; and I didn't write out 6! in the denominator. This is because both numerator and denominator have 6! in them, so they cancel out. So, there are 210 ways you can toss a coin 10 times and get 4 heads.

EXAMPLE. Candidates A, B, C and D are running for office. Vote for two.

The possible choices are: AB AC AD BC BD CD, i.e. there are 6 possible combinations. Confirming this with the above formula, we get

$$\frac{N!}{r!(N-r)!}=\binom{4}{2}=\frac{4!}{2!(4-2)!}=\frac{(4)(3)(2)(1)}{(2)(1)(2)(1)}=\frac{12}{2}=6$$

EXAMPLE. There are 100 applicants for 3 job openings. The number of possible combinations is

$$\frac{N!}{r!(N-r)!}=\binom{100}{3}=\frac{100!}{3!97!}=\frac{100*99*98}{3*2}=\frac{970{,}200}{6}=161{,}700$$

Again, note that, if you didn't take advantage of 97! appearing on both top and bottom, you'd have a much lengthier calculation.

> See Appendix E, Table 7, page 20 for $_NC_r$ values for various values of N and r. (Or see Hayes, Appendix E, Table IX, p. 948). Your calculator may have a function labeled nCr or something similar.

C. So putting everything together now: we know that any specific sequence that produces 4 heads in 10 tosses has a probability of 0.0009765625. Further, we now know that there are 210 such sequences. Ergo, the probability of 4 heads in 10 tosses is 210 * 0.0009765625 = 0.205078125.

We can now write out the complete formula for the binomial distribution:

In sampling from a stationary Bernoulli process, with the probability of success equal to p, the probability of observing exactly r successes in N independent trials is

$$\binom{N}{r}p^r q^{N-r}=\frac{N!}{r!(N-r)!}p^r q^{N-r}$$

Once again, N choose r tells you the number of sequences that will produce r successes in N tries, while $p^r q^{N-r}$ tells you what the probability of each individual sequence is.

To put it another way, the random variable X in a binomial distribution can be defined as follows:

Let Xi = 1 if the ith bernoulli trial is successful, 0 otherwise. Then,

X = ΣX_i, where the X_i's are independent and identically distributed (iid).

That is, X = the # of successes. Hence**, Any random variable X with probability function given by**

$$p(X=r;\, N, p)=\binom{N}{r} p^r q^{N-r} = \frac{N!}{r!(N-r)!} p^r q^{N-r}, \quad X=0,1,2,\ldots,N$$

is said to have a binomial distribution with parameters N and p.

EXAMPLE. In each of 4 races, the Democrats have a 60% chance of winning. Assuming that the races are independent of each other, what is the probability that:

a. The Democrats will win 0 races, 1 race, 2 races, 3 races, or all 4 races?
b. The Democrats will win at least 1 race
c. The Democrats will win a majority of the races

SOLUTION. Let X equal the number of races the Democrats win.

a. Using the formula for the binomial distribution,

$$\binom{4}{0} p^0 q^{4-0} = \frac{4!}{0!(4-0)!} .60^0 .40^4 = .40^4 = .0256,$$

$$\binom{4}{1} p^1 q^{4-1} = \frac{4!}{1!(4-1)!} .60^1 .40^3 = 4 * .60 * .40^3 = .1536,$$

$$\binom{4}{2} p^2 q^{4-2} = \frac{4!}{2!(4-2)!} .60^2 .40^2 = 6 * .60^2 .40^2 = .3456,$$

$$\binom{4}{3} p^3 q^{4-3} = \frac{4!}{3!(4-3)!} .60^3 .40^1 = 4 * .60^3 .40^1 = .3456,$$

$$\binom{4}{4} p^4 q^{4-4} = \frac{4!}{4!(4-4)!} .60^4 .40^0 = .60^4 = .1296$$

b. P(at least 1) = P(X ≥ 1) = 1 - P(none) = 1 - P(0) = .9744. Or, P(1) + P(2) + P(3) + P(4) = .9744.

c. P(Democrats will win a majority) = P(X ≥ 3) = P(3) + P(4) = .3456 + .1296 = .4752.

EXAMPLE. In a family of 11 children, what is the probability that there will be more boys than girls? Solve this problem WITHOUT using the complements rule.

SOLUTION. You could go through the same tedious process described above, which is what most students did when I first asked this question on an exam. You would compute P(6), P(7), P(8), P(9), P(10), and P(11).

Or, you can look at Appendix E, Table II (or Hays pp. 927-931). Here, both Hayes and I list binomial probabilities for values of N and r from 1 through 20, and for values of p that range from .05 through .50.

Thus, on page E-5, we see that for N = 11 and p = .50,

P(6) + P(7) + P(8) + P(9) + P(10) + P(11) = .2256 + .1611 + .0806 + .0269 + .0054 + .0005 = .50.

NOTE: Understanding the tables in Appendix E can make things a lot simpler for you!

EXAMPLE. [WE MAY SKIP THIS EXAMPLE IF WE RUN SHORT OF TIME, BUT YOU SHOULD STILL GO OVER IT AND MAKE SURE YOU UNDERSTAND IT]

Use Appendix E, Table II, to once again solve this problem: In each of 4 races, the Democrats have a 60% chance of winning. Assuming that the races are independent of each other, what is the probability that:

a. The Democrats will win 0 races, 1 race, 2 races, 3 races, or all 4 races?
b. The Democrats will win at least 1 race
c. The Democrats will win a majority of the races

SOLUTION. It may seem like you can't do this, since the table doesn't list p = .60. However, all you have to do is redefine success and failure. Let success = P(opponents win a race) = .40. The question can then be recast as finding the probability that

a. The opponents will win 4 races, 3 races, 2 races, 1 race, or none of the races?
b. The opponents will win 0, 1, 2, or 3 races; or, the opponents will not win all the races
c. The opponents will not win a majority of the races

We therefore look at page E-4 (or Hayes, p. 927), N = 4 and p = .40, and find that

a. P(4) = .0256, P(3) = .1536, P(2) = .3456, P(1) = .3456, and P(0) = .1296.
b. P(0) + P(1) + P(2) + P(3) = 1 - P(4) = .9744
c. P(1) + P(0) = .3456 + .1296 = .4752

In general, for p > .50: To use Table II, substitute 1 - p for p, and substitute N - r for r. Thus, for p = .60 and N = 4, the probability of 1 success can be found by looking up p = .40 and r = 3.

D. *Mean of the binomial distribution.* Recall that, for any discrete random variable, E(X) = Σxp(x). Therefore, $E(X_i)$ = Σxp(x) = 0 * (1 - p) + 1 * p = p, that is, the mean of any