M 404

Institute of Distance and Open Learning
Gauhati University

M.A./M.Sc. in Mathematics
Semester 4

Paper IV
Mathematical Statistics

Contents :

Math Paper IV

Course Co-ordination:

Prof. Amit Choudhury — Director, IDOL, GU

Dr. Chandra Lekha Mahanta — Department of Mathematics, Gauhati University

Format Editing :

Dipankar Saikia — Editor, Study Materials, IDOL, GU

Cover Design :

Bhaskar Jyoti Goswami : GU IDOL

CHAPTER-1

Probability

Introduction:

Probability theory is the branch of mathematics concerned with probability, the analysis of random phenomena. Probability theory began in seventeenth century in French when the two great French mathematicians, Blaise Pascal and Pierre de Format, corresponded over two problems from games of chance. In common parlance, the term Probability refers to the chance of happening or not happening of an event. Today, the theory of probability has been extensively developed and there is hardly any discipline – physical or social – where it is not being extensively used.

Random experiments:

It we perform certain experiments under identical conditions we expect of arrive at results that are essentially same provided we can control the value of the variables that may affect the out -come of the experiment. But in many cases we are not in a position to control the value of same variable. As a result though under identical conditions experiments are performed, the result will vary from one experiment to the next. These experiments are called random experiment.

Examples of random experiments.

1. Let the experiment be tossing of a coin. The result will be either head (H) or tall (T). But we cannot exactly predicts what the result will be . The results will depend on chance. There are various factor that will influence the result and all these factors cannot be controlled . This is an example of random experiment.

If coin is tossed twice, the result will be {(H,H), (H,T),(T,H),(T,T)}. This is also example of random experiments.

2. Let us consider the random experiment of throwing a six faced unbiased cubical die. As the die is perfect, we are sure that one of the faces will come up with one of the numbers 1,2,3,4,5,6.

Sample sapce and Event : A **sample space** is a collection of possible outcomes of a random experiment. A sample space may consist of a single outcome or a group of outcomes taken together. Each possible outcome or element in a sample space is called a **sample point or an elementary event**. An event will be denoted by capital letter of English alphabet.

A series of events $A_1, A_2, \ldots, A_k$ will be called **exhaustive**, if at least one of them is sure to happen in any trial of random experiment. For example when a coin is tossed, either head (H) or tail must occur. Hence the event is exhaustive.

Two events A and B are said to be **mutually exclusive**, if in occurrence of one precludes the occurrences of the other. In other words the two events cannot occur simultaneously. For example in casting of a die if 6 occurs then other members viz 1,2,3, 4,5 cannot occur. Hence it is an example of mutually exclusive events. A series of events $A_1, A_2, \ldots, A_k$ are said to be **equally likely**, if one of them cannot be expected to occur in preference to others in a single trial of random experiment . For example in tossing of a coin each face is equally likely to occur.

Classical definition of probability:

Let 'n' be exhaustive, mutually exclusive and equally like cases and m of them are favourable to an event A then the probability of A is defined by $\frac{m}{n}$ and is denoted by P(A) or p. So that

$$P(A)=\frac{m}{n}$$

since $0 \leq m \leq n$, P(A) lies between 0 and 1. When there is no elementary event favourable to A, then P(A)=0 since m=0.

Example 1. Let a six-faced die be cast. Find in probability that (i) the number shown on the die is odd (ii) the number shown in the die is divisible by 2 and 3 (iii) the number shown on the die is divisible by 2 and 5

When a die is cast there are six cases 1,2,3,4,5 or 6. These are exhaustive, mutually exclusive and equally like.

(i) If A be the event, then the favourable cases for the event A are 1,3,5.

$$P(A)=\frac{3}{6}=\frac{1}{2}.$$

(ii) If B be the event, then there is only one favourable case viz 6.

$$P(B)=\frac{1}{6}.$$

(iii) Out of 1,2,3,4,5 or 6 no one is divisible by 2 and 5. Hence if C denotes this event

$$P(C)=0.$$

Example2. A card is drawn at random from a well shuffled pack of playing cards, Find in probability then it is (i) a king (ii) a queen of spade (iii) a heart.

In a pack of cards there are 52 cards and any card may be drawn and hence total number of cases is 52.

(i) Let A denote the event. There are four kings and so number of favourable cases for drawing a king is 4.

$$P(A)=\frac{4}{52}=\frac{1}{13}.$$

(ii) Let B denote the event. There is only one queen of spade.

$$P(B)=\frac{1}{52}$$

(iii) Let the event be C. There are 13 hearts

$$P(C)=\frac{13}{52}=\frac{1}{4}.$$

Limitations of Classical definition :

The classical definition requires that 'n' is finite. But there are instances where it may be infinite. Secondly we assume elementary events to be equally likely. This is also need not be true always. For example in tossing a coin, there are two events head and tail. Unless the coin is unbiased they may not be equally likely. To get a perfectly unbiased coin is a very difficult condition. Hence probability of obtaining a head with any coin can be obtained from classical definition.

Statistical definition:

To overcome the difficulties encountered in classical definition, large number of trials of random experiment is considered . Let m be the number of occurrences of an event A associated with the n independent trials of random experiment. The probability of the event A is defined by

$$P(A) = \lim_{n\to\infty} \frac{m}{n}$$

Example:

In tossing of a coin 1000 times heads comes up 537 times. The probability of head coming up is

$$\frac{537}{1000}=0.537.$$

But this definition of probability also has drawbacks, because the large number involved in definition is vague.

As both the definitions suffer from certain defects, another approach has been put forward to over come all defects. The approach is known as axiomatic approach of probability. Before considering this approach, let us introduce another concepts know as sample space.

Sample Space:

A set S that consists of all possible outcomes of a random experiment is called a sample space. Each outcome is a sample point.

Example: Let us consider the result of casting a six-faced die. The outcome will be the appearance of one and only one of the numbers 1,2,3,4,5,6 on the upper most face of the die. In other words the six possibilities are such that no two or more of them can occur simultaneously and at least one of them must occur. The sample space S is {1,2,3,4,5,6}. If A denote the event of occurrence of even number then A={2,4,6} which is a subset of S.

Let us generalize the concept.

Let all possible outcomes of some particular experiment be $e_1, e_2, ..., e_n$ which are such that no two or more of them can occur simultaneously and exactly one of the outcomes must occur when in experiment is performed. Any set associated with an experiment which satisfied the above mentioned two properties is called a sample space. The elements or points of a sample space elementary events of the experiments.

If a sample space has finite number of points then it is a finite sample space. If it has many points as there are natural numbers then it is a countably infinite sample space.

Example 1: If we toss a coin twice, the four possible results are {HH,HT,TH,TT} which is the sample space.

Example 2: From an urn containing 4 balls of different colours red(R), Blue(B), white (W) and green(G); draw two balls simultaneously. The sample space for this experiment is {RB, RW, RG, BW, BG, WG}.

Note RB and BR represent the same outcome because we draw balls simultaneously. Instead of balls drawing simultaneously, let us now draw balls in succession with replacement. The sample space will be {RR, RB, RW, RG, BR, BB, BW, BG, WR, WB, WW, WG, GR, GB, GW, GG}.

Note: In this case RB and BR are not same.

Do the following:

1. Describe explicitly sample space of the following experiment.

   (i) Tossing of a coin three times.

   (ii) There successive draws(a) with replacement (b) without replacement from a container containing 6 coloured balls of which 1 is blue, 2 are red and 3 are white.

   (iii) Casting of 2 dice simultaneously.

   (iv) A card is drawn at random from an ordinary pack of 52 playing cards, if the consideration of suit is not considered and if the consideration of colour is not considered.

Axiomatic definition of Probability : Let S be a sample space. If S is discrete then all subsets correspond to events and conversely. If S is an discrete only special subsets correspond to events. To each event A in a class C of events, a real numbers P(A) is associated. Then P(A) is called the probability of the event A if the following axioms are satisfied.

Axiom 1. For every event A in the Class C, $P(A) \geq 0$.

Axiom 2. For certain event S in the Class C, $P(S)=1$.

Axiom 3. For any number of mutually exclusive events $A_1, A_2, \ldots$ in the class C,

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

In particular, for two events A_1 and A_2, $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

Let S be the sample space $\{e_1, e_2, \ldots, e_n\}$

and $P(\{e_i\}) = p_i$, $i = 1, 2, \ldots, n$ such that $0 \le p_i \le 1$ and $\sum_{i=1}^{n} p_i = 1$. Now $P(\{e_i\}) = \frac{1}{n}$.

If 'E' be any event consisting of r $(1 \le r \le n)$ elementary events then

$$P(E) = \frac{1}{n} + \frac{1}{n} + \ldots\ldots\ldots + (r \text{ times}) = \frac{r}{n}$$

$$\text{i.e} \quad P(E) = \frac{n(E)}{n(S)}$$

where $n(E)$=number of elements of E and $n(S)$ =number of elements of S.

Calculation of probabilities:

Example 1: Two dice are thrown. Find the probability of the event where E is

(i) Sum of the number shown by the two dice are 7.

(ii) the number on the two dice are equal.

(iii) number on the 2nd die is greater than the 1st.

The sample space S consists of 36 elementary events

(i) $E_1\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$. So, $P(E_1) = \frac{6}{36} = \frac{1}{6}$

(ii) $E_2 = \{(1,1),(2,2),(3,3;(4,4),(5,5),(6,6)\}$ So, $P(E_2) = \frac{1}{6}$

(iii) $E_3 = \{(1,2), (1,3), (1,4), (1,5), (1,6), (2,3), (2,4), (2,5), (2,6), (3,4), (3,5), (3,6), (4,5), (4,6), (5,6)\}$

$$\text{So,} \quad P(E_3) = \frac{15}{36} = \frac{5}{12}$$

Example 2. A coin is tossed three times in succession. Find the probability of the event that have

(i) two or more heads

(ii) exactly one head and two tails

(iii) number of heads equals in number of tails. The sample space is

{(HHH), (HHT), (HTH), (THH), (THT), (TTH), (TTT)}.

E_1={(HHH), (HHT), (HTH), (THH)}.

$$\text{So, } P(E_1) = \frac{4}{8} = \frac{1}{2}$$

E_2={HTT), (THT), (TTH)}. So, $P(E_2)=\frac{3}{8}$

E_3={(HHH), (TTT)}. So, $P(E_3)=\frac{2}{8}=\frac{1}{4}$

Example 3: From a pack of usual playing cards, a card is drawn at random and is noted. Calculate the probabilities of the following events:

(i) the drawn card is either a spade or a clubs.

(ii) the drawn card is a picture card

(iii) the drawn card is of denomination less then 10 and greater than 5.

The sample space consists of 52 elementary events corespanding to each of the 52 cards than can be drawn from the pack.

(i) Let E_1 be in event than the drawn card is a spade or a club. Then E_1 consists of (13+13)=26 elementary events

$$P(E_1)=\frac{26}{52}=\frac{1}{2}$$

(ii) E_2 be the event than drawn card is a picture card. Here aces are taken as picutre card. Then E_2 contains 4X4=16 elementary events

$$(P(E_2)=\frac{16}{52}=\frac{4}{13}.$$

(iii) E_3 be the event such than denomination lies between 5 and 10 . E_3 consists of 4X4=16 elementary events

$$P(E_3)=\frac{16}{52}=\frac{4}{13}.$$

Some theorems on probability:

Let $S=\{e_1,e_2,\ldots\ldots, e_n\}$ be a sample space of an experiment.

Theorem 1. $P(S)=1$.

Proof: By definition. $P(S)=P(\{e_1\})+P(\{e_2\})+\text{-----}+P(\{e_n\}) = p_1+p_2+\text{-----}+p_n =1$.

This theorem is very trivial.

Theorem 2. If A and B are two events such then $A \subseteq B$ than $P(A) \leq P(B)$.

Proof: $A \subseteq B \Rightarrow B$ contains all the elementary events of A.

The probability of B is equal to the probability of A plus the sum of the probabilities attached to those elementary events which belong to B but not to A is B-A. Since any probability is a non-negative number,

$$P(A) \leq P(B).$$

Theorem 3: For any event A, $0 \leq P(A) \leq 1$.

By axiom 1, $P(A) \geq 0$. Again for any event A, $A \subseteq S$.

By theorem 2, $P(A) \leq P(S)=1$ by theorem 1.

$$\text{Hence } 0 \leq P(A) \leq 1.$$

Note: $P(A)=0 \Rightarrow$ A is an impossible event and $P(S)=1 \Rightarrow$ S is sure event.

But converse is not true. For example let us consider the sample space $S = \{HH, HT, TH, TT\}$

Let us assign the following probabilities

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

$$P(HT)=0=P(TT)$$

If $A=\{HT,TT\}$ the $P(A)=0$, but A is not an impossible event

Again if $B=\{TH,HH\}$ then $P(B)=1$, though B is not a sure event.

Theorem 4: Probability of impossible event is zero.

Let P denote impossible event. We know $S=S \cup P$

$$\Rightarrow P(S)=P(S \cup P)=P(S)+P(P) \text{ by axiom 3}$$

$$\Rightarrow P(P)=0.$$

Theorem 5: $P(A_2-A_1)=P(A_2)-P(A_1)$

We know that $A_2=A_1 \cup (A_2-A_1)=$ where A_1 and A_2-A_1 are mutually exclusive.

$$P(A_2)=P[A_1 \cup (A_2-A_1)]$$

$$=P(A_1)+P(A_2-A_1)$$

$$\Rightarrow P(A_2-A_1)=P(A_2)-P(A_1)$$

Theorem 6: If A' is the complement of A then $P(A')=1-P(A)$

We know that $A \cup A'=S$.

$$\text{Hence } P(A \cup A')=P(S)$$

$$\Rightarrow P(A)+P(A')=1$$

$$\Rightarrow P(A')=1-P(A).$$

Composite event: When an event comprises of more then one sample, then the event is called a composite event. The event $A \cup B$ comprises of the sample points in the whole region bounded by A and B. $A \cup B$ is also denoted by A+B. Similarly, event $A \cap B$ is denoted by A.B.

Theorem 7 (Addition Theorem): For any two events A and B,

$$P(A \cup B)=P(A)+P(B)-P(AB).$$

Proof: Let the number of sample points in the sample space S be n. Let the number comprising the event A be m and that the event B be r. The number common to A and B be k. Then the event $A \cup B$ comprises of sample points that belong to A or B or to both. Hence $A \cup B$ will have m + r -k sample points. $P(A \cup B)$= sum of the probabilities of the m+ r-k sample points belonging to $A \cup B$.

P(A)=sum of the probabilities of m sample points belonging to A.

P(B)= Sum of the probabilities of r sample points belonging to B

P(AB)= Sum of the probabilities of k sample points belonging to AB

$$P(A \cup B)=P(A)+P(B)-P(AB).$$

Corollary : If A and B are mutually exclusive then AB=P so that P(AB)=0 Then.

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

Deduction: if A' denote the complement of A then AA'=P and $A \cup A'$=S Thus

$$P(A \cup A')=P(S)$$

$$\Rightarrow P(A)+P(A')=1$$

$$\Rightarrow P(A')=1-P(A) \text{ which is theorem 6.}$$

Generalisation: If A_1, A_2, A_3 be any there events then

$$P(A_1 \cup A_2 \cup A_3)=P(A_1)+P(A_2)+P(A_3)-P(A_1.A_2)-P(A_1A_3)-P(A_2A_3)+P(A_1A_2A_3)$$

It can be generalized to n events

Example 1. From an ordinary pack of cards a card is drawn at random. What is the probability that it is

(i) 3 of clubs or 6 of diamonds (ii) any suit except heart (iii) a jack or a queen

(iv) either a spade, or a heart or not a picture card.

(i) Let A denote 3 of clubs, and B denote 6 of diamonds. Now $A \cap B=\phi$.

$$P(A \cup B)=P(A)+P(B)=\frac{1}{52}+\frac{1}{52}=\frac{1}{26}$$

(ii) If H denote heart then H' will denote any suit except heart

$$P(H')=1-P(H)=1-\frac{1}{4}=\frac{3}{4}.$$

(iii) Let A denote that the card is Jack and B that it is queen. Now $A\cap B=\phi$.

$$P(A\cup B)=P(A)+P(B)=\frac{1}{13}+\frac{1}{13}=\frac{2}{13}.$$

(iii) Let S denote spade, H denote heart and A denote picture card. S and H are mutually exclusive $(S\cap H=\phi)$ but S and A, H and A are not mutually exclusive.

$$P(S\cap A)=\frac{3}{52} \quad \text{and} \quad P(H\cap A)=\frac{3}{52}$$

Also, $P(S)=\frac{13}{52}$, $P(H)=\frac{13}{52}$ and $P(A)=\frac{12}{52}$

$$P(S\cup H\cup A)=P(S)+P(H)+P(A)-P(S\cap H)-P(S\cap A)-P(H\cap A)+P(S\cap H\cap A)$$

$$=\frac{13}{52}+\frac{13}{52}+\frac{12}{52}-\frac{3}{52}-\frac{3}{52}=\frac{32}{52}=\frac{8}{13}$$

Example 2. Through certain election 3 persons are to be chosen out of five candidates A, B, C, D, E, with equal chances of being elected. Find the probability that

(i) A and B are elected or B and C are elected

(ii) E is elected or C and D are elected.

Let E_1 be the event that A and B are elected and E_2 be the event that B and C are elected.

$$S=\{ABC, ABD, ABE, ACD, ACE, ADE, ADE, BCD, BCE, BDE, CDE\}$$

(i) $E_1=\{ABC, ABD, ABE\}$, $E_2=\{ABC, BCD, BCE\}$.

$$E_1\cap E_2=\{ABC\}.$$

$$P(E_1)=\frac{3}{10}, \quad P(E_2)=\frac{3}{10}, \quad P(E\cap E_2)=\frac{1}{10}$$

$$P(E_1\cup E_2)=\frac{3}{10}+\frac{3}{10}-\frac{1}{10}=\frac{5}{10}=\frac{1}{2}$$

(ii) $E_3=\{ABE, ACE, ADE, BCE, BDE, CDE\}$, $E_4=\{ACD, BCD, CDE\}$

$$E_3\cap E_4=\{CDE\}$$

$$P(E_3)=\frac{6}{10}, \quad P(E_4)=\frac{3}{10}, \quad P(E_3\cap E_4)=\frac{1}{10}, \quad P(E_3\cup E_4)=\frac{6}{10}+\frac{3}{10}-\frac{1}{10}=\frac{8}{10}=\frac{4}{5}$$

Conditional probability:

Suppose birth day of one of your friend is the September. If you do not know anythiing more about his birth day then the probability of any day in september being his birthday is $\frac{1}{30}$. All the days of September are the sample points. Now if somebody tells you that his birth day is during the first 20 days of september than the probability of any day in september being birthday is $\frac{1}{20}$, the number of sample points in this case is 20. Thus additional information may change the probability of the happening of some events.

Let us now consider the problem of casting two dice. Let A be the event that the sum of members in the two dice is 8. Then the sample points are (2,6), (3,5), (4,4), (5,3), (6,2). Since there are 36 points in sample space we got $P(A)=\frac{5}{36}$. Let us now impose one additional condition that the 2nd die should show 4. Let this event be B. For this condition the sample space will be (1,4), (2,4), (3,4), (4,4), (5,4), (6,4). Under this additional condition, the event that the sum of the numbers shown on the faces of the two die is 8 is a conditional event and symbolically devoted by (A/B). Now the conditional probability

$$P(A/B)=\frac{1}{6}.$$

Let us now generalise the result. P(B/A) denote the probability of B given that A has already occurred. Since A is known to have occurred, it becomes a need sample space replacing the original S.

Let i = number of sample points common to A and B

j = number of sample points comprising A

n = total number of points in S.

Now, $\frac{i}{n}=P(AB)$, $\frac{j}{n}=P(A)$ and $\frac{i}{j}=P(B/A)$

$$P(B/A)=\frac{i}{j}=\frac{\frac{i}{n}}{\frac{j}{n}}=\frac{P(AB)}{P(A)}, \quad P(A)\neq 0. \Rightarrow P(AB)=P(A)P(B/A).$$

This is multiplication rule. This rule will hold only if P(A)>0

Example1.

A coin is tossed three times in succession. A is the event that there are at least two heads and B is the event that first throw gives a head. Find P(A/B).

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

A={HHH, HHT, HTH, THH}

B={HHH, HHT, HTH, HTT}

$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{\frac{3}{8}}{\frac{4}{8}}=\frac{3}{4}$$

Example 2.

Two dice Green (G) and red(R) are thrown. A is the event that the number of G is greater than number on R and B be the event than sum of the numbers on two dice is 7. Find $P(B/A)$ and $P(A/B)$

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

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

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

$$P(A/B)=\frac{P(AB)}{P(B)}=\frac{\frac{3}{36}}{\frac{6}{36}}=\frac{1}{2}$$

Example 3.

A box contains 5 black and 4 white balls. Two balls are drawn one by one without replacement. If the first drawn ball is black what is the probability that both the drawn balls are black

Let A be the event that first ball drawn is black. Then $P(A)=5/9$.

Let B be the event that 2nd ball is also black $P(B/A)=\frac{4}{8}=\frac{1}{2}$

$$P(AB)=P(A).P(P(B/A))=\frac{5}{9}.\frac{1}{2}=\frac{5}{18}.$$

Example 4.

If an event A must result in one of in two mutually exclusive events A_1 and A_2 then prove that

$$P(A)=P(A_1).P(A/A_1)+P(A_2).P(A/A_2)$$

we know $A=(A\cap A_1)\cup(A\cap A_2)$

$A\cap A_1$ and $A\cap A_2$ are mutually exclusive since A_1 and A_2 are mutually exclusive.

$$P(A)=P(A\cap A_1)+P(A\cap A_2)=P(A_1).P(A/A_1)+P(A_2)P(A/A_2).$$

Independent events: if the probability of B occurring is not affected by the occurrences or non occurrences of A then A and B are said to be independent events. In than case $P(B/A)=P(B)$ and

$$P(AB) = P(A) - P(B)$$

Conversely if $P(AB) = P(A).P(B)$ then A and B are independent.

Example.

Find the probability of a 4 turning up at least once in two tosses of a die.

Let A_1=event 4 on the first toss, A_2=event 4 on the 2nd toss.

Then $A_1 \cup A_2$=event at least one 4 turns up. A_1 and A_2 are not mutually exclusive but they are independent.

$$P(A_1 \cup A_2) = P_1(A_1) + P(A_2) - P(A_1A_2) = P(A_1) + P(A_2) - P(A_1)P(A_2)$$

$$= \frac{1}{6} + \frac{1}{6} - \frac{1}{6} \times \frac{1}{6} = \frac{11}{36}.$$

Some results on Independent events:

1. If A and B be two events such than $P(A) \neq 0, P(B) \neq 0$ and A is independent of B then B is independent of A.

A is independent of B $\Rightarrow P(A/B)=P(A)$.

$$\Rightarrow P(AB)=P(A).P(B) \Rightarrow P(BA)+P(A).P(B) \Rightarrow \frac{P(BA)}{P(A)}=P(B). \Rightarrow P(B/A)=P(B)$$

$\Rightarrow$ B is independent of A.

2. A and B are events such than $P(A) \neq 0, P(B) \neq 0$ then A and B will be independent events iff $P(AB)=P(A).P(B)$

A and B are independent $\Rightarrow P(A/B)=P(A) \Rightarrow \dfrac{P(AB)}{P(B)}=P(A) \Rightarrow P(AB)=P(A).P(B)$.

Conversely let $P(AB)=P(A).P(B)$.

Then $\dfrac{P(AB)}{P(A)}=P(B) \Rightarrow P(B/A)=P(B) \Rightarrow$ B is independent of A.

3. A and B be two independent events $A \cap B \neq P$.

A and B are independent $\Rightarrow P(A) \neq 0, P(B) \neq 0$

If possible let $A \cap B = P$. Then $P(A \cap B)=P(P) \Rightarrow P(A).P(B)=0$.

Which is a contradiction to the fact $P(A) \neq 0, P(B) \neq 0$. Hence $A \cap B \neq P$.

4. If A is independent of B then A' is independent of B.

$B=(A \cap B) \cup (A' \cap B)$ and $A \cap B$ and $(A' \cap B)$ are mutually exclusive.

$$P(B)=P(A\cap B)+P(A'B)=P(A).P(B)+P(A'\cap B)\Rightarrow P(A'B)=P(B)[1-P(A)]$$
$$=P(B).P(A')$$

This shows A' is independent of B.

Three events A_1, A_2, A_3 are independent if they are pairwise independents

$$P(A_iA_j)=P(A_i)P(A_j) \quad i\neq j \text{ where } i, j=1,2,3.$$
$$P(A_1A_2A_3)=P(A_1)P(A_2)P(A_3).$$

Baye's Theorem:

Let $B_1, B_2, ..., B_k$ be mutually exclusive events such that they form a partition of S i.e. $(B_1\cup B_2\cup ...\cup B_k)=S$. Let A be an event which can occur in conjunction with any of the events $B_i(i=1,2, ..., k)$.

The probabilities $P(B_i)$ $(i=1,2, ..., k)$ without any regard to the occurrences or not occurrences of A and conditional probabilities $P(A/B_i)(i=1,2, ..., k)$ are known. Then

$$P\left(B_i/A\right)=\frac{P(B_i).P\left(A/B_i\right)}{\sum_{r=1}^{k}P(B_r)P\left(A/B_r\right)}$$

Proof: Since A occur in conjunction with only one of the events B_i, $i=1,2, ..., k$

we have

$$A=AB_1+AB_2+...+AB_k.$$

Again AB_i are mutually exclusive.

Thus

$$P(A)=P(AB_1)+P(AB_2)+...+P(AB_k)$$
$$=P(B_1)P\left(A/B_1\right)+P(B_2).P\left(A/B_2\right)+P(B_k)P\left(A/B_k\right)$$
$$=\sum_{r=1}^{k}P(B_r).P\left(A/B_r\right) \qquad \text{(i)}$$

Again $$P(B_i/A)=\frac{P\left(B_i/A\right)}{P(A)}=\frac{P(A.B_i)}{P(A)} \qquad \text{(ii)}$$

$$P(AB_i)=P(B_iA)=P(B_i).P\left(A/B_i\right) \qquad \text{(iii)}$$

From (i),(ii) and (iii) we get,

$$P\left(\frac{B_i}{A}\right)=\frac{P(B_i).P\left(\frac{A}{B_i}\right)}{\sum_{r=1}^{k} P(B_r)P\left(\frac{A}{B_r}\right)}$$

Example. There are three urns u_1,u_2,u_3. The contents of the urns are

u_1: 1 white, 2 black, 3 red balls.

u_2: 2white, 1 black, 1 red balls.

u_3: 4 white, 5 black, 3 red balls.

Each urn is chosen at random and two balls are drawn. They are white and red. What is the probability that they come from urn u_2?

Let B_i denote the event that the ith urn is chosen. Since the urns are identical

$$P(B_1)=P(B_2)=P(B_3)=\frac{1}{3}.$$

Let A denote the event one white and one red ball are drawn.

$$P\left(\frac{A}{B_1}\right)=\frac{1}{5},\quad P\left(\frac{A}{B_2}\right)=\frac{1}{3},\quad P\left(\frac{A}{B_3}\right)=\frac{2}{11},\quad P(AB_1)=P(B_1)\,P\left(\frac{A}{B_1}\right)=\frac{1}{3}.\frac{1}{5}$$

$$P(AB_2)=P(B_2)\,P\left(\frac{A}{B_2}\right)=\frac{1}{3}.\frac{1}{3}.\qquad P(AB_3)=P(B_3)\,P\left(\frac{A}{B_3}\right)=\frac{1}{3}.\frac{2}{11}$$

$$P(A)=P(AB_1)+P(AB_2)+P(AB_3)=\frac{1}{3}\left(\frac{1}{5}+\frac{1}{3}+\frac{2}{11}\right)=\frac{1}{3}.\frac{118}{165}$$

$$P\left(\frac{B_2}{A}\right)=\frac{P(AB_2)}{P(A)}=\frac{\frac{1}{3}.\frac{1}{3}}{\frac{1}{3}.\frac{118}{165}}=\frac{165}{3\times 118}=\frac{55}{118}.$$

Example 2.

Box I contains 3 red and 2 blue marbles, box II contains 2 red and 8 blue marbles. A coin is tossed. If it is head, a marble is chosen from box I and if it is tail then the marble is chosen from box II. If the person who tosses the coin does not reveal it head or tail but reveals that a red marble is chosen, what is the probability that box I is chosen ?

Let R, I, II denote the event that a red marble is chosen, box I and box II is chosen respectively.

$$P(I)+P(II)=\frac{1}{2}\qquad P\left(\frac{R}{I}\right)=\frac{3}{5}\qquad P\left(\frac{R}{II}\right)=\frac{2}{10}.$$

By Bay'es theorem.

$$P\left(I/_{R}\right)=\frac{P(I)\,P\left(R/_{I}\right)}{P(I)P\left(R/_{I}\right)+P(II)P\left(R/_{II}\right)}$$

$$=\frac{\frac{1}{2}\cdot\frac{3}{5}}{\frac{1}{2}\cdot\frac{3}{5}+\frac{1}{2}\cdot\frac{2}{10}}=\frac{\frac{3}{5}}{\frac{3}{5}+\frac{2}{10}}=\frac{6}{6+2}=\frac{3}{4}.$$

Random Variable

Let to each point of a sample space we assign a number. Then we get a function defined on the sample space. This function is called a random (or stochastic) variable. Sometimes it is also called random function.

Let us consider the experiment of tossing two coins. The sample space is S= {HH, HT, TH, TT}. The outcome is qualitative and are described by attributes. Now let us find a numerical valued function of the sample points. Let us consider the number of heads shown and devote this by X. We will exhibit it in a table

Sample point	HH	HT	TH	TT
X	2	1	1	0

Corresponding to the sample point HH, we set X=2, to the sample points HT (or TH) we set X=1, and to the point TT we set X=0

Let us now consider three coin tossing experiment. The sample space is

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

Let us now find real valued function by considering the number of heads shown and denote it by X.

Sample point	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
X	3	2	2	2	1	1	1	0

In the first case, $P(X=2)=\frac{1}{4}$, $P(X=1)=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$, $P(X=0)=\frac{1}{4}$.

In the 2nd case $P(X=3)=\frac{1}{8}$, $P(X=2)=\frac{3}{8}$, $P(X=1)=\frac{3}{8}$, $P(X=0)=\frac{1}{8}$.

Let us how consider an experiment of throwing 2 dice. The sample space has 36 points. Let X be the sum of the numbers shown in the two dice. For each sample point we get a numerical value.

There are some sample points which may correspond to the same numerical values. For example the points (1,3),(2,2) and (3,1) will give X=4 but X=12 is obtained from only one point (6,6), X cannot have values 0 and 1.

$P(X=2)=\frac{1}{36}$, $P(X=3)=\frac{1}{18}$, $P(X=4)=\frac{1}{12}$, $P(X=5)=\frac{1}{9}$

$P(X=6)=\frac{5}{36}$, $P(X=7)=\frac{1}{6}$, $P(X=8)=\frac{5}{36}$, $P(X=9)=\frac{1}{9}$

$P(X=10)=\frac{1}{12}$ $P(X=11)=\frac{1}{18}$, $P(X=12)=\frac{1}{36}$.

Here we discuss random variables which assumes only integral values. Such random variables are known as discrete random variables.

Discrete Random Variables:

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

$$P(X=x_k)=f(x_k) \qquad k=1,2 \ldots\ldots$$

$$\text{Note that } \sum_{k=1}^{\infty} P(X=x_k)=1.$$

Let us now introduce the probability function by $P(X=x)=f(x)$. For $x=x_k$, $P(X=x)=f(x)$ reduces to $P(x=x_k)=f(x_k)$, which for other values of x, $f(x)=0$. Thus we get $f(x)\geq 0$ and $\sum_x f(x)=1$.

The distribution function (cumulative) for a random variable X is defined by $F(x)=P(X\leq x)$ where x is any real numbers $(-\infty<x<\infty)$

$$\lim_{x\to-\infty} F(x)=0, \qquad \lim_{x\to\infty} F(x)=1$$

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

Continuous random variables:

A non-discrete random variable X is said to be continuous if its distribution function may be represented as

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

where f(u) has the properties 1. $f(x)\geq 0$ 2. $\int_{-\infty}^{\infty} f(u)du=1$

If X is a continuous random variable, the probability that X tosses on any are particular value is zero

where as informal probability is $\int_a^b P(a<X<b)=f(u)du.$

It now follows that if X is a continuous random variable, then the probability that X takes on any one particular value is zero, but when X lies between two different values it is given by

$$P(a<X<b)=\int_a^b f(u)du.$$

Mathematical Expectation:

For a discrete random variable X, having the values x_1, x_2, x_3------x_u the mathematical expectation is expected value or expectation of X is defined as

$$E(X) = x_1 P(X=x_1) + x_2 P(X=x_2) + \ldots\ldots + x_u P(X=x_u) = \sum_{i=1}^{u} x_i P(X=x_i)$$

$$= \sum_{i=1}^{u} x_i p_i \text{ where } p_i = P(X=x_i)$$

If all the probabilities are equal and each equal to $\frac{1}{n}$ then

$$E(X) = \frac{x_1 + x_2 + \ldots + x_n}{n}$$ which is the A.M of $x_1, x_2, \ldots, x_n$ and so $E(X) = \bar{x}$

Example 1. A dice is thrown. Find the expected value of the number X shown on the face.

X can assume the values 1,2,3,4,5,6 with equal Probability $\frac{1}{6}$

$$E(X)=1.\frac{1}{6}+2.\frac{1}{6}+3.\frac{1}{6}+4.\frac{1}{6}+5.\frac{1}{6}+6.\frac{1}{6}=7/2=3.5$$

Example 2. Two dice are thrown. If X is the sum of the numbers shown on the faces of the two dice, find E(X) X can take values 2,3,4,5,6,7,8,9,10,11, and 12 with corresponding probabilities

$$\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36}$$ (Already discussed)

$$E(X)=\frac{1}{36}(2+6+12+20+30+42+40+36+30+22+12)=7$$

The expected value may not be any value of X. Though in the second example we see that the expected value 7 is one of the values of X, but in the first example the expected number 3.5 is different from the values of the variable X.

E(X) may or may not exist when X takes infinite number of values. The expected value, E(X) when exists is called the population mean of the probability distribution and is denoted by μ. Hence $E(X)=\mu=\bar{x}$

Expected value of functions of random variables:

The expectation of a function G(X) of a random variable is defined by

$$E[g(X)]=\sum_{i=1}^{n} g(x_i)P(X=x_i)=\sum_{i=1}^{n} g(x_i)p_i$$

where $x_1, x_2, x_3, \ldots, x_n$ are the values of the variable X and $p_1, p_2, \ldots, p_n$ are the corresponding probabilities.

Example: A random variable X has uniform distribution over the integers 1,2, ..., 10 with equal probability. Find the expected value of X and X^2.

X assumes the values 1,2, ..., 10 with equal probability $\frac{1}{10}$.

$$E(X)=\frac{1}{10}(1+2+\ldots+10)=5.5$$

$$E(X^2)=\frac{1}{10}(1+2^2+3^2+\ldots+10^2)=38.5.$$

Properties of expectation:

(i) E(c)=c where c is a constant

$$E(c)=\sum_{i=1}^{n} CP(X=c)=C\sum_{1}^{n} P(X=C)=C \text{ since } \sum_{1}^{n} P(X=c)=1$$

(ii) E(aX)=aE(X).

Let X be a random variable and 'a' be a constant. X assumes values $x_1, x_2, \ldots, x_n$ with probabilities $P(X=k_i)=p_i$

Let g(X) be function of X. Then $E[g(X)]=E(ax)=\sum_{i=1}^{n} ax_i P(x=x_i)$

$$=a\sum_{i=1}^{n} x_i P(x=x_i)=a E(X)$$

(iii) $E(ax+b) = aE(x) + b$ where 'a' and 'b' are constants.

Let X be a random variable which assumes values $x_1, x_2 \cdots x_n$ with probability $P(X=x_i)=p_i$

$$E(aX+b)= \sum_{i=1}^{n} (ax_i+b)P(X=x_i)= \sum_{i=1}^{n} ax_i P(X=x_i) + \sum_{i=1}^{n} b\,P(X=x_i)$$

$$=a \sum_{i=1}^{n} x_i P(X=x_i) + b \sum_{i=1}^{n} P(X= x_i)$$

$$= aE(X) + b.$$

(iv) $E[g(x)+c]=E[g(x)]+c$ by property (iii)

(v) $E[cg(x)]=cE[g(x)]$ by property (ii)

Joint distributions : Let X and Y be two discrete random variables. Then the joint probability function of X and Y are defined by

$P(X=x, Y=y) = f(x,y)$. where

1. $f(x,y) \geq 0$
2. $\sum_{x} \sum_{y} f(x,y)=1.$

If X assumes any one of the values $x_1, x_2, \ldots, x_n$ and Y assumes any one of the values $y_1, y_2, \ldots, y_m$ then the probability of the event that $X=x_i$ and $Y=y_j$ is given by

$$P(X=x_i, Y=y_j)=f(x_i,y_j).$$

When both the variables are continuous the joint probability function for the random variable X and Y is defined by $f(x,y) \geq 0$ and $\int_{-\infty}^{\infty} \int_{\infty}^{\infty} f(x,y)\,dxdy=1.$

Expectation of sum: If X an Y are two random variables then $E(X+Y) = E(X) + E(Y)$.

Proof: Let $f(x,y)$ be the joint probability function of X and Y. Then

$$E(X+Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i+y_j)f(x_i,y_j)$$

where $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$ are the values of X and Y and $f(x_i,y_j) = P(X=x_i, Y=y_j)$

$$E(x+y)= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i f(x_i,y_j) + \sum_{i=1}^{n} \sum_{j=1}^{m} y_j f(x_i, y_j) =E(X) + E(Y).$$

Independent random variables:

Let X and Y be two discrete random variables. If the events X=x and Y=y are independent events for all x and y then X and Y are independent random variables. In such case,

$$P(X=x, Y=y)=P(X=x)P(Y=y) \quad \text{or} \quad f(x,y)=f_1(x).\ f_2(y)$$

Conversely if for all x and y the joint probability function f(x,y) can be expressed as the product of a function of x alone and a function of y alone then X and Y are independent.

Mathematical Expectation of product:

If X and Y are independent random various then E(XY)=E(X).E(Y).

Proof: Let f(x,y) be the joint probability function of the discrete random variables. Since X and Y are independent random variables,

$$f(x,y)=f_1(x).f_2(y)$$

where $f_1(x)$ is the probability function of X and $f_2(y)$ is that of Y.

$$E(XY)=\sum_x\sum_y xy\, f(x,y)=\sum_x\sum_y xy\, f_1(x) f_2(y)=\sum_x x\, f_1(x).\sum_y y\, f_2(y)=E(X).E(Y)$$

Variance: Let X be a random variable with expected value E(X)=μ. The variance of X is defined as $E[(x-\mu)^2]$ and is denoted by Var (X)

$$\begin{aligned} Var(X)&=E[(X-\mu)^2] \\ &=E(X^2-2\mu X+\mu^2) \\ &=E(X^2)-2\mu E(X)+\mu^2 \\ &=E(X^2)-2\mu^2+\mu^2 \\ &=E(X^2)-\mu^2 \\ &=E(X^2)-[E(X)]^2 \end{aligned}$$

From definition it is clear that variance is a non-negative number. The positive square root of varriance is called standard deviation. The variance is also denoted by σ^2

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

Standard deviation (S.d) = σ. S.d is vey impertant concept of statistics.

Probability distribution of functions of random variables:

Let us consider a problem of drilling for oil. From record on the average only one in 10 drilled wells strikes oil. Let x be the number of drilling unit the first strike of well. Here x is a random variable taking values 1,2,3,..... Let the experiment of drilling be independent.

The event [X=1] denotes that oil is striken at in the first drilling.

$$P(X=1)=0.1.$$

If the first drilling is a failure and second drilling yields oil then X=2.

Then the two events are independent and probability of first is 0.9

$$\text{So,} \quad P(X=2) = (0.9)\times(0.1) = 0.09.$$

similarly $P(X=3)=(0.9)\times(0.9)\times(0.1)=0.081$ and so an.

$$\text{In general } P(X=n)=(0.9)^{(n-1)}(0.1),\ n=1,2,.....$$

Let us now define distribution function as $F(x)=P[X\leq x]$

Then $F(1)=P(X\leq 1)=0.1$

$F(2)=P(X\leq 2)=P(X=1)+P(X=2)=0.1+0.09=0.19$

$F(3)=P(x\leq 3)=0.1+0.19+0.081=0.371$ and so an.

In general $F(x)=P(X\leq k)$

Thus $F(x)$ is cumulative distribution function

Let X be a discrete random variable with probability function $f(x)$. Then $Y=g(X)$ is also discrete random variable and probability function of Y is

$$h(y)=P(Y=y)=\sum P(X=x)=\sum f(k).$$

If X takes values $x_1, x_2, ..., x_n$ and Y the values $y_1, y_2,...., y_m$ $(m\leq n)$ then

$$y_1h(y_1)+y_2h(y_2)+.......+y_mh(y_m)=g(x_1)f(x_1)+g(x_2)f(x_2)+......+g(x_n)f(x_n).$$

$$E[g(X)]=\sum g(x_i)f(x_i).$$

It X is a continuous random variable having probability density $f(x)$ then

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

If X and Y are two continuous random variables having joint probability density function $f(x,y)$ the

$$E[g(x,y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy.$$

Example: Two dice are thrown. If X is the sum of the numbers shown on the faces of the two dice, find the s.d. of the sum obtained.

$E(X)=7$ (Already calculated)

$$E(X^2)=2^2.\frac{1}{36}+3^2.\frac{2}{36}+4^2.\frac{3}{36}+5^2.\frac{4}{36}+6^2.\frac{5}{36}+7^2.\frac{6}{36}+8^2.\frac{5}{36}+9^2.\frac{4}{36}+10^2.\frac{3}{36}+11^2.\frac{2}{36}+12^2.\frac{1}{36}$$

$= 54.83$

$$\sigma^2=E(X^2)-[E(x)]^2=54.83-7^2=5.83,$$

$$\text{s.d}=\sqrt{5.83}=2.41$$

Theorems on variance:

Theorem 1. If a be a constant and x be a random variable then

$$Var(cx)=c^2 var(X)$$

$$Var(cX)=E[(cX)^2]-E[(cX)]^2$$
$$=E[(c^2X^2)-c^2[E(X)]^2$$
$$=c^2E(X^2)-c^2[E(X)]^2$$
$$=c^2[E(X^2)-\{E(X)\}^2]=c^2 Var(X)$$

cor. $Var(a+bx)=b^2 var(x)$

Theorem 2. $E[(X-a)^2]$ is minimum when $a=E(X)$.

$$E[(X-a)^2]=E[\{(X-\mu)+(\mu-a)\}^2] \qquad \text{where } \mu=E(x).$$
$$=E[(x-\mu)^2+2(x-\mu)(\mu-a)+(\mu-a)^2]$$
$$=E(x-\mu)^2+2(\mu-a)E(x-\mu)+(\mu-a)^2$$
$$=E(x-\mu)^2+(\mu-a)^2 \qquad \text{since } E(x-\mu)=E(x)-\mu=0$$

Thus $E[(x-a)^2]$ will attain minimum value when $a=\mu$ i.e. when $a=E(x)$.

Theorem 3. $Var(X+Y)=Var(X)+Var(Y)$

Let $E(X)=\mu_x$ and $E(Y)=\mu_y$

Then $Var(X+Y)=E[\{(x+y)-(\mu_x+\mu_y)\}^2]=E[\{(x-\mu_x)+(y-\mu_y)\}^2]$

$$=E[(x-\mu_x)^2]+2E[(x-\mu_x)(y-\mu_y)]+E[(y-\mu_y)^2]$$

We know that $E[(x-\mu_x)(y-\mu_y)]=E[(x-\mu_x)].E[(y-\mu_y)]=0$ if X and Y and $(X-\mu_x)$ and $(Y-\mu_y)$ are independent.

Hence $Var(X+Y)=Var(X)+Var(Y)$.

Similarly it can be proved that $Var(X-Y)=Var(X)-Var(Y)$.

Example 1. A coin is tossed three times and x is the number of heads shown. Find E(x) and var (x).

The sample space S is

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

x assumes values 3,2,1,0 with probabilities $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$. respectively.

$$E(x)=3.\frac{1}{8}+2.\frac{3}{8}+1.\frac{3}{8}+0.\frac{1}{8}=\frac{1}{8}(3+6+3)=\frac{3}{2}.$$

$$E(x^2)=9.\frac{1}{8}+4.\frac{3}{8}+1.\frac{3}{8}+0.\frac{1}{8}=3$$

$$Var(x)=E(x^2)-[E(x)]^2=3-\frac{9}{4}=\frac{3}{4}$$

Example 2. In a lottery there are 200 prizes of Rs 500.00, 20 prizes of Rs 2500.00 and 5 prizes of Rs 10,000.00. Assuming there 10,000 tickets are to be issued and sold what is the fair price to pay for a ticket?

Let X be a random variable denoting the amount of money to be won on a tickets. The values of X together with corresponding probabilities are shown below

X:	500	2500	10000
P(X):	$\frac{200}{10000}$	$\frac{20}{10000}$	$\frac{5}{10000}$

$$E(x)=500\times 0.02+2500\times 0.002+10,000\times 0.0005=20$$

Thus fair price to pay for a ticket is Rs 20.00

Bernoullian Trails :

If a random experiment be such that there are only two possible out-course ('Success' and 'failure' type), the trail corresponding to this type to random experiment is called a Bernoullian trial.

A random variable X which assumes only two values say 0 and 1 is called a Bernoulli random variable X.

Let $X = 1$ when the trial gives a success with probability p

$\quad\quad 0$ otherwise with probability $1-p=q$.

In notation, $P(X=1)=p$

$P(X=0)=q$

where $0<p<1$ and $p+q=1$.

The above probabilities define an probability distribution of the random variable X.

$$E(X)=1.p+0.q=p. \qquad E(X^2)=1^2.p+0^2.q=p.$$

$$Var(X)=E(X^2)-\{E(X)\}^2=p-p^2=p(1-p)=pq.$$

If we are to find the probability $P(X=x)$ we shall see that in one trial, if x be the number of success then $(1-x)$ will be number of failures. Then $P(X=x)=p^x q^{1-x}$ $\quad x=0,1,2,.....$

For example, in throwing a die, if we regard the appearance of an even number as success and appearance of the odd number as failure then with $p=\frac{1}{2}$, $q=\frac{1}{2}$ we have

$$P(X=1)=\frac{1}{2}, \quad P(X=0)=\frac{1}{2}$$

where X=1 represent success and X=0 represent failure.

In the same experiment if we regard appearance of six as success and appearance of other numbers as failure then

$$P(X=1)=\frac{1}{6} \text{ and } P(X=0)=\frac{5}{6}$$

Again if two dice are thrown and appearance of six is a success and appearance of other numbers a failure then

$$P(X=2)=P(SS)=\left(\frac{1}{6}\right)^2$$

$$P(X=1)=P(SF)=P(FS)=\frac{1}{6}.\frac{5}{6}+\frac{5}{6}.\frac{1}{6}=\frac{5}{18}$$

$$P(X=0)=P(FF)=\left(\frac{5}{6}\right)^2$$

Moments :

The r th moment of random variable X about any point x = c is defined as $E(X^r)$ and is denoted by μ'_r. This is sometimes known as r th raw moment.

$$\mu'_r = E(X-c)^r = \sum_i (x_i - c)^r.p_i = \frac{1}{N}\sum_i f_i(x_i - c)^r, \quad N = \sum_i f_i, \quad p_i = \frac{f_i}{N}$$

$$= \frac{1}{N}\sum_i f_i d_i^{\,r}, \quad \text{where } d_i = x_i - c.$$

In Particular 2nd moment about c will be

$$\mu'_2 = E(X-c)^2 = \sum_i (x_i - c)^2.p_i$$

The rth moment about the origin is defined as

$$\mu'_r = E(X^r) = \sum_i x_i^{\,r}.P(X = x_i) = \sum_i x_i^{\,r}.p_i \quad \text{where } \sum_i p_i = 1$$

or $$\mu'_r = \frac{1}{N}\sum_i f_i x_i^{\,r}, \quad \sum_i f_i = N \quad \text{since } p_i = \frac{f_i}{N}.$$

In particular, $\mu_1' = E(X) = \frac{1}{N}\sum_i f_i x_i$ = Arithmetic mean.

$\mu_2' = E(X^2) = \Sigma x_i^2 p_i$ is 2nd moment about origin

similarly $\mu_3' = E(X^3) = \Sigma x_i^3 p_i$ is 3rd moment about the origin and so on.

Replacing c by the arithmetic mean $\bar{x}$ in the formula of rth momnt about a point x = c, we get the r th moment of random variable X about the mean $\bar{x}$. This moment is called r th central moment and is denoted by μ_r. Thus

$$\mu_r = E(X-\bar{x})^r = \frac{1}{N}\sum_i f_i (x_i - \bar{x})^r \qquad r = 0,1,2,......$$

Now
$$\mu_0 = E(X-\bar{x})^0 = \Sigma p_i = 1.$$
$$\mu_1 = E(X-\bar{x}) = E(X) - \bar{x} = \bar{x} - \bar{x} = 0$$
$$\mu_2 = E(X-\bar{x})^2 = \sigma^2$$

This shows that the second central moment or second moment about the mean is the variance.

In case of continuous variable

$$\mu_r = \int_{-\infty}^{\infty} (x-\mu)^r f(x) dx$$

Relationship between moment about mean and about any point :

We know that if $d_i = x_i - c$, then $\bar{x} = A + \frac{1}{N}\sum_i f_i d_i = A + \mu_1'$, where A is the assumed mean.

Now, By definition,

$$\mu_r = E(X-\mu)^r = \frac{1}{N}\sum_i f_i (x_i - \bar{x})^r$$

$$= \frac{1}{N}\sum_i f_i (x_i - c + c - \bar{x})^r$$

$$= \frac{1}{N}\sum_i f_i (d_i + A - \bar{x})^r, \quad \text{where } d_i = x_i - A$$

$$= \frac{1}{N}\sum_i f_i (d_i - \mu_1')^r$$

$$= \frac{1}{N}\sum_i f_i (d_i^r - {}^rC_1 d_i^{r-1}\mu_1' + {}^rC_2 d_i^{r-2}\mu_1'^2 - {}^rC_3 d_i^{r-3}\mu_1'^3 + ... + (-1)^r \mu_1'^r)$$

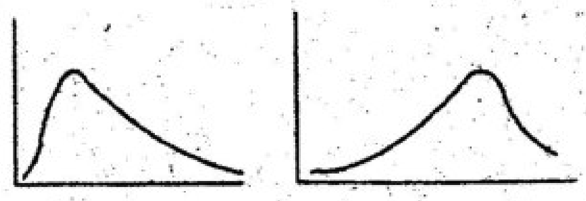
$$= \mu'_r - {}^rC_1\mu'_{r-1}\mu'_1 + {}^rC_2\mu'_{r-2}\mu'^2_1 - {}^rC_3\mu'_{r-3}\mu'^3_1 + \dots + (-1)^r\mu'^r_1$$

In particular, $\mu_2 = \mu'_2 - \mu'^2_1$, $\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1$, $\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1$ and so on.

Skewness and kurtosis :

Let us now apply different moments in measuring skewness and kurtosis.

Skewness : Often a distribution is not symmetric about any value instead it has one of its tails longer then the other. If the longer tail occurs to the right then the distribution is said be skewed to the right. If the longer tail occurs to the left then it is said to be skewed to the left. Measures describing this asymmetry are called coefficient of skewness.



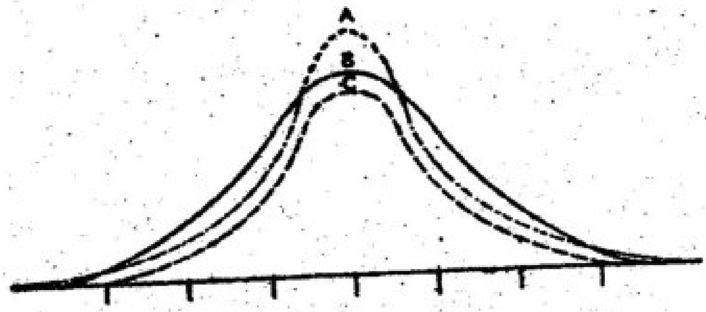
Skewness to the right Skeness to the left.

Let us define the parameter β_1 in terms of moments as $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

The measure $\gamma_1 = \pm\sqrt{\beta_1} = \pm\frac{\mu_3}{\sqrt{\mu_2^3}} = \pm\frac{\mu_3}{\sigma^3}$

will give the measure of skewness. The measure will be positive or negative according as the distribution is skewed to the right or left respectively. For symmetric distribution $\gamma_1 = 0$.

Kurtosis : In some cases a distribution may have its values concentrated near the mean so that the distribution has a large peak. In other cases the distribution may be relatively flat. Measures of degree of peakedness of a distribution are called coefficient of kurtosis.



Let us define another parameter β_2 in terms of moments as $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$. The measure $\gamma_2 = \beta_2 - 3$ will give the measure of kurtosis.

Moment generating function (m.g.f):

The moment generating function of the discrete probability distribution of a discrete random variable X about a point a is defined by

$$M_{x=a}(t) = E\left(e^{t(X-a)}\right) = \sum_i e^{t(x_i-a)} p_i$$

where t is a real parameter.

$$M_{x=a}(t) = \sum_i \left[1 + t(x_i - a) + \frac{t^2}{2!}(x_i - a)^2 + \frac{t^3}{3!}(x_i - a)^3 + \dots + \frac{t^r}{r!}(x_i - a)^r + \dots\right] p_i$$

$$= \sum_i p_i + t\sum_i (x_i - a)p_i + \frac{t^2}{2!}\sum_i (x_i - a)^2 p_i + \dots\dots + \frac{t^r}{r!}\sum_i (x_i - a)^r p_i + \dots$$

$$= 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \dots\dots + \frac{t^r}{r!}\mu_r' + \dots \quad (1)$$

where μ_r' is the moment of order 'r' about a. Thus $M_{x=a}(t)$ generates moments and hence the name.

Differentiating the relation (1) 'r' times w.r.t 't' we have

$$\mu_r' = \frac{d^r}{dt^r}\left[M_{x=a}(t)\right]_{t=0}$$

Again

$$M_{x=a}(t) = E\left(e^{t(X-a)}\right) = \sum_i e^{t(x_i-a)} p_i = e^{-at}\sum_i e^{tx_i} p_i = e^{-at} M_0(t)$$

where $M_0(t)$ is the m.g.f about the origin.

Thus if f(x) be probability distribution function then $M_0(t)=E(e^{tx})$

$$M_0(t) = \sum_i e^{tx_i} f(x) \qquad \text{for discrete variable}$$

and $$M_0(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx \qquad \text{for continuous variable}$$

Theorem : If X and Y are independent random variables having moment generating functions $M_x(t)$ and $M_y(t)$ respectively then $M_{x+y}(t) = M_x(t).M_y(t)$.

Proof: Since X and Y are independent random variables, any function of X and Y are also independent. Hence e^{tx} and e^{ty} are independent.

$$M_{x+y}(t)=E[e^{t(x+y)}]=E[e^{tx}.e^{ty}]=E(e^{tx}).E(e^{ty})=M_x(t).M_y(t).$$

Change of origin and Scale;

Let us introduce a new variable u defined by

$$u_i = \frac{x_i - a}{h}$$

Then

$$M_X(t) = E(e^{tX}) = \sum_i e^{tX_i} p_i = \sum_i e^{t(a+u_i h)} p_i = e^{at} \sum_i e^{u_i ht} p_i$$

$$\Rightarrow M_X\left(\frac{t}{h}\right) = e^{\frac{at}{h}}.M_u(t) \quad \text{writing} \quad \left(\frac{t}{h}\right) \text{ for } t$$

Example: The random variable X can assume the values 1 and -1 with probability $\frac{1}{2}$ each. Find the m.g.f and first four moments about the origin.

$$E\left(e^{tX}\right) = e^{t.1}\left(\frac{1}{2}\right) + e^{t.(-1)}\left(\frac{1}{2}\right) = \frac{1}{2}\left(e^t + e^{-t}\right)$$

Again $\quad e^t = 1 + \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + ...\quad$ and $\quad e^{-t} = 1 - \frac{t}{1!} + \frac{t^2}{2!} - \frac{t^3}{3!} + ...$

$$\Rightarrow e^t + e^{-t} = 2.\left(1 + \frac{t^2}{2!} + \frac{t^4}{4!} + ...\right)$$

Now

$$M_X(t) = E(e^{tX}) = \frac{1}{2}(e^t + e^{-t}) = 1 + \frac{t^2}{2!} + \frac{t^4}{4!} + ...$$

Also

$$M_X(t) = 1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \mu_4' \frac{t^4}{4!} + ... + \mu_r' \frac{t^r}{r!} + ...$$

By comparison, $\quad \mu = 0, \quad \mu_2' = 1, \quad \mu_3' = 0, \quad \mu_4' = 1, \quad ......$

This shows that all odd moments are zero and all even moments are 1.

Charactaristic functions :

Let us replace t by 'iw' in m.g.f.

Then

$$M_X(iw) = E(e^{iwX})$$

This function is called characteristic function and is denoted by $\varphi_x(w)$.

Hence $\varphi_x(w) = \Sigma e^{iwX} f(x)$ for discrete variable and

$\varphi_x(w) = \int_{-\infty}^{\infty} e^{iwX} f(x)\, dx$ for continuous variable.

Since $|e^{iwx}| = 1$ and $|\phi_X(w)| \le 1$ the integral always converges absolutely.

From

$$M_X(t) = 1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \mu_4' \frac{t^4}{4!} + \ldots\ldots + \mu_r' \frac{t^r}{r!} + \ldots\ldots$$

we get,

$$\phi_X(w) = 1 + i\mu_1' w - \mu_2' \frac{w^2}{2!} - i\mu_3' \frac{w^3}{3!} + \ldots\ldots + i^r \mu_r' \frac{w^r}{r!} + \ldots\ldots$$

where $\mu_r' = E[X^r]$ is the the 'r'th moment of X about the origin.

Theorem : If $\phi_X(w)$ is the characteristic function of the random variable X and a and b ($b \neq 0$) are constants then the characteristic function of $\frac{X+a}{b}$ is $\phi_{\frac{X+a}{b}}(w) = e^{\frac{aiw}{b}} \phi_X\left(\frac{w}{b}\right)$

Proof: Let $u = \frac{X+a}{b}$.

Then $\phi_X(w) = \sum_x e^{iwX} p(x) = \sum_u e^{iw(bu-a)} p(u) = e^{-iaw} \sum_u e^{ibwu} p(u)$

Replacing w by $\frac{w}{b}$ we have $\phi_X\left(\frac{w}{b}\right) = e^{-ai\frac{w}{b}} \sum_u e^{iwu} p(u)$

$$\Rightarrow \phi_{\frac{X+a}{b}}(w) = e^{\frac{aiw}{b}} \phi_X\left(\frac{w}{b}\right)$$

Theorem : If X and Y are independent random variables with charactaristic functions $\phi_X(w)$ and $\phi_Y(w)$ respectively then $\phi_{X+Y}(w)=\phi_X(w).\phi_Y(w)$

Proof : Since X and Y are independent random variables, any function of X and any function of Y are independent. Hence e^{iwX} and e^{iwY} are independent.

$$\phi_{X+Y}(w)=E[e^{iw(X+Y)}]$$

$$=E[e^{(iwX+iwY)}]=E[e^{(iwX}.e^{iwY)}]$$

$$=E(e^{(iwX)})\cdot E(e^{iwY})=\phi_X(w).\phi_Y(w)$$

Properties :

1. $\phi_X(0)=1$.

Putting $w=0$ in $\phi_X(w)=\Sigma e^{iwX} f(x)$ we get $\phi_X(0)=\Sigma f_x(x)=1$, since sum to the probabilities is 1

2. $\phi_X(w)$and $\phi_X(-w)$ are conjugate.

We know, $\phi_X(w)=\Sigma e^{iwX} f(x)$. Replacing w by $-w$ we get, $\phi_X(-w)=\Sigma e^{-iwX} f(x)$

This shows that $\phi_X(w)$ and $\phi_X(-w)$ are conjugate to each other.

Probability generating function (p.g.f):

When the random variable X takes only integral values, the probability generating function P(S) in determined by $P(S)=p_0+p_1s+p_2s^2+......=\sum_{u=0}^{\infty} p_n s^n=E(S^X)$

The coefficient of S^n in the expansion of P(s) gives $P(X=n)$.

$$\text{Now} \quad \left(\frac{\partial P}{\partial S}\right)_{s=1}=\sum_{n=0}^{\infty} n\,P_n=\mu_1', \qquad \left(\frac{\partial^2 P}{\partial S^2}\right)_{s=1}=\sum_{0}^{\infty} n(n-1)P_n=\mu_2' - \mu_1'$$

Then

$$\mu_2'=\mu_2'-\mu_1'^2=\left(\frac{\partial^2 P}{\partial S^2}+\frac{\partial P}{\partial S}-\left(\frac{\partial P}{\partial S}\right)^2\right)_{S=1}$$

Theorem : The probability generating function of the Sum of two independent random variables is equal to the product of their p.g.f.

Proof : Let X_1 and X_2 be two independent random variables.

Then $P_{X_1+X_2}(S)=E\left(S^{X_1+X_2}\right)=E\left(S^{X_1}.S^{X_2}\right)=E\left(S^{X_1}\right)E\left(S^{X_2}\right)$ since X_1 and X_2 are independent.

$$=P_{X_1}(S).P_{X_2}(S)$$

Deductions: 1. $P_{X+1}(s)=\sum_{n=0}^{\infty} p_{n} s^{n+1}=s\sum_{n=0}^{\infty} p_n s^n=sP(s)$

2. $P_{2X}(s)=\sum_{n=0}^{\infty} p_n s^{2n}=\sum_{n=0}^{\infty} p_n (s^2)^n=P(s^2)$

Covariance : Let X and Y be two random variables with respective mean μ_X and μ_Y. In case of two variables X and Y we define another quantities known as co-variable by

$$\sigma_{XY}=\text{cov}(X,Y)=E[(X-\mu_X)(Y-\mu_Y)]$$

Theorem 1 : $\sigma_{XY}=E(XY)-E(X)E(Y)$

Proof : $\sigma_{XY}=\text{Cov}(X,Y)$

$=E[(X-\mu_X)(Y-\mu_Y)]$

$=E[XY-\mu_X Y-\mu_Y X+\mu_X\mu_Y)]$

$=E(XY)-\mu_X E(Y)-\mu_Y E(X)+E(\mu_X\mu_Y)$.

$=E(XY)-\mu_X\mu_Y-\mu_Y\mu_X+\mu_X\mu_Y$

$=E(XY)-\mu_X\mu_Y$

$=E(XY)-E(X)E(Y)$

Theorem 2: If X and Y are independent random variables then $\sigma_{XY}=\text{Cov}(X,Y)=0$

Proof. : Since X and Y are independent, $E(XY)=E(X)E(Y)$.

Now $\text{Cov}(X,Y)=E(XY)-E(X)E(Y)=E(X)E(Y)-E(X)E(Y)=0$

Chebyshev's Inequality :

Let X be a random variable with mean μ and variance σ^2 (finite). Then for given $\varepsilon>0$.

$$P(|X-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{\varepsilon^2}$$

Proof : If f(x) be the density function of X then

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

since the integral is non negative the value of the integral can only decrease when the rang of integration is diminished. Therefore,

$$\sigma^2\geq\int_{|x-\mu|\geq\varepsilon}(x-\mu)^2 f(x)dx\geq\varepsilon^2\int_{|x-\mu|\geq\varepsilon}f(x)dx$$

Again $P[|X-\mu|\geq\varepsilon]=\int_{|x-\mu|\geq\varepsilon}f(x)dx$.

$$\therefore\sigma^2\geq\varepsilon^2 P(|X-\mu|\geq\varepsilon)\Rightarrow P(|X-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{\varepsilon^2}.$$

Deduction : If $\varepsilon = k\sigma$ then $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

Taking k=2, $P(|X-\mu| \geq 2\sigma) \leq 0.25$ or $P(|X-\mu| < 2\sigma) \geq 0.75$

Law of Large Numbers :

Let $X_1, X_2, \ldots\ldots, X_n$ be 'n' mutually independent random variables each having finite main μ and variance σ^2.

Let $S_n = X_1 + X_2 + \ldots\ldots + X_n$ $\quad$ (n=1,2, ..., n)

Then for given $\varepsilon > 0$, $\quad \lim_{n\to\infty} P(|\frac{S_n}{n} - \mu| \geq \varepsilon) = 0$

Proof : We have $E(x_1) = E(x_2) = \ldots\ldots = E(x_n) = \mu$ and $Var(X_1) = Var(X_2) = \ldots\ldots = Var(X_n) = \sigma^2$

$$\text{Then, } E\left(\frac{S_n}{n}\right) = E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{1}{n}[E(X_1) + E(X_2) + \ldots + E(X_n)]$$

$$\text{So, } E\left(\frac{S_n}{n}\right) = \frac{1}{n}[n\mu] = \mu$$

$Var(S_n) = Var(X_1 + X_2 + \ldots\ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$,

since $X_1, X_2 \ldots X_n$ are independent:

$$\Rightarrow Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2} Var(S_n) = \frac{\sigma^2}{n}$$

Now put $X = \frac{S_n}{n}$ in Chebyshev's inequality. Then $P\left(|\frac{S_n}{n} - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$

Taking limit as $n \to \infty$ we have

$$\lim_{n\to\infty} P\left(|\frac{S_n}{n} - \mu| \geq \varepsilon\right) = 0$$

Deduction : We have $P\left(|\frac{S_n}{n} - \mu| < \varepsilon\right) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$

Taking limit as $n \to \infty$ we have,

$$\lim_{n\to\infty} P\left(|\frac{S_n}{n} - \mu| < \varepsilon\right) = 1$$

Bernonlli's Theorem : Let X be number of successes in 'n' independents trials with constant probability 'p' of success. Then given $\varepsilon > 0$, $\delta > 0$ we can find 'N' depending on ε and δ such that

$$P\left(\left|\frac{X}{n}-p\right|<\varepsilon\right)>1-\frac{pq}{n\varepsilon^2} \quad \text{where } q = 1 - p.$$

Proof : Let as associate with trials 1,2,3, ..., n random variables $X_1, X_2, ..., X_n$ as

$$X_i = \begin{cases} p, & x_i = 1 \\ q, & x_i = 0 \end{cases}$$

$$\text{So that } X = X_1 + X_2 + ... + X_n$$

Since trials are independent, variables are independent.

$E(X_i) = 0.q + 1.p = p,$

$Var(X_i) = E[(X_i - p)^2]$

$= (0-p)^2 q + (1-p)^2 p = p^2 q + q^2 p = pq(p+q) = pq.$

$Var(X) = Var(X_1 + X_2 + ... + X_n) = Var(X_1) + Var(X_2) + ... + Var(X_n) = npq.$

$E(X) = E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n) = np.$

$$E\left(\frac{X}{n}\right) = p, \quad Var\left(\frac{X}{n}\right) = \frac{1}{n^2} Var(X) = \frac{pq}{n}$$

By Chebyshev's inequality,

$$P\left(\left|\frac{x}{n}-\mu\right|<\varepsilon\right)\geq 1-\frac{pq}{n\varepsilon^2} \Rightarrow P\left(\left|\frac{X}{n}-\mu\right|<\varepsilon\right)\geq 1-\frac{pq}{n\varepsilon^2} \text{ since } \mu = p$$

Also $P\left(\left|\frac{x}{n}-p\right|\geq\varepsilon\right)\leq \frac{pq}{n\varepsilon^2}$.

Taking limit as $n \to \infty$ we get

$$\lim_{n\to\infty} P\left(\left|\frac{x}{n}-p\right|\geq\varepsilon\right)=0$$

which is the law of large number for Bernoulli's trial.

Central limit theorem :

Let $X_1, X_2, ..., X_n$ be independent random variables that are identically distributed and have finite mean μ and variance σ^2.

$$\text{Let } S_n = X_1 + X_2 + ... + X_n \qquad (n = 1, 2, 3, ...)$$

Then

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{u^2}{2}}du$$

The word identically distributed means all have the same density function i.e the random variable $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ which is standardized variable coresponding to S_n is asymptotically normal.

$$E(S_n)=E(X_1+X_2+\ldots\ldots+X_n)=E(X_1)+E(X_2)+\ldots\ldots+E(X_n)=n\mu.$$
$$Var(S_n)=Var(X_1)+Var(X_2)+\ldots..+Var(X_n)=n\sigma^2$$

The standardized normal variable corresponding to S_n is

$$S_n^* = \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}}$$

The m.g.f for S_n^* is

$$E\left(e^{tS_n^*}\right) = E\left[e^{t\left(\frac{S_n-n\mu}{\sigma\sqrt{n}}\right)}\right] = E\left[e^{t\left(\frac{X_1-\mu}{\sigma\sqrt{n}}\right)}.e^{t\left(\frac{X_2-\mu}{\sigma\sqrt{n}}\right)}\ldots\ldots.e^{t\left(\frac{X_n-\mu}{\sigma\sqrt{n}}\right)}\right]$$

$$= E\left[e^{t\left(\frac{X_1-\mu}{\sigma\sqrt{n}}\right)}\right].E\left[e^{t\left(\frac{X_2-\mu}{\sigma\sqrt{n}}\right)}\right].\ \ldots\ldots\ldots\ E\left[e^{t\left(\frac{X_n-\mu}{\sigma\sqrt{n}}\right)}\right]$$

$$= \left[E\left[e^{t\left(\frac{X_1-\mu}{\sigma\sqrt{n}}\right)}\right]\right]^n$$

since X_i s are independent and identically distributed.

Now by expantion of e^x we have,

$$E\left[e^{t\left(\frac{X_1-\mu}{\sigma\sqrt{n}}\right)}\right] = E\left[1 + t\left(\frac{X_1-\mu}{\sigma\sqrt{n}}\right) + \frac{1}{2!}t^2\left(\frac{X_1-\mu}{\sigma\sqrt{n}}\right)^2 + \frac{1}{3!}t^3\left(\frac{X_1-\mu}{\sigma\sqrt{n}}\right)^3 + \ldots.\right]$$

$$= E(1) + \frac{t}{\sigma\sqrt{n}}E\left(X_1 - \mu\right) + \frac{t^2}{2\sigma^2 n}E\left(X_1-\mu\right)^2 + \ldots.$$

$$= 1 + \frac{t}{\sigma\sqrt{n}}.0 + \frac{t^2}{2\sigma^2 n}\sigma^2 + \ldots. = 1 + \frac{t^2}{2n} + \ldots.$$

Thus

$$E\left(e^{tS_n^*}\right)=\left[1+\frac{t^2}{2n}+\ldots\right]^n. \quad \text{Limit of this as } n\to\infty \text{ is } 1.$$

Then as $\frac{S_n-n\mu}{\sigma\sqrt{n}}\sim N(0, 1)$

we have $\lim_{n\to\infty} P\left(a\le\frac{S_n-n\mu}{\sigma\sqrt{n}}\le b\right)=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{u^2}{2}}du$

Solved Examples.

Ex 1. Find first four moments

(i) about the origin

(ii) abut the mean for a random variable X having density function

$$f(x)=\begin{cases} x(9-x^2) & \text{if} \quad 0\le x\le 3 \\ 0 & \text{otherwise}\end{cases}$$

$$\mu_1'=E(x)=\frac{4}{81}\int_0^3 x^2(9-x^2)dx=\frac{4}{81}\left[9\frac{x^3}{3}-\frac{x^5}{5}\right]_0^3=\frac{4}{81}\left[9\frac{3^3}{3}-\frac{3^5}{5}\right]=\frac{8}{5}$$

$$\mu_2'=E(x^2)=\frac{4}{81}\int_0^3 x^3(9-x^2)dx=\frac{4}{81}\left[9\frac{x^4}{4}-\frac{x^6}{6}\right]_0^3=\frac{4}{81}\left[9\frac{3^4}{4}-\frac{3^6}{6}\right]=3$$

$$\mu_3'=E(x^3)=\frac{4}{81}\int_0^3 x^4(9-x^2)dx=\frac{4}{81}\left[9\frac{x^5}{5}-\frac{x^7}{7}\right]_0^3=\frac{4}{81}\left[9\frac{3^5}{5}-\frac{3^7}{7}\right]=\frac{216}{35}$$

$$\mu_4'=E(x^4)=\frac{4}{81}\int_0^3 x^5(9-x^2)dx=\frac{4}{81}\left[9\frac{x^6}{6}-\frac{x^8}{8}\right]_0^3=\frac{4}{81}\left[9\frac{3^6}{6}-\frac{3^8}{8}\right]=\frac{27}{2}$$

Now $\mu_1=0$

$$\mu_2=\mu_2'-\mu_1'^2=3-\left(\frac{8}{5}\right)^2=\frac{11}{25}=\sigma^2$$

$$\mu_3=\mu_3'-3\mu_2'\mu_1'+2\mu_1'^3=\frac{216}{35}-3.3.\frac{8}{5}+2.\left(\frac{8}{5}\right)^3=-\frac{32}{875}$$

$$\mu_4=\mu_4'-4\mu_3'\mu_1'+6\mu_2'\mu_1'^2-3\mu_1'^4=\frac{27}{2}-4\left(\frac{216}{35}\right)\left(\frac{8}{5}\right)+6.3\left(\frac{8}{5}\right)^2-3\left(\frac{8}{5}\right)^4=\frac{3693}{8750}.$$

Ex 2. A random variable X has density function given by

$$f(x)=\begin{cases}2e^{-2x} & x\geq 0\\ 0 & x<0\end{cases}$$

Find (i) moment generating function and

(ii) first four moments about origin.

Solution : We know,

$$M(t)=E\left(e^{tx}\right)=\int_{-\infty}^{\infty}e^{tx}\,f(x)dx$$

$$M(t)=\int_{0}^{\infty}e^{tx}\,(2e^{-2x})dx=\left[\frac{2}{t-2}.e^{(t-2)x}\right]_0^{\infty}=\frac{2}{2-t}\quad\text{assuming}\quad t<2$$

Thus if $|\ t\ |<2$,

$$\frac{2}{2-t}=\frac{1}{1-\frac{t}{2}}=1+\frac{t}{2}+\frac{t^2}{2^2}+\frac{t^3}{2^3}+\frac{t^4}{2^4}+\ldots..$$

Again, $M(t)=1+\mu t+\frac{t^2}{2!}\mu_2'+\frac{t^3}{3!}\mu_3'+\frac{t^4}{4!}\mu_4'+\ldots\ldots..$

Comparing we have. $\mu=\frac{1}{2}$ $\mu_2'=\frac{1}{2}$, $\mu_3'=\frac{3}{4}$. $\mu_4'=\frac{3}{2}$

Ex 3. A random variable X can assume values 1 and -1 with probability $\frac{1}{2}$ each.

Find (i) m.g.f (ii) first four moments (iii) characteristic function

$$\text{m.g.f} = M_x(t) = E(e^{tx}) = e^{t(1)}\frac{1}{2} + e^{t(-1)}\frac{1}{2} = \frac{1}{2}(e^t + e^{-t})$$

Again $e^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \ldots\ldots\ldots$ and $e^{-t} = 1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} + \frac{t^4}{4!} - \ldots\ldots\ldots$

Then $e^t + e^{-t} = 2(1 + \frac{t^2}{2!} + \frac{t^2}{4!} + \ldots\ldots\ldots)$

$$\Rightarrow \frac{1}{2}(e^t + e^{-t}) = 1 + \frac{t^2}{2!} + \frac{t^2}{4!} + \ldots\ldots\ldots$$

Also $Mx^{(t)} = 1 + \mu t + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \frac{t^4}{4!}\mu_4' + \ldots\ldots\ldots\ldots\ldots$

Comparing we get $\mu = 0$, $\mu_2' = 1$, $\mu_3' = 0$, $\mu_4' = 1$.

Characteristic function is given by

$$E(e^{iwx}) = e^{iw(1)}\frac{1}{2} + e^{iw(-1)}\frac{1}{2}$$

So, $E(e^{iwx}) = \frac{1}{2}(e^{iw} + e^{-iw}) = \cos w$

Ex.4 X is a random variable with probability generating function p(s).

Find the probability generating function of (i) X + 1 and (ii) 2 X.

We have $P_X(s) = \sum_{k=0}^{\infty} P_k S^k$

Probability generating function of $(X+1) = \sum_{k=0}^{\infty} P_k S^{k+1} = S\sum_{k=0}^{\infty} P_k S^k = SP(s)$

Probability generating function of $2x = \sum_{k=0}^{\infty} P_k S^{2k} = \sum_{k=0}^{\infty} P_k (S^2)^k = P(S^2)$

Ex 5. Use Chebyshev's inequality determine how many tosses of a fair coin will be required so that the probability will be at least 0.9 and the relative frequency of the number of heads will be between 0.4 and 0.6.

Solution . Let X be the number of heads in n tosses. Then $\frac{X}{n}$ is the relative frequency of heads.

$$E\left(\frac{X}{n}-1\right)=\frac{1}{2}=0.5=\mu$$

By the given condition, $P\left[\left|\frac{X}{n}-0.5\right|\le 0.1\right]\ge 1-\frac{pq}{n\varepsilon^2}\ge 1-\frac{\frac{1}{2}\cdot\frac{1}{2}}{n\varepsilon^2}$

Note that $\varepsilon=0.1$. So, $P\left[\left|\frac{X}{n}-0.5\right|\le 0.1\right]\ge 1-\frac{1}{4n(0.1)^2}$

$$\therefore \quad \frac{1}{4n(0.1)^2}\le 0.1 \qquad \Rightarrow n\ge\frac{1}{4.10^{-3}}=\frac{1000}{4}=250$$

Hence the required number of tosses will be at least 250

Example 6. **A random variable X has density function given by**

$$f(x)=\begin{cases}2e^{-2x} & x\ge 0\\ 0 & x<0\end{cases}$$

Find $P[\,|X-\mu|>1\,]$. Use Chebyshev's inequality to obtain the upper bound of $P[\,|X-\mu|>1\,]$.

Solution : From example 2, $\mu=\frac{1}{2}$

$$\text{Then } P[\,|X-\mu|<1\,]=P[\,|X-\tfrac{1}{2}|<1\,].$$

$$P\left(\frac{1}{2}<X<\frac{3}{2}\right)=\int_0^{\frac{3}{2}}2e^{-2x}dx=2\left[\frac{e^{-2x}}{-2}\right]_0^{\frac{3}{2}}=-(e^{-3}-1)=1-e^{-3}$$

$$\therefore P[|X-\tfrac{1}{2}|\ge 1]=1-(1-e^{-3})=e^{-3}$$

again from example 2,

$$\sigma^2 = \mu'_2 - \mu^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

Chebyshev's inequality with $\varepsilon = 1$ gives

$$P[|X - \frac{1}{2}| \geq 1] \leq \sigma^2 = 0.25$$

Example 7. A random variable X have the density function

$$f(x) \begin{cases} \frac{4x(9-x^2)}{81} & 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Find the coefficient of skewness and kurtosis.

Solution : From example 1, we have $\sigma^2 = \frac{11}{25}$, $\mu_3 = -\frac{32}{875}$ and $\mu_4 = \frac{3693}{8750}$.

$$\text{Coefficient of skewness} = \frac{\mu_3}{\sigma^3} = -0.1253$$

$$\text{Coefficient of kurtosis} = \frac{\mu_4}{\sigma^4} = 2.172$$

Thus these is a moderate skewness to the left and it some what less peaked then normal distribution which has a kurtosis 3.

●●●

Unit 2

Distribution.

Introduction :

Theoretical or probability distributions are such distributions which are not obtained by actual observations or experiments but are mathematically deduced on certain assumptions. The importance of theoretical distribution cannot be overemphasized. They provide us data on the basis of which the results of actual observations or experiments can be assessed. In fact, where theoretical distributions are available, there is no need of having observed distributions.

Discrete Probability distribution:

A random variable that takes are a finite or countably infinite number of values is called a discrete random variable.

Let X be a discrete random variable with the possible values that it can assume be $x_1, x_2, x_3, \ldots\ldots$

Let the probabilities of these values be given by

$$P(X=x_k) = f(x_k) \quad k=1, 2, 3, \ldots\ldots\ldots\ldots$$

It is convenient to introduce the probability function, also referred to as probability distribution given by $P(X=x) = f(x)$.

$f(x)$ is known as probability mass function also.

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

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

The distribution function F for a random variable X is defined by

$F(x) = P(X \leq x)$ where x is any real number.

The distribution function F(x) has the properties :

1. F(x) is non decreasing 2. $\lim_{n \to 0} F(x) = 0$ and $\lim_{n \to \infty} F(x) = 1.$

Binomial distribution.

Suppose we have an experiment such as tossing a coin or drawing a coral from a pack of cards. Each toss or draw is called a trial . Suppose that a trial is repeated, so that we have a series of u-trials. Let us call the occurrence of an events a 'success' and its non occurrence a 'failure'. Let p be in probability of success and q be the probability of failure in a single trial, so that p+q=1. In some cases this probability will not change from one trial to the next. Such trials are said to be independent and are often called **Bernoulli trials.** The number of successes in trials may be 0,1,2, ..., r,...,n and is obviously a random variable.

Let us find the probability that the variable takes a particular value r (say). Then the probability of r success will be associated with (n-r) failures. Now the probability of r successes in a preassigned order will be $p^r(1-p)^{n-r}$ or $p^r q^{n-r}$. We are interested in any r trails being successes and since r trials can be chosen out of n trails in ${}^nC_r$ mutually exclusive ways, by theorem of total probability the chance P(r) of r successes n a series of n independent trails is given by

$P(X=r)=f(r)={}^nC_r p^r q^{n-r}$ for $r=0,1,2,3,\ldots\ldots, n$.

Thus the number of successes can take the values 0,1,2,3,..., n with corresponding probabilities

$$q^n,\ {}^nC_1 q^{n-1}p,\ {}^nC_2 q^{n-2}p^2,\ldots,\ {}^nC_r q^{n-r}p^r,\ldots\ldots,p^n$$

Since $f(r)\geq 0\ \forall\ r$ and $\sum_r f(r)=\sum_{r=0}^{n} {}^nC_r\, q^{n-r}\, p^r=(q+p)^n=1$, f(r) is a probability mass function of random variable X.

Since different terms of the expansion $(q+p)^n$ gives the different probabilities for r =0,1,2,.... the above probability distribution is known as **Binomial distribution**.

The binomial distribution contains two independent constants viz n and p. They are called parameters of the distribution.

If $p=q=\frac{1}{2}$, the binomial distribution is symmetrical and when $p\neq q$ is is a skew distribution.

Moments of the binomial distribution.

Let us take an arbitrary origin at '0' success. Hence by definition of moments we have.

$$\mu_1'=\sum_{r=0}^{n} r.\,{}^nC_r\, p^r q^{n-r}$$

$$=0.q^n+1.\,{}^nC_1 pq^{n-1}+2.\,{}^nC_2 p^2q^{n-2}+\ \ldots\ +np^n$$

$$=np[q^{n-1}+{}^{n-1}C_1 pq^{n-2}+{}^{n-1}C_2 p^2q^{n-3}+\ldots+p^{n-1}\}$$ (using properties of binomial coefficient)

$$=np(q+p)^{n-1}=np.$$ Thus mean is np.

$$\mu_2'=\sum_{r=0}^{n} r^2.\,{}^nC_r\, p^r q^{n-r}$$

$$=\sum_{r=0}^{n}[r+r(r-1)]\,{}^nC_r\, p^r q^{n-r}$$

$$=\sum_{r=0}^{n} r\ \ {}^nC_r\, p^r q^{n-r}+\sum_{r=0}^{n} r(r-1)\ {}^nC_r\, p^r q^{n-r}$$

$$=np\mu_1'+n(n-1)p^2\sum_{r=2}^{n}{}^{n-2}C_{r-2}\,p^{r-2}q^{n-r}$$

$$=np+n(n-1)p^2(p+q)^{n-2}$$

$$=np+n(n-1)p^2=np[1+np-p]=np(q+np)=npq+n^2p^2.$$

$$\mu_2=\mu_2'-\mu_1'^2=npq+n^2p^2-(np)^2=npq$$

Hence variance is npq.

Since npq< np, as q < 1 it follows that for the binomial distribution, mean > variance.

$$\mu_3' = \sum_{r=0}^{n} r^3 . {}^nC_r p^r q^{n-r}$$

$$= \sum_{r=0}^{n} r+3r(r-1)+r(r-1)(r-2]\, {}^nC_r p^r q^{n-r}$$

$$= np(q+p)^{n-1}+3n(n-1)p^2(q+p)^{n-2}+n(n-1)(n-2)p^3(q+p)^{n-3}$$

$$= np +3n(n-1)p^2+n(n-1)(n-2)p^3.$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$= np+3n(n-1)p^2+n(n-1)(n-2)p^3-3(npq+n^2p^2)np+2n^3p^3$$

$$= np[1+3np-3p+(n^2-3n+2)p^2-3npq-3n^2p^2+2n^2p^2]$$

$$= np(1+3np-3p+n^2p^2-3np^2+2p^2-3npq-3n^2p^2+2n^2p^2)$$

$$= np(1-p)(1-2p)$$

$$= npq(p+q-2p)$$

$$= npq(q-p)$$

$$\mu_4' = \sum_{r=0}^{n} r^4\ {}^nC_r p^r q^{n-r}$$

$$= \sum_{r=0}^{n} \{r + 7r(r-1) + 6r(r-1)(r-2) + r(r-1)(r-2)(r-3)\}\ {}^nC_r\ p^r q^{n-r}$$

$$= np+7n(n-1)p^2+6n(n-1)(n-2)p^3+n(n-1)(n-2)(n-3)p^4$$

$$\mu_4 = \mu_4' - 4\mu_2'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$= np[1+7(n-1)p+6(n-1)(n-2)p^2+(n-1)(n-2)(n-3p^3]-4(npq+n^2p^2)np+$$

$$6(npq+n^2p^2)n^2p^2-3n^4p^4$$

$$= 3n^2p^2q^2+npq(1-6pq) \qquad \text{(after simplification)}$$

Now $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{n^2p^2q^2}{n^3p^3q^3} = \frac{(q-p)^2}{npq}$

$$\gamma_1 = \sqrt{\beta_1} = \frac{q-p}{\sqrt{npq}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3n^2p^2q^2 + npq(1-6pq)}{n^2p^2q^2} = 3 + \frac{1-6pq}{npq}$$

$$\gamma_2 = \beta_2 - 3 = \frac{1-6pq}{npq}$$

Again $q-p = q + p - 2p = 1-2p$.

For symmetrical distribution $\beta_1 = 0 \Rightarrow q = p = \frac{1}{2}$.

Skewness is positive if $\gamma_1 > 0$ in if $1-2p > 0$.

For positive skewness $p < \frac{1}{2}$

For negative skewness $p > \frac{1}{2}$

If the number of trials increases indefinitely i.e. if $n \to \infty$ then $\beta_1 \to 0$, $\beta_2 \to 3$, $\gamma_1 \to 0$, $\gamma_2 \to 0$.

Moment generating function of binomial distribution :

If X is a binomially distributed random variable then probability mass function is given by

$$f(x) = P(X=x) = {}^nC_x p^x q^{n-x}$$

Then the moment generating function is given by

$$M(t) = E(e^{tx}) = \sum_{x=0}^{n} e^{tx} f(x) = \sum_{x=0}^{n} e^{tx}\, {}^nC_x p^x q^{n-x} = \sum_{x=0}^{n} {}^nC_x (pe^t)^x q^{n-x}$$

$$M(t) = (q+pe^t)^n \qquad (1)$$

Differentiating (1) w.r.t. 't' and putting t=0,

$$\mu_1' = np.$$

Differentiating (1) w.r.t 't' twice and putting t=0.

$$\mu_2' = n^2p^2 + npq \qquad \text{and so on.}$$

Recurrence relation for binomial probabilities

we know $b(x;n,p) = f(x) = {}^nC_x p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$

$$b(x+1,np) = f(x+1) = {}^nC_{x+1} p^{x+1} q^{n-x-1} = \frac{n!}{(x+1)!(n-x-1)!} p^{x+1} q^{n-x-1}$$

$$= \frac{n-x}{x+1} \frac{n!}{x!(n-x)!} p^x q^{n-x} . \frac{p}{q}$$

$$= \frac{n-x}{x+1}\left({}^nC_x p^x q^{n-x}\right)\frac{p}{q}$$

$$b(x+1;n,p) = \frac{n-x}{x+1}.\frac{p}{q}.f(x) \Rightarrow f(x+1) = \frac{n-x}{x+1}.\frac{p}{q}.f(x)$$

This is called recurrence relation for probabilities, since f(x+1) can be calculated if f(x), n and p are known.

Theorem 1 : The rth order moment about arbitrary point of binomial distribution is given by

$$\mu'_r = \left(p\frac{\partial}{\partial p}\right)^r (p+q)^n$$

we know, $(p+q)^n = \sum_{r=0}^{n} {}^nC_r p^r q^{n-r}$

Differentiating this relation partially w.r.t p. we get

$$\frac{\partial}{\partial p}(p+q)^n = \sum_{r=0}^{n} {}^nC_r p^{r-1} q^{n-r}.r \Rightarrow p\frac{\partial}{\partial p}(p+q)^n = \sum_{r=0}^{n} r.{}^nC_r p^r q^{n-r} = \mu'_1$$

Differentiating again w.r.t p,

$$\frac{\partial}{\partial p}\left[p\frac{\partial}{\partial p}(p+q)^n\right] = \sum_{r=0}^{n} r^2\, {}^nC_r p^r q^{n-r} \Rightarrow \left(p\frac{\partial}{\partial p}\right)^2 (p+q)^n = \mu'_2$$

Thus the result is established for $r = 1,2$

using method of induction we get. $\left(p\frac{\partial}{\partial p}\right)^r (p+q)^n = \mu'_r$

Theorem 2. For binomial distribution the recurrence relation for central moment is

$$\mu_{r+1} = pq\left(nr\mu_{r-1} + \frac{d\mu_r}{dp}\right)$$

we have, $\mu_r = \sum_{x=0}^{n} {}^nC_x p^x q^{n-x} (x-np)^r$

Differentiating w.r.t p,

$$\frac{d\mu_r}{dp} = \sum_{x=0}^{n} {}^nC_x\, xp^{x-1}q^{n-x}(x-np)^r - \sum_{x=0}^{n} {}^nC_x\, p(n-x)q^{n-x-1}(x-np)^r$$

$$- \sum_{x=0}^{n} {}^nC_x\, p^x q^{n-x}\, nr\,(x-np)^{r-1}$$

$$= \sum_{x=0}^{n} {}^nC_x\, p^{x-1}q^{n-x-1}(x-np)^r[xq-np+xp] - nr\,\mu_{r-1}$$

$$\Rightarrow pq\frac{d\mu_r}{dp} = \sum_{x=0}^{n} {}^nC_x\, p^x q^{n-x}(x-np)^{r+1} - (nr\mu_{r-1})pq$$

$$= \mu_{r+1} - npq^r\mu_{r-1}$$

$$\Rightarrow \mu_{r+1} = pq\left(\frac{d\mu_r}{dp} + nr\mu_{r-1}\right)$$

Deductions.

1. Putting r = 1, $\mu_2 = pq\left(\frac{d\mu_1}{dp} + n\mu_0\right)$ since $\mu_1 = 0$ and $\mu_0 = 0$ we have $\mu_2 = npq$

2. Putting r = 2, $\mu_3 = pq\left(\frac{d\mu_2}{dp} + n^2.\mu_1\right) = pq(nq - np + 0) = npq(q-p)$

3. Putting r = 3, $\mu_4 = pq\left(\frac{d\mu_3}{dp} + n.3.\mu_2\right) = pq[n(6p^2-6p+1)+3n\,np(1-p)] = npq[1-6p(1-p)+3n\,pq]$

Theorem 3. Most probable number of successes in a series of n independent trials, the probability of successes in each trial being p.

If X=x be the most probable number of success then $P(X=x-1) \leq P(X=x) \geq P(X=x+1)$

$$\Rightarrow {}^nC_{x-1}p^{x-1}q^{n-x+1} \leq {}^nC_x p^x q^{n-x} \geq {}^nC_{x+1}p^{x+1}q^{n-x-1}$$

$$\Rightarrow \frac{n!}{(x-1)!(n-x+1)!}p^{x-1}q^{n-x+1} \leq \frac{n!}{x!(n-x)!}p^x q^{n-x} \geq \frac{n!}{(x-1-1)!(x+1)!}p^{x+1}q^{n-x-1}$$

$$\Rightarrow \frac{x}{n-x+1}\cdot\frac{q}{p} \leq 1 \geq \frac{n-x}{x+1}\cdot\frac{p}{q}.$$

That gives $xq \leq np-xp+p$ and $xq+q \geq np-xp$. ie $x \leq np+p$ and $x \geq np-q$

Hence $np-q \leq x \leq np+p$

$\Rightarrow (n+1)p-1 \leq x \leq (n+1)p.$

Case 1 . When $(n+1)p$ is an integer. Since x is an integer there will be two most probable number of successes.

Case 2 . When $(n+1)p$ is not an integer then let $(n+1)p=m+\lambda$ where m is integer and λ is proper fraction. then $m-1+\lambda \leq x \leq m+\lambda$.

This suggests that there is a single most probable values.

The most probable number of successes gives the mode of the binomial distribution. If $(n+1)p-1$ is an integer then the distribution is bimodal, otherwise it has a single mode.

Example 1 . Find the probability that in a family of 4 children there will be (a) at least one boy (b) at least one boy and at least one Girl.

Let us assume that the probability of male birth is $\frac{1}{2}$

$$P(1\text{boy}) = {}^4C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 = \frac{1}{4} \qquad P(2\text{ boys}) = {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{3}{8}$$

$$P(3\text{boys}) = {}^4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 = \frac{1}{4} \qquad P(4\text{boys}) = \left(\frac{1}{2}\right)^4 = \frac{1}{16}.$$

Thus P(at least one boy) = P(1 boy) + P(2boys) + P(3 boys) + P(4 boys)

$$= \frac{1}{4} + \frac{3}{8} + \frac{1}{4} + \frac{1}{16} = \frac{15}{16}$$

P(at least one boy and at least one girl) = P(at least one boy). P(at least one girl)

$$= \frac{15}{16} \cdot \frac{15}{16}$$

$$= 0.8789$$

Example 2. The mean of a binomial distribution is 5 and s.d is 3. Verify whether the given statement is true.

From given condition np =5 and npq=9

Hence $q = \frac{9}{5} = 1.8$

But q cannot be greater than 1. Hence the statement is wrong.

Example 3. Reliability of a missile is reported to be 0.9. Assuming the test firings to be independent obtain the probability that in 4 test firings (a) 3 or more are failures (b) not more than one is failure.

The number of failures in 4- test - firing has binomial distribution with parameters u=4 and p=1-9=.1

$P(X=4)=(0.1)^4=0.0001$

$P(X=3)={}^4C_3(0.1)^3(0.9)=0.0036$

$P(X=2)={}^4C_2(0.1)^2(0.9)^2=0.048$

$P(X=1)={}^4C^1(0.1)(0.9)^3=0.2916.$

$P(X=0)=(0.9)^4=0.6561$

Hence

(i) Probability of 3 or more failures=$P(X \geq 3)$

$=P(X=3) + P(X=4) =0.0037$

(b) Probability of not more than one failures =$P(X \leq 1)$

$=P(X=1) + P(X=0) =0.9477.$

Example 4 .Two dice are thrown until a seven is obtained. Find the most probable number of throws.

Probability of obtaining a total of seven is $\frac{6}{36} = \frac{1}{6}$.

Let X be the random variable denoting the number of throws required.

Then X may take any values 0,1,2,3 , ...

$$P(X=x)=\left(\frac{5}{6}\right)^{x-1}=\frac{1}{6}\ x=0,1,2,3,......$$

$$E(X)=\frac{1}{6}\sum_{x=0}^{\infty}x\left(\frac{5}{6}\right)^{x-1}=\frac{1}{6}\left[1+2\left(\frac{5}{6}\right)^{1}+3\left(\frac{5}{6}\right)^{2}+.........\right]=\frac{1}{6}\left(1-\frac{5}{6}\right)^{-2}=\frac{1}{6}\times 6^{2}=6.$$

If $X=x$ be the most probable value, $P(X=x)$ will be maximum when x-1 is minimum.

$\therefore$ $x-1=0 \Rightarrow x=1$ as the most probable number.

Example 5 .X and Y are two independent binomial variates with parameters $\left(8,\frac{1}{2}\right)$ and $\left(7,\frac{1}{2}\right)$ respectively. Show that X+Y will also a binomial variate with parameter $\left(15,\frac{1}{2}\right)$. Hence find the probability that X+Y=2

$$M_{X+Y}(t)=\sum_{x=0}^{8}\sum_{y=0}^{7}e^{tx}e^{ty}\,{}^{8}C_{x}\left(\frac{1}{2}\right)^{8}\,{}^{7}C_{y}\left(\frac{1}{2}\right)^{7}$$

$$=\sum_{x=0}^{8}{}^{8}C_{x}\left(\frac{1}{2}e^{t}\right)^{x}\cdot\left(\frac{1}{2}\right)^{8-x}\sum_{y=0}^{7}{}^{7}C_{y}\left(\frac{1}{2}e^{t}\right)^{y}\left(\frac{1}{2}\right)^{7-y}.$$

$$=\left(\frac{1}{2}+\frac{1}{2}e^{t}\right)^{8}\cdot\left(\frac{1}{2}+\frac{1}{2}e^{t}\right)^{7}=\left(\frac{1}{2}+\frac{1}{2}e^{t}\right)^{15}$$

Which is the m.g.f of a binomial variate with parameters $\left(15,\frac{1}{2}\right)$. Hence X+Y follows a binomial distribution with parameters $\left(15,\frac{1}{2}\right)$. $P(X+Y=Z)={}^{15}C_{2}\left(\frac{1}{2}\right)^{15}=105\times\left(\frac{1}{2}\right)^{15}.$

Poisson Distribution.

Let us consider the binomial distribution.$P(X=x)={}^{n}C_{x}p^{x}q^{n-x}$ enables us to calculate the value of the probability of x successes in n Bernoulli trials, with probability of success p in each trial. For small values of n we can calculate the probability easily. But when n becomes large enough, though it can be calculated, the calculation would be tedious and time consuming. Of course tables of binomial probabilities are available which give probabilities for certain values of n and p. But no table will give the probabilities for all possible values of n and p. Moreover we want to study the limiting behaviour of binomial distribution.

The binomial distribution exhibits an interesting limiting behaviour for large n and small p where np is constant.

This limiting form of Binomial distribution will be known as Poisson distribution,

Derivation of Poisson distribution.

In binomial distribution, the probability of x successes is given by ${}^{n}C_{x}p^{x}q^{n-x}$.

Let us consider the limiting case when p is very small and n is large enough so that the average number of successes np is a finite constant m.

Now $P(x)={}^{n}C_{x}p^{x}q^{n-x}$.

$$=\frac{n!}{x!(n-x)!}\left(\frac{m}{n}\right)^{x}\left(1-\frac{m}{n}\right)^{n-x} \qquad \text{since } p=\frac{m}{n}.$$

$$=\frac{m^{x}}{x!}\left(1-\frac{m}{n}\right)^{n}\frac{\left(1-\frac{m}{n}\right)^{-x}}{n^{x}}\cdot\frac{n!}{(n-x)!}. \qquad \text{we shall use Stirling's formula for } n!.$$

The formula is

$$\lim_{n\to\infty}(n!)\cong\sqrt{2\pi}\;e^{-n}n^{n+\frac{1}{2}}$$

Now taking limit of P(x) as $n\to\infty$ we get.

$$\lim_{n\to\infty}P(x)=\frac{m^{x}}{x!}e^{-m}\lim_{n\to\infty}\left[\frac{1}{n^{x}}\cdot\frac{m!}{(n-x)!}\right]$$

$$=\frac{m^{x}e^{-m}}{x!}\lim_{n\to\infty}\frac{\sqrt{2\pi}\;e^{-n}\,n^{\frac{n+1}{2}}}{\sqrt{2\pi}\;e^{-n+x}(n-x)^{\frac{n-x+1}{2}}}\cdot\frac{1}{n^{x}}$$

$$=\frac{m^{x}e^{-m}}{x!\,e^{x}}\cdot\lim_{n\to\infty}\left\{\frac{1}{\left(1-\frac{x}{n}\right)^{n}\left(1-\frac{x}{n}\right)^{\frac{-x+1}{2}}}\right\}$$

$$= \frac{m^x e^{-m}}{x!} \quad \text{since } e^x . e^{-x} = 1, x \text{ being finite}$$

The chance for 0,1, 2, 3, ... successes are e^{-m}, me^{-m},

$$\frac{m^2 e^{-m}}{2!} \text{ .... respectively.}$$

In other words the random variate assumes the values 0,1,2, ... with corresponding probabilities e^{-m}, me^{-m}, $\frac{m^2 e^{-m}}{2!}$, ..., $\frac{m^x e^{-m}}{x!}$, ...

The probability distribution of the number of successes is called Poisson distribution.

The sum of the probabilities = $\sum_{x=0}^{\infty} \frac{m^x e^{-m}}{x!}$

$$= e^{-m} \sum_{x=0}^{\infty} \frac{m^x}{x!}$$

$$= e^{-m}.e^{m}$$

$$= 1$$

Thus we have verified that the limiting function

$$P(X=x) \rightarrow \frac{m^x e^{-m}}{x!} \text{ of}$$

$P(X=x) = {}^nC_x p^x q^{n-x}$ which is a probability function.

The Poisson distribution involves only one parameter m.

Definition : A random variable X, taking a set of values 0,1,2,3 ----- is said to have Poisson distribution with parameter m if for m (>0)

$$P(X=x) = \begin{cases} \frac{e^{-m} m^x}{x!}, & x = 0, 1, 2, ... \\ 0 & \text{otherwise} \end{cases}$$

The random variable X having Poisson distribution is known as Poisson random variable.

The above expression gives the probability of x occurrence of an event, the probability of one occurrence is small. An event with a small probability of occurrence is a rare event. Hence Poisson distribution is also known as distribution of rare events.

Applications of Poisson distribution :

A Poisson distribution is a widely used model, particularly to explain the probabilistic behaviour of events with small probability of occurrence. The Poisson model is good for the number of occurrence of rare events. The following are the some examples.

1. The number of accidents occurring in large factory during week.
2. The number typographical errors per page in typed materials.
3. The number of bacteria observed under a microscope in small volume of liquid.
4. The number of monthly deaths due to a rare disease like AIDS in a city.
5. The number defective screws per box of 100 screws. In each of the above examples, the probability of success is small and n is large.

Moments of Poisson distribution :

Let X be a Poisson variate with probability

$$P(X=x)=\frac{m^x e^{-m}}{x!} \qquad x = 0, 1, 2, 3, ...$$

$$\mu_1' = \sum_{x=0}^{\infty} x.\frac{m^x e^{-m}}{x!} = 0.e^{-m} + 1.\frac{m.e^{-m}}{1!} + 2.\frac{m^2 e^{-m}}{2!} + 3.\frac{m^3 e^{-m}}{3!} + ...$$

$$= m\,e^{-m}\left[1+\frac{m}{1!}+\frac{m^2}{2!}+........\right]$$

$$= m\,e^{-m}.e^{m} = m. \qquad \text{Hence mean} = m.$$

$$\mu_2' = \sum_{x=0}^{\infty} x^2\frac{m^x e^{-m}}{x!} = \sum_{x=0}^{\infty}\left[\{x+x(x-1)\}\frac{m^x e^{-m}}{2!}\right]$$

$$= \sum_{x=0}^{\infty}\frac{xm^x e^{-m}}{x!} + \sum_{x=0}^{\infty} m^2\frac{m^{x-2} e^{-m}}{(x-2)!}$$

$$= m + m^2\left\{\left(1+m+\frac{m^2}{2!}+-------\right).e^{-m}\right\}$$

$$= m + m^2.e^{-m}.e^{m} = m + m^2$$

$\mu_2 = \mu_2' - \mu_1' = m + m^2 - m^2 = m$. So, Variance = m.

Thus in case of Poisson distribution, mean = variance.

$$\mu_3' = \sum_{x=0}^{\infty} x^3 \frac{m^x e^{-m}}{x!} = \sum_0^{\infty} [x+3x(x-1)+x(x-1)] \frac{m^x e^{-m}}{x!}$$

$$= \sum_0^{\infty} \frac{m^x e^{-m}}{(x-1)!} + 3\sum_0^{\infty} \frac{m^x e^{-m}}{(x-2)!} + \sum_0^{\infty} \frac{m^x e^{-m}}{(x-3)!}$$

$$= me^{-m}.e^{m} + 3m^2 e^{-m} \sum_2^{\infty} \frac{m^{x-2}}{(x-2)!} + m^3 e^{-m} \sum_0^{\infty} \frac{m^{x-3}}{(x-3)!}$$

$$= m.1 + 3m^2 e^{-m} e^{m} + m^3 e^{-m}. e^{m}$$

$$= m + 3m^2 + m^3$$

$$\mu_3 = \mu_3' - 3\mu_2' \mu_1' + 2\mu_1'^3$$

$$= (m + 3m^2 + m^3) - 3(m + m^2)m + 2m^3.$$

$$= m + 3m^2 + m^3 - 3m^2 - 3m^3 + 2m^3 = m.$$

$$\mu_4' = \sum_0^{\infty} x^4 \frac{m^x e^{-m}}{x!} = \sum_0^{\infty} [x+7x(x-1)+6x(x-1)(x-2)+x(x-1)(x-2)(x-3)] \frac{m^x e^{-m}}{x!}$$

$$= \sum_0^{\infty} \frac{m^x e^{-m}}{(x-1)!} + 7\sum_0^{\infty} \frac{m^x e^{-m}}{(x-2)!} + 6\sum_0^{\infty} \frac{m^x e^{-m}}{(x-3)!} + \sum_0^{\infty} \frac{m^x e^{-m}}{(x-4)!}$$

$$= m + 7m^2 + 6m^3 + m^4$$

$$= m(m^3 + 6m^2 + 7m + 1)$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$= m(m^3+6m^2+7m+1) - 4(m+3m^2+m^3)m + 6(m+m^2)m^2 - 3m^4$$

$$= m\,[m^3 + 6m^2 + 7m + 1 - 4m + 12m^2 - 4m^3 + 6m^2 + 6m^3) - 3m^3]$$

$$= m(3m+1)$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{m^2}{m^3} = \frac{1}{m},$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{m(3m+1)}{m^2} = \frac{3m+1}{m} = 3 + \frac{1}{m}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{m}},$$

$$\gamma_2 = \beta_2 - 3 = 3 + \frac{1}{m} - 3 = \frac{1}{m}.$$

Thus when $m \to \infty$ both γ_1 and $\gamma_2 \to 0$. Again when $m \to \infty$ $\beta_1 \to 0$ and $\beta_2 \to 3$.

Coefficient of skewness is $\frac{1}{\sqrt{m}}$. coefficient of kurtosis is $3+\frac{1}{m}$.

Recurrence relation:

1. Let $f(x) = P(X=x) = \frac{m^x e^{-m}}{x!}$ $\quad x = 0, 1, 2, 3, \ldots$

$$\text{Then } f(x) = \frac{m}{x} f(x-1)$$

Proof: Since $f(x) = \frac{m^x e^{-m}}{x!} = \frac{m}{x}\left(\frac{m^{x-1}e^{-m}}{(x-1)!}\right) = \frac{m}{x} . f(x-1)$

$$\text{Thus } f(x) = \frac{m}{x} f(x-1).$$

2. $\mu_{r+1} = m\left[r\mu_{r-1} + \frac{d\mu_r}{dm}\right]$

Proof: Since $P(X=x) = \frac{m^x e^{-m}}{x!}$ $\quad x = 0, 1, 2, 3, \ldots$

$$\mu_r = \sum_{x=0}^{\infty} (x-m)^r \frac{m^x e^{-m}}{x!}$$

$$\frac{d\mu_r}{dm} = \sum_{x=0}^{\infty} -r(x-m)^{r-1} \frac{m^x e^{-m}}{x!} + \sum_{x=0}^{\infty} (x-m)^r x \frac{m^{x-1} e^{-m}}{x!} \sum_{x=0}^{\infty} (x-m)^r \frac{m^x e^{-m}}{x!}$$

$$= -r\mu_{r-1} + \sum_{x=0}^{\infty} (x-m)^r \frac{m^x e^{-m}}{x!}\left(\frac{x}{m} - 1\right)$$

$$= -r\mu_{r-1} + \frac{1}{m} \sum_{x=0}^{\infty} (x-m)^{r+1} \frac{m^x e^{-m}}{x!}$$

$$= -r\mu_{r-1} + \frac{1}{m} \mu_{r+1}$$

$$\Rightarrow \mu_{r+1} = m\left[r\mu_{r-1} + \frac{d\mu_r}{dm}\right]$$

Deduction : we know $\mu_0=1$ and $\mu_1=0$

Putting r =1, $\mu_2 = m\,[0 + 1.\mu_0] = m$

Putting r =2, $\mu_3 = m\,[1 + 2.\mu_1] = m$

Putting r =3, $\mu_4 = m\,[1 + 3.\mu_2] = m(1+3m)$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{1}{m} \text{ and } \beta_2 = \frac{\mu 4}{\mu_2^2} = 3 + \frac{1}{m}.$$

Moment generating function.

$$M_{x=0}(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} p(x)$$

$$= \sum_{x=0}^{\infty} e^{tx} . \frac{m^x e^{-m}}{x!}$$

$$= e^{-m} \sum_{x=0}^{\infty} e^{tx} \frac{(me^t)^x}{x!}$$

$$= e^{-m} e^{me} = e^{m(e^t - 1)}$$

Theorem : It two independent random variables X_1 and X_2 have Poisson distributions with means m_1 and m_2 respectively then their sum X_1+X_2 is a Poisson variable with mean m_1+m_2

Let $M_1(t)$ and $M_2(t)$ be the m.g.f of x_1 and x_2 and let $M(t)$ be the m.g.f of their sum.

$M_1(t)=\exp[m_1(e^t-1)]$,

$M_2(t)=\exp[m_2(e^t-1)]$

Now, m.g.f of their sum X_1+X_2 is given by

$M(t) = M_1(t) \times M_2(t) = \exp[m_1(e^t-1)].\exp[m_2(e^t-1)] = \exp[(m_1+m_2)(e^t-1)$

This is the m.g.f of Poisson distribution with mean m_1+m_2. Hence the result.

Mode of Poisson distribution .

This mode is that value of x for which $\frac{e^{-m} m^x}{x!}$ is greater than the term that precedes it and the term that follows it.

Hence

$$\frac{m^{x-1}e^{-m}}{(x-1)!} \leq \frac{m^{x}e^{-m}}{x!} \leq \frac{m^{x+1}e^{-m}}{(x+1)!} \Rightarrow 1 \leq \frac{m}{x} \leq \frac{m^2}{(x+1)(x)}$$

So, $x \leq m$ and $x+1 \geq m$.

Thus $m-1 \leq x \leq m$.

Thus if m is an integer, there will be two modal values (m-1) and m.

If m is not an integer these will be a single mode at x=[m], the greater integer contained in m.

Illustrative examples.

Example 1. If the probability that an individual will suffer a bad reaction from injective of a given serum is 0.001, determine the probability that out of 2000 individuals (a) exactly 3 (b) more than 2 individual will suffer a bad reaction.

Let X denote the number of individual suffering bad reaction.

X is Bernoulli distributed but since bad reactions are rare events, we can suppose that X is Poisson distributed.

$$P(X=x) = \frac{m^{x}e^{-m}}{x!} \text{ where } m = (2000)(0.001) = 2$$

(a) $P(X=3) = \frac{2^3 e^{-2}}{3!} = 0.18$

(b) $P(X>2) = 1-P(X=0)-P(X=1)-P(X=2) = 1 - \frac{2^0 e^{-2}}{0!} - \frac{2e^{-2}}{1!} - \frac{2^2 e^{-2}}{2!}$

$$= 1 - e^{-2}(1+2+2) = 1 - 5e^{-2}$$

Example 2. If the number of accidents occurring in an industrial plant during a day is given by Poisson random variable with parameter 3, find (a) probability that no accident occurs in a day (b) the expected number of accident per day

$$P(X=x) = \frac{e^{-3}3^x}{x!} \quad x = 0, 1, 2, \ldots\ldots\ldots$$

(a) $P(X=0)\ e^{-3} = 0.05$ (b) $E(X)=3$ and var $(X)=3$.

Example 3. Find the probability that at must 5 defective fuses will be found in a box of 200 fuses if experience shows that 2 percent of such bases are defective.

$$m=np=200(.02)=4$$

$$=\sum_{0}^{5} \frac{e^{-4}4^x}{x!}$$

$$P(X \leq 5) = e^{-4}\left(1+4+\frac{4^2}{2!}+\frac{4^3}{3!}+\frac{4^4}{4!}+\frac{4^5}{5!}\right)$$

$$=0.785$$

Example 4. A car hire company has two luxury buses, which it hives out day by day. The number of demands for such a bus an each day is distributed as a Poisson distribution with mean 1.2. Calculate the proportion of days on which some demands are refused.

Let X be random variable denoting number of demands on each day.

Required probability $=P(X>2)=1-P(X\leq 2)=1-\sum_{0}^{2} e^{-1.2}\frac{(1.2)}{x!}$

$$=1-\left[e^{-1.2}\left\{1+\frac{1.2}{1!}+\frac{(1.2)^2}{2!}\right\}\right]$$

$$=1-e^{-1.2}(1+1.2+0.72)$$

Example 5. Screws are packed in boxes containing 300 screws each. On an average 1% of the screws are defective. What is the probability that a box contain 3 or more detectives?

$$np = 300\times(.01)=3.$$

Required probability $= P(X\geq 3)=1-P(X\leq 3)$

$$=1-e^{-3}\left(1+3+\frac{3^2}{2!}\right)$$

$$=1-\frac{17}{2}.e^{-3}$$

Hypergeometric Distribution

Concept of Hypergeometric variable :

We shall explain through example.

Let a box contains 3 red and 5 black balls- one ball is drawn at random. Probability of getting a red ball is $\frac{3}{8}$. If we want to make a second draw, there will be two ways (i) replacing the ball drawn (ii) without replacing the ball drawn.

Let us consider the first case. Since the number of ball remains the same, the result of the first draw has no effect on the second draw and probability of drawing a red ball will be same $\frac{3}{8}$. If 4 such drawings with replacement are made and X is the number of red balls drawn then X will be a binomial random variable. Hence

$$P(X=k)={}^{n}C_{k}p^{k}q^{n-k} \quad \text{where } k=0, 1, 2, 3, 4 \text{ and } n=4.$$

Let us now consider the second case, where drawn ball is not replaced. Definitely it will effect the successive drawn and hence trials are not independent. For second draw there will 7 ball, for 3rd there will 6 balls and so on. After 8 drawn the box will be empty. As it is not a Bernoulli trials, the random variable will be different from binomial. Such random variables will be called hypergeometric random variable.

If two drawings are made, we may get 2 red balls, 1 red and 1 black balls , 2 black balls. If X is the number of red balls then X can take values 2,1, or 0.

Let us take the case X=1, then out of the two ball 1 is red and the other is black, Two balls out of (3+5)=8 can be drawn in ${}^{8}C_{2}$ ways; 1 red ball can be drawn in ${}^{3}C_{1}$ ways, and 1 black ball can be drawn in ${}^{5}C_{1}$ ways. Each way to drawing red and black ball can be combined and the number of ways of getting 1 red, 1 black ball will be ${}^{3}C_{1}.{}^{5}C_{1}$.

$$P(X=1)=\frac{{}^{3}C_{1}.\,{}^{5}C_{1}}{{}^{8}C_{2}}$$

Let us now generalise the problem.

Suppose that r balls are drawn are at a fine without replacement from a box containing m red and n black balls. Let X be the number of red balls drawn . Then X=x implies that of r drawn x are red and r-k are black ball. We have

$$P(X=x)=\begin{cases}\dfrac{^{m}C_{x}\ ^{n}C_{r-x}}{^{m+n}C_{r}}; & x=0,1,2,\ldots r;\quad r\le m,\ r\le n\\ 0 & \text{otherwise}\end{cases}$$

where $\sum_{x=0}^{r} P(X=x)=1$ since $\sum_{x=0}^{r}\ ^{m}C_{x}.\ ^{n}C_{r-x}=\ ^{m+n}C_{r}.$

If r<n then X assumes values from x=0 to x=min (r,m)

If r>n and Z=r-n then X assumes values from x=r(>0) to x= min (r,m)

The above distribution is known as hypergeometric distribution with three parameters m,n,r.

Mean and variance

$$E(x)=\sum_{x=0}^{r} x\ P(X=x)=\sum_{x=0}^{r} x\ \frac{^{m}C_{x}\ ^{n}C_{r-x}}{^{m+n}C_{r}}$$

$$=\ ^{m+n}C_{r}\sum_{x=1}^{r}\ ^{m-1}C_{x-1}\ ^{n}C_{r-x} \qquad \text{since } ^{m}C_{x}=\frac{m}{x}.\ ^{m-1}C_{x-1}$$

$$=\frac{m}{^{m+n}C_{r}}\sum_{y=0}^{r-1}\ ^{m-1}C_{y}\ ^{n}C_{r-y-1}$$

$$=\frac{m}{^{m+n}C_{r}},\ ^{m-1+n}C_{r-1}=m\ \frac{|\underline{m+m-1}}{|\underline{r-1}\ |\underline{m+n-1}}.\ \frac{|\underline{m+n-1}\ |\underline{r}.}{|\underline{m+n}}=\frac{mr}{m+n}$$

So, $\mu_1'=\text{mean}=\dfrac{mr}{m+n}$

$$\mu_2'=E(x^2)=\sum_{x=0}^{r} x^2\ P(x)=\sum_{x=0}^{r} x(x-1)\ P(x)+\sum_{x=0}^{r} x\ P(x)$$

$$= \sum_{x=0}^{r} x(x-1)\, P(x) = \frac{\sum_{x=o}^{r} x(x-1)\ {}^{m}C_x\ {}^{n}C_{r-x}}{{}^{m+n}C_r}$$

$$= \frac{m(m-1)}{{}^{m+n}C_r} \sum_{x=2}^{r} {}^{m-2}C_{x-2}\ {}^{n}C_{r-x}$$

$$= \frac{m(m-1)}{{}^{m+n}C_r} \sum_{y=0}^{r} {}^{m-2}C_y\ {}^{n}C_{r-2-y}$$

$$= \frac{m(m-1)}{{}^{m+n}C_r}\ {}^{m-2+n}C_{r-2}$$

$$= m(m-1) \frac{|\underline{m+n-2}}{|\underline{r-2}\ |\underline{m+n-r}} \cdot \frac{|\underline{r}\ |\underline{m+n-r}}{|\underline{m+m}} = \frac{m(m-1)\, r\, (r-1)}{(m+n)(m+m-1)}$$

$$\mu_2' = \frac{m(m-1)\, r\, (r-1)}{(m+n)(m+m-1)} + \frac{mr}{m+n}$$

$$\therefore\ \mu_2 = \frac{m(m-1)\, r\, (r-1)}{(m+n)(m+m-1)} + \frac{mr}{m+n} - \left(\frac{mr}{m+n}\right)^2$$

$$= \frac{mr}{(m+n)^2} \left[\frac{(m-1)(r-1)}{m+n-1} \cdot (m+n) + (m+n) - mr \right]$$

$$= \frac{mr}{(m+n)^2}\ \frac{1}{m+n-1} \left[m^2r - mr - m^2 + m + mnr - nr + mn + n + m^2 + mn - m + mn + n^2 - n - m^2r - mnr + mr\right]$$

$$\mathrm{Var}\,(x) = \frac{mnr\,(m+n-r)}{(m+n)^2\ m+n-1}$$

Multinomial Distribution

Definition :

Let us suppose that the event $A_1, A_2, \ldots\ldots, A_k$ are mutually exclusive and can occur with respective probabilities $p_1, p_2, \ldots\ldots, p_k$ where $p_1 + p_2 + p_3 + \ldots\ldots + p_k = 1$. If $x_1, x_2, \ldots\ldots, x_k$ are random variable respectively giving the number of times that $A_1, A_2, \ldots\ldots, A_k$ occur in a total of n trials, so that

$X_1+X_2+.........+X_k = n$ then

$$P(X_1 = x_1, X_2 = x_2, ..., X_k = x_k) = \begin{cases} \dfrac{\lfloor x}{\lfloor n_1 \lfloor n_2 ---- \lfloor n_k} p_1^{n_1}.p_2^{n_2}.....p_k^{n_k} & \\ 0 & \text{otherwise} \end{cases}$$

where $n_1+n_2+........+n_k = n$ is the joint probability function to the random variables $x_1, x_2, ..., x_k$.

This is a generalisation of binomial distribution and is known as multinomial distribution, since it is the general term in the multinomial expansion of $(p_1+p_2+........+p_k)^n$.

This is not a univariate distribution but a multivariate distribution involving k variates $n_1, n_2, ..., n_k$. Out of these k variates only (k-1) are independent.

Illustrative examples

Example 1. If a fair die is thrown, the probability of getting 1, 2, 3, 4, 5 and 6 points exactly twice each is

$$P(x_1=1, x_2=2,....., x_6=6) = \frac{\lfloor 12}{(\lfloor 2)^6}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2$$

$$= \frac{1925}{559872}$$

Example 2. A box contains 5 red balls, 4 white balls, and 3 blue balls. A ball is selected at random from the box, its colour is noted and then ball is replaced. Find the probability that out of 6 balls selected 3 are red, 2 are while and 1 is blue.

$$P(\text{ red at any drawing }) = \frac{5}{12} \qquad P(\text{white at any drawing }) = \frac{4}{12}$$

$$P(\text{ blue at any drawing }) = \frac{3}{12}$$

$$\text{Then } P(3 \text{ red}, 2 \text{ white}, 1 \text{ blue}) = \frac{\lfloor 6}{\lfloor 3 \lfloor 2 \lfloor 1}\left(\frac{5}{12}\right)^2\left(\frac{4}{12}\right)^2\left(\frac{3}{12}\right)^2 = \frac{625}{5184}.$$

Example 3. A box contains 10 red, 6 black and 9 white balls. 5 balls are drawn at random without replacement. What is the probability that there are

(i) 3 red, 1 black and 1 white balls. (ii) 1 red, 3 black and 1 white balls.

It will be problem of hypergeometric distribution.

(i) Required probability $= \dfrac{{}^{10}C_3\, {}^{6}C_1\, {}^{9}C_1}{{}^{25}C_5} = \dfrac{\lfloor 10}{\lfloor 7 \lfloor 3} \cdot \dfrac{\lfloor 6}{\lfloor 1 \lfloor 5} \cdot \dfrac{\lfloor 9}{\lfloor 1 \lfloor 8} \times \dfrac{\lfloor 5}{\lfloor 25}$

$$= \frac{10.9.8}{3.2} . 6.9. \frac{1}{25.24.23.22.21 \lfloor 20} = \frac{1}{8755. \lfloor 20}$$

(ii) Required probability $= \dfrac{{}^{10}C_1\, {}^{6}C_2\, {}^{9}C_1}{{}^{25}C_5} = \dfrac{10.15.9}{{}^{25}C_5} = \dfrac{1350}{{}^{25}C_5}$

Continuous probability distribution.

We have so far dealt with discrete probability distribution. We shall now consider a few probability models for continuous variables. Such variables can take any value in an interval. Normally these are associated with measurement data. Examples are measurement of height, weight, amount of rainfall, temperature etc. The basic difference between a discrete and a continuous variable is that former involves counting and the latter involves measuring. These variables (X) can take any value in a given interval in $a \leq x \leq b$. These variables are called continuous variates and their probability distributions are known as continuous probability distribution.

The distribution of continuous random variable can be represented by a curve such that the total area between the curve and the x-axis is equal to 1.

It f(x) a continuous function of x, defines the probability distribution of a random variate X by the relation that the probability of the value of the variate falling in the interval $(x - \frac{1}{2} du)$ to $(x + \frac{1}{2} du)$ is expressible in the form f(x) dx.

$$P\left(x - \frac{du}{2} \leq x \leq x + \frac{du}{2}\right) = f(x)dx.$$

f(x) is called probability density function and f(x)dx is called probability differential. The continuous curve y=f(x) is called probability curve.

Thus the probability density function f(x) of a continuous random variable is such that

(i) $f(x) \geq 0$

(ii) total area of the region between the curve y=f(x) and the x-axis is 1.

(iii) $P(a<x<b)$= area between the curve y=f(x) and x-axis bounded by the ordinate x=a and x=b.

In general, y=f(x) is a straight-line.

Normal distribution.

So far we have discussed discrete probability distribution. Here we propose to study a very important continuous random variable and its distribution, normal distribution. As in the previous unit viz Poisson distribution here also we shall study limiting case of binomial distribution when x is very large but unlike Poisson distribution the probability of successes p is finite. The common properties of normal distribution will be studied here.

The normal distribution is by for the most used distribution for drawing inferences from statistical data because of the following reason:

1. Number of evidences are accumulated to show that normal distribution provides a good fit or describe the frequencies of occurrences of many variable and functions in the field of physical and social sciences.
2. Normal distribution is of great value in educations evaluation and educational research. Normal distribution is not an actual distribution of scores an any test of ability or academic achievement, but it is a mathematical modes.

In solving problems we take values of different probabilities from standard table which are available in any statistics book. We have not given those tables here, as it is a material for theoretical studies . But students are advised to take help of those tables

The density function of Normal distribution is given by

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & -\infty < x < \infty \\ 0 & \text{otherwise} \end{cases}$$

where μ and σ are mean and standard deviation of the distribution. The corresponding distribution function is given by

$$F(x) = P(X \le x) = f(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

The normal distribution has two parameters, the mean μ and standard deviation σ.

The actual shape of the frequency curve y=f(x) is bell-shaped. It is shown the figure below

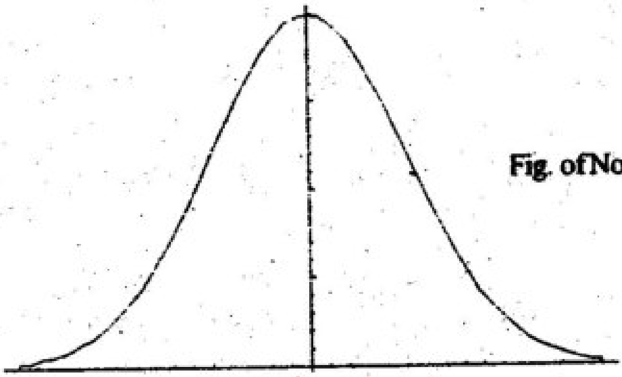


Fig. of Normal Probability Curve

The curve is symmetrical about the point x = m. The total area between the curve and x-axis from $-\infty$ to ∞ is 1. The frequency curve of normal distribution is known as normal curve.

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} .\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)dx = 1.$$

Derivation of normal probability distribution function.

The normal distribution can be regarded as the limiting form of binomial distribution when the number of trials n is very large but p is not very small.

In the binomial distribution, the probability for the variate to take the value r is

$$P(n,r) = {}^nC_r\, p^r q^{n-r} = \frac{n!}{r!(n-r)!} p^r q^{n-r}.$$

Applying Stirling's formula.

$$P_r = \lim_{n\to\infty} P(n,r) = \lim_{n\to\infty} \frac{\sqrt{2\pi}\, e^{-n}\, n^{n+\frac{1}{2}}\, p^r\, q^{n-r}}{\sqrt{2\pi}\, e^{-r}\, r^{r+\frac{1}{2}}\, \sqrt{2\pi}\, e^{-(n-r)}\, {n-r}^{n-r+\frac{1}{2}}}$$

Let $N = \left(\frac{r}{np}\right)^{r+\frac{1}{2}} \left(\frac{n-r}{nq}\right)^{n-r+\frac{1}{2}}$ and $Z = \frac{r-np}{\sqrt{npq}}$.

Then $P_r = \lim_{n\to\infty} \frac{1}{N} \cdot \frac{1}{\sqrt{2\pi npq}}$.

Now $Z = \frac{r-np}{\sqrt{npq}} \Rightarrow \frac{r}{np} = 1+2\sqrt{\frac{q}{np}}$ and $\frac{n-r}{nq} = 1-Z\sqrt{p/nq}$

Then, $\log_e N = \left(np+2\sqrt{npq}+\frac{1}{2}\right)\log_e\left(1+2\sqrt{\frac{q}{np}}\right)+\left(nq-2\sqrt{npq}+\frac{1}{2}\right)\log_e\left(1-2\sqrt{\frac{p}{nq}}\right)$

Assuming that $|Z| <$ the smaller of the two quantities $\sqrt{\frac{np}{q}}$ and $\sqrt{\frac{nq}{p}}$ and expanding we get,

$$\log_e N = \left(np+Z\sqrt{npq}+\frac{1}{2}\right)\left[Z\sqrt{\frac{q}{np}}-\frac{2q}{2np}+O(n^{-3/2})\right]$$

$$+\left(nq-Z\sqrt{npq}+\frac{1}{2}\right)\left[-Z\sqrt{\frac{p}{nq}}-\frac{2^2 p}{2nq}+O(u^{-3/2})\right]$$

$$=\frac{2}{2\sqrt{n}}\left(\sqrt{\frac{q}{p}}-\sqrt{\frac{p}{q}}\right)+\frac{z^2}{2}-\frac{z^2}{4n}\left(\frac{q}{p}+\frac{p}{q}\right)+O(n^{-3/2}).$$

Thus when $n\to\infty$, $\log_e N \to \frac{Z^2}{2}$ i.e. $N \to e^{\frac{1}{2}Z^2}$.

As r runs through integral values the increments in 2 are each equal to $(npq)^{-\frac{1}{2}}$ which we denote by dz when $n \to \infty$. Thus if dp denotes the probability for the variate Z to lie in the interval $z-\frac{1}{2}dz$ to $z+\frac{1}{2}dz$

we have $dP = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ $\qquad -\infty < z < \infty$

Replacing z by $\frac{x-np}{\sqrt{npq}}$, $\quad dp = \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-\frac{1}{2}\left(\frac{x-np}{\sqrt{npq}}\right)^2}$ $\qquad -\infty < x < \infty$

But $np = \mu$, the mean and $\sqrt{npq} = \sigma$ the s.d.

$$dp=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \qquad -\infty< x <\infty$$

$$=f(x)$$

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}.\exp\left[-\frac{1}{2}.\frac{(x-\mu)^2}{\sigma^2}\right]$$

The continuous variate X which is distributed with probability density f(x) is called normal variate with mean μ and s.d σ and denoted by $N(\mu,\sigma)$

Standard normal variable:

Let X be an $N(\mu,\sigma)$ so that $E(x)=\mu$ and var $(x)=\sigma^2$

Let us now transfer X to another random variable Z such that $Z=\frac{X-\mu}{\sigma}$. Then Z has also a normal distribution.

$$E(z)=E\left(\frac{x-\mu}{\sigma}\right)=E\left(\frac{x}{\sigma}\right)-E\left(\frac{\mu}{\sigma}\right)=\frac{E(x)}{\sigma}-\frac{\mu}{\sigma} \text{ since } \mu \text{ and } \sigma \text{ are constant}$$

$$=\frac{\mu}{\sigma}-\frac{\mu}{\sigma}=0$$

$$Var(Z)=Var\left(\frac{x}{\sigma}-\frac{\mu}{\sigma}\right)=\frac{1}{\sigma^2}\,Var(x)=\frac{\sigma^2}{\sigma^2}=1.$$

Thus Z is a N (0,1) is a normal distribution with mean 0 and s.d. 1.

$$Z=\frac{X-\mu}{\sigma}\Rightarrow X=\mu+Z\sigma$$

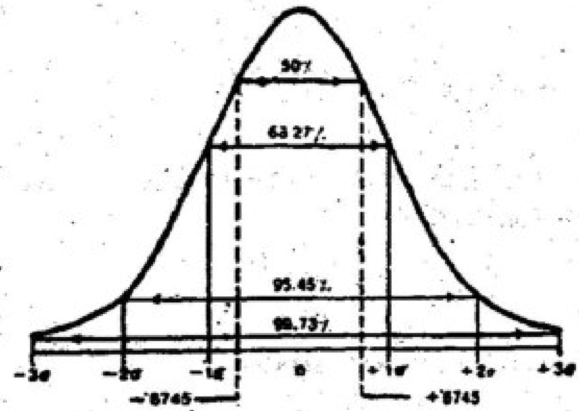
So that $P(a< x <b)=P(a<\mu+z\sigma<b)=P\left(\frac{a-\mu}{\sigma}<Z<\frac{b-\mu}{\sigma}\right)$.

Features of normal curve:

The normal curve is symmetrical about the ordinate $x=\mu$. The ordinate at $x=\mu$ divides the area under the normal curve into two equal parts. Hence the median of the distribution coincides with the mean and mode. No portion of the curve lies below x-axis since the normal probability function cannot be negative.

The normal curve changes its direction from convex to concave at a point of inflexion. If we draw perpendiculars from these two points in the x-axis these two will meet x-axis at a distance 1σ from the $x = \mu$:

Approximately 68.26% area of the curve falls within the limits of $\pm 1\sigma$ unit from the mean. The total area under the normal curve may be considered to approach 100% probability.



In the above graph we have indicated the areas within $1\sigma, 2\sigma$ and 3σ of the mean. They are respectively 68.26%, 95.45% and 99.73% of the total area.

$P(-1 \le z \le 1) = 0.6827$, $P(-z \le Z \le 2) = 0.9545$, $P(-3 < z < 3) = 0.9973$, 50% area of the curve lies to the left side of maximum central ordinate and 50% lies to the right side.

Properties of normal distribution:

1. Mean deviation from the mean . μ.

$$\text{m. d from } \mu = \int_{-\infty}^{\infty} |x-\mu| \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$Z = \frac{x-\mu}{\sigma} \text{ so that } x = \mu + 2\sigma.$$

$$\text{m.d from } \mu = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma|z| e^{-\frac{1}{2}z^2} .\sigma\, dz$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z|\, e^{-\frac{1}{2}z^2}\, dz$$

$$=\frac{\sigma}{\sqrt{2\pi}}\left[\int_{-\infty}^{0} ze^{-\frac{1}{2}z^2}\,dz+\int_{0}^{\infty} ze^{-\frac{1}{2}z^2}\,dz\right]$$

$$=\frac{2\sigma}{\sqrt{2\pi}}\int_{0}^{\infty} ze^{-\frac{1}{2}z^2}\,dz=\sigma\sqrt{2/\pi}=0.7979\sigma(\text{approx.})=\frac{4}{5}\sigma.$$

Moutments about the mean μ.

$$\mu_{2n+1}=\int_{-\infty}^{\infty}(x-\mu)^{2n+1}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx$$

$$=\frac{\sigma^{2n+1}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^{n+1}e^{-\frac{1}{2}Z^2}\,dz,\qquad Z=\frac{x-\mu}{\sigma}$$

=0 since the integral is an odd function of z

$$\therefore\ \mu_3=\mu_5=\ldots\ldots\ldots=0$$

Thus all odd moments about the mean is zero.

$$\beta_1=\frac{\mu_3^{\,2}}{\mu_2^{\,3}}=0$$

Hence in normal distribution these is no skewness.

$$\mu_{2n}=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-\mu)^{2n}\exp\left[\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

$$=\frac{\sigma^{2n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} Z^{n-1}ze^{-\frac{1}{2}z^2}\,dz,\qquad Z=\frac{x-\mu}{\sigma}$$

$$=\frac{\sigma^{2n}}{\sqrt{2\pi}}\left[Z^{u-1}e^{-\frac{1}{2z^2}}\right]_{-\infty}^{\infty}+\frac{\sigma^{2n}}{\sqrt{2\pi}}(2n-1)\int_{-\infty}^{\infty} Z^{2n-2}e^{-\frac{1}{2}z^2}\,dz$$

$$= (2n-1)\sigma^2 \frac{\sigma^{2n-2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} Z^{n-2} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$ Since Ist integral in limit is zero.

$=(2n-1)\sigma^2\mu_{2n-2}$

Putting n=0,1,2, ... and noting $\mu_0=1$ we have

$\mu_2=\sigma^2$, $\mu_4=3\sigma^2.\mu_2=3\sigma^4$ and so on.

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{\sigma^4} = 3.$$

Moment generating function.

(a) w.r.t origin :

$$M_{x=0}(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx}.\frac{1}{\sigma\sqrt{2\pi}} \text{Exp}\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{[t(\mu+\sigma z)]}.e^{\left(-\frac{z^2}{2}\right)}dz, \qquad z = \left(\frac{X-\mu}{\sigma}\right)$$

$$= e^{\mu t+\frac{1}{2}t^2\sigma^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t\sigma)^2} dt$$

$$= e^{\mu t+\frac{1}{2}t^2\sigma^2}$$

(b) w.r.t mean μ:

$$M_{x=\mu}(t) = E(e^{t(x-\mu)}) = e^{-\mu t}E(e^{tx}) = e^{\frac{1}{2}t^2\sigma^2} \text{ by (a)}$$

Now $$e^{\frac{1}{2}t^2\sigma^2} = 1 + \frac{\frac{1}{2}t^2\sigma^2}{1!} + \frac{\left(\frac{1}{2}t^2\sigma^2\right)^2}{2!} + \dots\dots\dots + \frac{\left(\frac{1}{2}t^2\sigma^2\right)^n}{n!} + \dots\dots$$

since these is no odd powers of t, the coefficients of $t^{2n+1}=0$ for n=0,1,2,.......

$$\therefore \mu_{2n-1}=0 \quad \text{and } \mu_{2n}=\text{coefficient of } \frac{t^{2n}}{(2n)!}=\frac{\left(\frac{1}{2}\sigma^2\right)^n (2n)!}{n!}=1.3.5.\ldots\ldots(2n-1)\sigma^{2n}$$

Mean and mode of normal distribution.

We have ,

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

Taking log of both sides $\log f(x)=-\log\sigma\sqrt{2\pi}-\frac{1}{2}\cdot\frac{1}{\sigma^2}(x-\mu)^2$

Differentiating w.r.t x, $\frac{f'(x)}{f(x)}=-\frac{1}{\sigma^2}(x-\mu)\Rightarrow f'(x)=-\frac{x-\mu}{\sigma^2}.f(x).$

Differentiating again, $f''(x)\,x-\frac{1}{\sigma^2}[f(x)+(x-\mu)f'(x)]=-\frac{f(x)}{\sigma^2}\left[1-\left(\frac{x-\mu}{\sigma}\right)^2\right]$

$$f'(x)=0\Rightarrow x=\mu$$

Also $f''(x)]_{x=\mu}=-\frac{1}{\sigma^2}f(x)]_{x=\mu}=-\frac{1}{\sigma^2}\cdot\frac{1}{\sigma\sqrt{2\pi}}<0$

Hence mean = μ= mode.

Median of normal distribution .

Let a be the median

Then $\int_{-\infty}^{a} f(x)\,dx=\frac{1}{2}$

$$\int_{-\infty}^{a} f(x)\,dx=\int_{-\infty}^{\mu}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)dx+\int_{\mu}^{a}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)dx$$

$$=\frac{1}{2}+\frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{a}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)dx \quad \text{since } \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mu}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx=\tfrac{1}{2}$$

So, $\frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{a} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}dx = 0$ i.e. $\mu = a$. Thus for normal distribution mean = median = mode.

Recurrence Relation :

We know,

$$\mu_{2r} = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-\mu)^{2r}\exp\left(-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2\right)dx$$

differentiating w.r.t σ,

$$\frac{d\mu_{2r}}{d\sigma} = -\frac{1}{\sigma^2\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-\mu)^{2r}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx + \frac{1}{\sigma^4\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-\mu)^{2r+2}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

$$\Rightarrow \sigma^3\frac{d\mu_{2r}}{d\sigma} = -\sigma^2\mu_{2r} + \mu_{2r+2}$$

$$\Rightarrow \mu_{2r+2} = \sigma^2\mu_{2r} + \sigma^3\frac{d\mu_{2r}}{d\sigma}$$

Putting $r = 0, 1, 2, \ldots\ldots\ldots$ we get

$$\mu_2 = \sigma^2\mu_0 + \sigma^3.0 = \sigma^2 \qquad \text{since } \mu_0 = 1.$$

$$\mu_4 = \sigma^2\mu_2 + \sigma^3.\frac{d}{d\sigma}(\sigma^2) = \sigma^4 + 2\sigma^4 = 3\sigma^3 \qquad \text{and so an.}$$

Area under normal curve.

The probability P for the interval from the mean μ to a value x is given by the definite integral

$$P(\mu \le X < x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{\mu}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)}dx = \frac{1}{\sqrt{2\pi}}\int_{0}^{z} e^{-\frac{z^2}{2}}dz \quad \text{where} \quad z = \frac{x-\mu}{\sigma}$$

The values of P for values of z at intervals of 0.01 have been tabulated in standard table. The value given by the above definite integral is called the normal probability integral.

The probability that a random value of the normal variate will fall within the interval $x = \mu - \sigma$ to $x = \mu + \sigma$ is

$$P(x-\mu < x < x+\mu) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-\sigma}^{\mu+\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$\text{or } P(-1<Z<1) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-\frac{1}{2}z^2} dz = \frac{2}{\sqrt{2\pi}} \int_{0}^{1} e^{-\frac{z^2}{2}} dz = 2\times 0.3413 = 0.6826 \text{ from the table.}$$

The probability that a random value of the normal variate will deviate more than σ from the mean is (1-0.6826)=0.3174 which is less then $\frac{1}{3}$.

The probability that a random value of the normal variate deviates more then 2σ from the mean is

$$P\left(\left|\frac{x-\mu}{\sigma}\right| \geq 2\right) = 1 - P[\mu-2\sigma<x<\mu+2\sigma] = 1 - \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-2\sigma}^{\mu+2\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \text{ is}$$

$$1-P(-2\leq z\leq 2) = 1-2\frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{2} e^{-\frac{z^2}{2}} dz \quad \text{where} \quad Z = \frac{x-\mu}{\sigma}$$

$$= 1-2(0.4772) \text{ from table } = 0.0456$$

$$\therefore P(|z|>2) = 0.0456 \text{ which is less than } 5\%$$

For further illustration let X be N(100,10)

Then $Z = \frac{X-100}{10}$ is N(0,1).

For P(X<130), X=130 and $Z = \frac{130-100}{10} = 3$.

P(X<130)=P(Z<3)=0.9987 (From table)

when X=120, $z = \frac{120-100}{10} = 2$

P(X>120)=P(z>2)=1-P(z<2)=1-0.9772=0.0228.

P(120< x <130)=P(2< z <3) =P(z < 3)-P(z < 2) =0.9987-0.9772 =0.0215

Thus from the tables of probabilities P(z < c) where z is N(0,1) we can find probabilities of X where X is N(μ,σ) lying in any interval.

Illustrative Examples related to application .

Example 1.

In a normal distribution it is known that 8% of the items are below 4.4, while 90% of then are between 4.4 and 18. Find the mean and s.d. of the distribution $P(X<4.4)=0.08$ and $P(4.4<x<18)=0.90$.

Solution : From the table of areas under normal curve we get corresponding z-values to be -1.4 and 2.

$$\text{Now } Z=\frac{x-\mu}{\sigma} \Rightarrow X=\mu+z\sigma. \quad \therefore 4.4=\mu-1.4\sigma \quad \text{and } 18=\mu+2\sigma.$$

From these two equations. $13.6=3.4\sigma \Rightarrow \sigma=4$ and $\mu=10$.

Example 2.

If skulls are classified as A.B.C according to the length -breadth index is under 75, between 75 and 80, over 80. Find the mean and s.d of series in which A are 58%, B are 38% and C are 4% being given

that if $$f(t)=f(x)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}\,dx \text{ then } f(0.20)=0.08 \text{ and } f(1.75)=0.46$$

Solution : If $t=0.20$, the area of the curve from $t=0$ to t is 0.08 in the area to the left of the ordinate t is $(.50+.08)=0.58$

The area corresponding to $x=75$

$$\therefore t=\frac{75-\mu}{\sigma}$$

$$\text{Hence } \frac{75-\mu}{\sigma}=0.20 \Rightarrow 75\text{-}0.20\sigma=\mu \qquad (1)$$

For $t=\frac{80-\mu}{\sigma}$, the area to the right of the ordinate at $x=80$ is given to be 0.04. Hence the area to the left of this ordinate is $(1-.04)=0.96$.

Thus the area of the curve from 0 to $t=0.96-0.50=0.46$. The corresponding value of t is 1.75

$$\text{Hence } \frac{80-\mu}{\sigma}=1.75 \Rightarrow \mu=80\text{-}1.75\sigma \qquad (2)$$

Solving (1) and (2) We get $75\text{-}20\sigma=80\text{-}1075\sigma \Rightarrow 1.55\sigma=5 \Rightarrow \sigma=3.2$ and $\mu=74.4$

Gamma Distribution

Definition: The continuous random variate X which id distributed with probability density function f(x) given by

$$f(x)=\begin{cases}\dfrac{m^{\tau}}{\lceil\tau}.x^{\tau-1}e^{-mx}, & \text{if} \quad 0<x<\infty, \quad \tau\geq 0, \quad m\geq 0 \\ 0 & \text{otherwise}\end{cases}$$

is called Gamma variate with parameters τ and m and its distribution is called Gamma distribution.

$$\int_0^{\infty} f(x)dx=\frac{m^{\tau}}{\lceil\tau}\int_0^{\infty} e^{-mx}\,x^{\tau-1}\,dx \qquad \text{put } mx=z$$

$$=\frac{m^{\tau}}{\lceil\tau}\int_0^{\infty} e^{-z}\left(\frac{z}{m}\right)^{\tau-1}\frac{dz}{m}$$

$$=\frac{m^{\tau}}{\lceil\tau}\int_0^{\infty} e^{-z}\frac{z^{\tau-1}}{m^{\tau}}\,dz=\frac{1}{\lceil\tau}\int_0^{\infty} e^{-z}z^{\tau-1}dz=\frac{1}{\lceil\tau}.\lceil\tau=1$$

Hence f(x) is a probability density function.

Mean and variance

$$\mu_1'=E(X)=\frac{m^{\tau}}{\lceil\tau}\int_0^{\infty} x\,e^{-mn}\,x^{\tau-1}dx.$$

$$=\frac{m^{\tau}}{\lceil\tau}\int_0^{\infty} e^{-mx}\,x^{\tau}dx=\frac{m^{\tau}}{\lceil\tau}\int_0^{\infty} e^{-z}\left(\frac{z}{m}\right)^{\tau}\frac{dz}{m}$$

$$=\frac{m^{\tau}}{\lceil\tau}\int_0^{\infty}\frac{1}{m^{\tau+1}}.e^{-z}z^{\tau}\,dz=\frac{1}{\lceil\tau}.\frac{1}{m}.\lceil\tau+1=\frac{\tau}{m}.$$

$$\mu_2' = \frac{m^\tau}{\lceil\tau} \int_0^\infty x^2 e^{-mx} x^{\tau-1} dx$$

$$= \frac{m^\tau}{\lceil\tau} \int_0^\infty e^{-z} \left(\frac{z}{m}\right)^{\tau+1} \frac{dz}{m} = \frac{1}{\lceil\tau} \cdot \frac{1}{m^2} \int_0^\infty e^{-z} z^{\tau+1} dz$$

$$= \frac{1}{\lceil\tau} \cdot \frac{1}{m^2} \lceil\overline{\tau+2} = \frac{\tau(\tau+1)}{m^2}$$

$$\text{Var }(x) = \mu_2' - \mu_1'^2 = \frac{\tau(\tau+1)}{m^2} - \frac{\tau^2}{m^2} = \frac{\tau}{m^2}.$$

In Gamma distribution mean = variance if $m = 1$.

In general.

$$\mu_r' = \frac{m^\tau}{\lceil\tau} \int_0^\infty e^{-mx} x^{\tau-1+r} dx$$

$$= \frac{m^\tau}{\lceil\tau} \cdot \int_0^\infty e^{-z} \frac{Z^{\tau+r-1}}{m^{\tau+r-1}} \cdot \frac{dz}{m} = \frac{1}{\lceil\tau} \cdot \frac{1}{m^r} \cdot \int_0^\infty e^{-z} z^{\tau+r-1} dz$$

$$= \frac{1}{\lceil\tau} \cdot \frac{1}{m^r} \lceil\overline{\tau+r}$$

$$r=3 \Rightarrow \mu_3' = \frac{1}{\lceil\tau} \cdot \frac{1}{m^3} \tau(\tau+1)(\tau+2) \lceil\tau = \frac{\tau(\tau+1)(\tau+2)}{m^3}$$

$$r=4 \Rightarrow \mu_4' = \frac{1}{\lceil\tau} \cdot \frac{1}{m^4} \cdot \lceil\overline{\tau+4} = \frac{\tau(\tau+1)(\tau+2)(\tau+3)}{m^4}$$

Now,

$$\mu_3 = \mu_3' - 3\mu_2' \mu_1' + 2\mu_1'^3$$

$$= \frac{\tau(\tau+1)(\tau+2)}{m^3} - 3\frac{\tau(\tau+1)}{m^2} - \frac{\tau}{m} + 2\frac{\tau^3}{m^3} = \frac{\tau(\tau+1)}{m^3}(\tau+2-3\tau) + \frac{2\tau^3}{m^3}$$

$$= \frac{2\tau}{m^3}.$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\,\mu'_2{}^2 - 3\mu'_1{}^4$$

$$= \frac{\tau(\tau+1)(\tau+2)(\tau+3)}{m^4} - 4\frac{\tau(\tau+1)(\tau+2)}{m^3}\,\frac{\tau}{m} + 6\frac{\tau(\tau+1)}{m^2}\cdot\frac{\tau^2}{m^2} - 3\,\frac{\tau^4}{m^4}$$

$$= \frac{\tau}{m^4}[(\tau+1)(\tau+2)(\tau+3-\tau)+3\tau^2(2\tau+2-\tau)]$$

$$= \frac{\tau}{m^4}[3(\tau+1)(\tau+2)(1-\tau)+3\tau^2(\tau+2)]$$

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

Measures of Skewness and Kurtosis :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{4\tau^2}{m^6}\cdot\frac{m^6}{\tau^3} = \frac{4}{\tau}. \qquad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\tau(\tau+2)}{m^4}\cdot\frac{m^4}{\tau^2} = 3+\frac{6}{\tau}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{2}{\sqrt{\tau}} \quad \text{and} \quad \gamma_2 = \beta_2 - 3 = \frac{6}{\tau}$$

The distribution is therefore positively skewed and leptokurtic.

If $\tau \to \infty$ then $\beta_1 \to 0, \beta_2 \to 3$ and $\gamma_1 \to 0\ \gamma_2 \to 0$

This suggest that the limiting form of Gamma distribution is symmetric and meso-kurtic.

Beta Distribution-

Definition : the continuous random variate X which is distributed with probability density function f(x) given by

$$f(x) = \begin{cases} \dfrac{1}{\beta(\tau,m)}\, x^{\tau-1}(1-x)^{m-1} & 0 \le x \le 1,\ \tau, m>0 \\ 0 & \text{otherwise} \end{cases}$$

is a Beta variate of 1st kind with parameters τ and m.

$$\text{Now} \quad \int_0^1 f(x)dx = \frac{1}{\beta(\tau m)} \int_0^1 x^{\tau-1} (1-x)^{m-1} \, dx$$

$$= \frac{1}{\beta(\tau, m)} . \beta(\tau, m) = 1.$$

Hence f(x) is a probability density function.

Mean and variance

$$\mu_1' = E(X) = \frac{1}{\beta(\tau,m)} \int_0^1 x . x^{\tau-1} (1-x)^{m-1} \, dx$$

$$= \frac{\beta(\tau+1,m)}{\beta(\tau,m)} = \frac{\Gamma(\tau+1)\Gamma(m)}{\Gamma(\tau+m+1)} . \frac{\Gamma(\tau+m)}{\Gamma(\tau)\Gamma(m)} = \frac{\tau}{\tau+m}$$

$$\text{So,} \quad \text{Mean} = \frac{\tau}{\tau+m}$$

$$\mu_2' = \frac{1}{\beta(\tau,m)} \int_0^1 x^2 . x^{\tau-1} (1-x)^{m-1} \, dx$$

$$= \frac{1}{\beta(\tau,m)} . \beta(\tau+2,m) = \frac{\Gamma(\tau+2)\Gamma(m)}{\Gamma(\tau+m+2)} . \frac{\Gamma(\tau+m)}{\Gamma(\tau)\Gamma(m)}$$

$$= \frac{\tau(\tau+1)}{(\tau+m+1)(\tau+m)}$$

$$Var(X) = \mu_2 = \mu_2' - \mu_1^2$$

$$= \frac{\tau(\tau+1)}{(\tau+m+1)(\tau+m)} - \frac{\tau^2}{(\tau+m)^2} = \frac{\tau}{(\tau+m+1)(\tau+m)^2}\left(\tau^2 + m\tau + \tau + m - \tau^2 - \tau m - \tau\right)$$

$$= \frac{\tau m}{(\tau+m)^2(\tau+m+1)}$$

Moments :

$$\mu_r^1 = \frac{1}{\beta(\tau,m)} \int_0^1 x^{\tau+\gamma-1}(1-x)^{m-1}\,dx$$

$$= \frac{1}{\beta(\tau,m)}\,\beta(\tau+\gamma,m)$$

Putting $r=3,\ \mu_3' = \frac{1}{\beta(\tau,m)}\,\beta(\tau+3,m) = \frac{(\tau+2)(\tau+1)\tau}{(\tau+m+2)(\tau+m+1)(\tau+m)}$

$$r=4,\quad \mu_4' = \frac{1}{\beta(\tau,m)}\,\beta(\tau+4,m) = \frac{(\tau+3)(\tau+2)(\tau+1)\tau}{(\tau+m+3)(\tau+m+2)(\tau+m+1)(\tau+m)}$$

Now, $\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$

$$= \frac{(\tau+2)(\tau+1)\tau}{(\tau+m+2)(\tau+m+1)(\tau+m)} - 3\,\frac{\tau(\tau+1)}{(\tau+m+1)(\tau+m)}\cdot\left(\frac{\tau}{\tau+m}\right) + 2\,\frac{\tau^3}{(\tau+m)^3}$$

$$= \frac{2\tau m(m-\tau)}{(\tau+m)^3(\tau+m+1)(\tau+m+2)}$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 \;=\; \frac{3\tau m\left\{\tau m(\tau+m-6)+2(\tau+m)^2\right\}}{(\tau+m)^4(\tau+m+1)(\tau+m+2)(\tau+m+3)}$$

Skewness and kurtosis

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{4(m-\tau)^2(\tau+m+1)}{\tau m(\tau+m+2)^2} \qquad \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3(\tau+m+1)\,\tau m\,(\tau+m-6)+2(\tau+m)^2}{\tau m(\tau+m+2)(\tau+m+3).}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{2(m-\tau)\sqrt{\tau+m+1}}{\sqrt{\tau m}\;(\tau+m+2)} \qquad \gamma_2 = \beta_2 - 3.$$

B- distribution of 2nd kind

The continuous random variate X which is distributed with probability density function f(x) defined by

$$f(x)=\begin{cases}\dfrac{1}{\beta(\tau,m)}\cdot\dfrac{x^{\tau-1}}{(1+x)^{\tau+m}} & \text{if } x\geq 0,\ \ \tau,\ \ m>0\\ 0 & \text{if } x<0.\end{cases}$$

is a β- variate of second kind with parameter τ and m.

$$\int_0^\infty f(x)\ dx=\frac{1}{\beta(\tau,m)}\int_0^\infty\frac{x^{\tau-1}}{(1+x)^{\tau+m}}\ dx$$

$$=\frac{1}{\beta(\tau,m)}\int_0^1 y^{m-1}(1-y)^{\tau-1}dy \quad \text{put } 1+x=\frac{1}{y}$$

$$=\frac{1}{\beta(\tau,m)}\beta(\tau,m)=1.$$

Hence f(x) defines a probability density function.

Moments

$$\mu'_r=\int_0^\infty\frac{1}{\beta(\tau,m)}\cdot\frac{x^{\tau+r-1}}{(1+x)^{\tau+m}}dx$$

$$=\frac{1}{\beta(\tau,m)}\int_0^1 y^{m-r-1}(1-y^{\tau+r-1}dy \quad \text{putting} \quad \frac{1}{y}=1+x$$

$$=\frac{1}{\beta(\tau,m)}\int_0^1 z^{\tau+r-1}(1-z)^{m-r-1} \quad \text{where } z=1-y$$

$$=\frac{\beta(\tau+r,\ m-r)}{\beta(\tau,m)} \quad \text{if } m>r$$

$$r=1\Rightarrow\mu'_1=\frac{\beta(\tau+1,m-1)}{\beta(\tau,m)}=\frac{\overline{|\tau+1}\ \overline{|m-1}}{|\tau+m}\cdot\frac{\overline{|\tau+m}}{|\tau\ |m}$$

$$\text{Mean} = \frac{\tau}{m-1}.$$

Hence mean of the distribution exist only if m >1

$$r=2 \Rightarrow \mu_2' = \frac{\beta(\tau+2;m-2)}{\beta(\ell.m)} = \frac{\overline{|\tau+2}\,\overline{|\tau+m}}{\overline{|\tau+m}}\cdot\frac{\overline{|\tau+m}}{\overline{|\tau}\,\overline{|m}} = \frac{(\tau+1)\tau}{(m-1)(m-2)}.$$

$$\text{Var }(x) = \mu_2' - \mu_1'^2$$

$$= \frac{\tau(\tau+1)}{(m-1)(m-2)} - \frac{\tau^2}{(m-1)^2} = \frac{\tau}{(m-1)^2(m-2)}\,[(\tau+1)(m-1)-\tau(m-2)$$

$$= \frac{\tau}{(m-1)^2(m-2)}\,(\tau+m-1)$$

Differential Equation for Pearsonian system.

Karl Pearson dealing with generalized frequency curves, derived some types of frequency curves in addition to normal curves. He considered the following characteristics of a frequency distribution.

1. A frequency distribution generally starts at zero and rises to a single maximum (mode) and then falls again . Hence generally a frequency distribution is unimodal.

If $y=f(x)$ be a frequency curve then,

$$\frac{dy}{dx} = 0 \text{ for some } x = -a.$$

2. At the ends of the frequency curve there is often high contact with the x-axis.

$$\frac{dy}{dx} = 0 \text{ where } y = 0.$$

3. Generally first four moments are sufficient to determine the distribution.

The condition (1) and (2) give,

$$\frac{dy}{dx} = -\frac{y(x+a)}{F(x)} \qquad (1)$$

where F(x) is an arbitrary function of x, not vanishing at $x = -a$.

Expanding F(x) by Maclaurius series,

$$F(x) = b_0+b_1x+b_2x^2+b_3x^3+\dots\dots\dots$$

Empirically we retain only first three terms in the expansion of F(x).

Hence $F(x) = b_0 + b_1 x + b_2 x^2$

Equation (1)

$$\Rightarrow \frac{dy}{dx} = -\frac{y(x+a)}{b_0 + b_1 x + b_2 x^2} \qquad (2)$$

This is the differential equation for determination of Pearson's type of curves.

Solutions of the D.E.

The general form of (2) is

$$(b_0 x^n + b_1 x^{n+1} + b_2 x^{n+2})\, dy = -y(x^{n+1} + a x^n)\, dx \qquad (3)$$

Integrating (3) by parts over the total range of variate x,

$$\left(b_0 x^n + b_1 x^{n+1} + b_2 x^{n+2}\right) y \Big]_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} \left[n b_0 x^{n-1} + (n+1) b_1 x^n + (n+2) b_2 x^{n+1} \right] y dx$$

$$= -\int_{-\infty}^{\infty} y\left(x^{n+1} + a x^n\right) dx. \qquad \text{provided the integral exists.}$$

Assuming that the first bracket an the left hand side vanishes at either limit and applying the definition of moments we have.

$$n b_0 \mu_{n-1} + (n+1) b_1 \mu_n + (n+2) b_2 \mu_{n+1} = \mu_{n+1} + a \mu_n \qquad (4)$$

Measuring x from the mean of the curve and putting n = 0,1,2,3 we get

$b_1 \mu = a\, \mu_0 \Rightarrow b_1 = a$ since $\mu_0 = 1, \mu_1 = 0$

$b_0 \mu_0 + 3 b_2 \mu_2 = \mu_2$

$3 b_1 \mu_2 + 4 b_2 \mu_3 = \mu_3 + a \mu_2$

$3 b_0 \mu_2 + 4 b_1 \mu_3 + 5 b_2 \mu_4 = \mu_4 + a \mu_3$

Eliminating b_0 from 2nd and 4th relation,

$$3 b_1 \mu_3 + b_2 (5 \beta_2 \sigma^4 - 9 \sigma^4) = (\beta_2 - 3) \sigma^4$$

Eliminating b_1 form 3rd and 5th relation,

$$b_2 = \frac{2\beta_2-3\beta_1-6}{2(5\beta_2-6\beta_1-9)}$$

From 2nd relation, $$b_0 = \frac{\sigma^2(4\beta_2-3\beta_1)}{2(5\beta_2-6\beta_1-9)}$$

From 3rd relation $$b_1 = \frac{\sigma\sqrt{\beta_1}\,(\beta_2+3)}{2(5\beta_2-6\beta_1-9)}$$

where $$\mu_2=\sigma^2,\quad \beta_1=\frac{\mu_3^2}{\mu_2^3} \quad \text{and} \quad \beta_2=\frac{\mu_4}{\mu_2^2}.$$

Solution of the equation (2) depends on the nature of the roots of the equation $b_0+b_1x+b_2x^2=0$. The discriminant of the equation is $b_1^2-4b_0b_2$.

Let us define $k=\frac{b_1^2}{4b_0b_2}$ so that discriminant becomes $4b_0b_2\,(k\text{-}1)$

Different types of curves

Case 1 : Roots of $b_0+b_1x+b_2x^2=0$. are real and of opposite sign.

Product of the roots $=\frac{b_0}{b_2}=\frac{b_0^2}{b_0b_2}$ and $b_0b_2<0$. Hence $K<0$.

Shifting the origin to the mode $x=-a$ we have.

$$\frac{1}{y}\frac{dy}{dx}=\frac{x}{B_0+B_1x+B_2x^2}=\frac{-x}{B_2(x-\alpha_1)(x-\alpha_2)} \qquad (5)$$

where α_1 and α_2 are the roots of the equation . Expressing in partial fractions,

$$\frac{-x}{B_2(x-\alpha_1)(x-\alpha_2)}=\frac{m_1}{x-\alpha_1}+\frac{m_2}{x-\alpha_2}$$

$$\Rightarrow -x=B_2[m_1(x-\alpha_2)+m_2(x-\alpha_1)]$$

Putting $x=\alpha_1$ and α_2 we get,

$$-\alpha_1=B_2 m_1(\alpha_1-\alpha_2)$$

$$\text{and} \quad -\alpha_2=B_2 m_2(\alpha_1-\alpha_2)$$

Dividing $\dfrac{\alpha_1}{\alpha_2} = -\dfrac{m_1}{m_2} \Rightarrow m_1\alpha_2+m_2\alpha_1=0$ (6)

Eliminating α_1 and α_2 from above two relations, $B_2(\alpha_1-\alpha_2)(m_1+m_2)=-(\alpha_1-\alpha_2)$

$$\Rightarrow m_1+m_2=-\frac{1}{B_2} \quad (7)$$

Now (5) becomes $\dfrac{1}{y}\cdot\dfrac{dy}{dx}=\dfrac{m_1}{x-\alpha_1}+\dfrac{m_2}{x-\alpha_2}$

Intergrating

$$\log y=m_1\log(x-\alpha_1)+m_2\log(x-\alpha_2)+\text{constant.}$$

Let us choose

$$\alpha_1=-a_1 \quad \text{and} \quad \alpha_2=a_2$$

Then,

$$\log y=m_1\log(x+a_1)+m_2\log(x-a_2)+\text{constant.}$$

$$=m_1\log a_1(1+\frac{x}{a_1})+m_2a_2\log(-a_2)(1-\frac{x}{a_2})+\text{constant}$$

$$\Rightarrow y=y_0\left(1+\frac{x}{a_1}\right)^{m_1}\left(1-\frac{x}{a_2}\right)^{m_2} \text{ where } y_0 \text{ is a constant} \quad (8)$$

where $m_1a_2-m_2a_1=0 \Rightarrow \dfrac{m_1}{a_1}=\dfrac{m_2}{a_2}$. **(8) is the frequency curve of type 1.**

Case II

Roots of the equation $b_0+b_1x+b_2x^2=0$ are equal but of opposite sign.

Here $b_1=0$ and b_0 and b_2 are of opposite sign so that $\dfrac{b_0}{b_2}>0$

Accordingly k=0 and Pearson differential equation reduces to

$$\frac{1}{y}\cdot\frac{dy}{dx} = -\frac{x}{b_2\left(x^2+\frac{b_0}{b_2}\right)} = -\frac{x\cdot}{b_2(x^2+c^2)} \quad \text{where } c^2 = \frac{b_0}{b_2} \text{ Integrating,}$$

Integrating, $\log y = -\frac{1}{2b_2}\log(x^2+c^2) + \log y_0$

$$\Rightarrow y = y_0\,(x^2+c^2)^{-\frac{1}{2b_2}} \Rightarrow y = y_0\left(1+\frac{x^2}{c^2}\right)^{-m} \quad \text{where } -c \leq x < c$$

This is pearsonian curve of type II.

Corollary : If $m=0$ then $y=$ constant and the distribution becomes rectangular.

Case III.

One root of $b_0+b_1x+b_2x^2=0$ is infinite.

In this case $b_2=0$ and $b_1 \neq 0$ so that $k \to \infty$

Then Pearsonian differential equation becomes

$$\frac{1}{y}\cdot\frac{dy}{dx} = -\frac{x}{b_0+b_1x}.$$

$$\Rightarrow \frac{dy}{y} = -\frac{xdx}{b_1(x+a)} \quad \text{where } c=\frac{b_0}{b1}$$

$$= -\frac{1}{b_1} + \frac{c}{b_1(x+c)}$$

Integrating,

$$\log y = \text{constant}\left(-\frac{x}{b_1}\right) + \frac{c}{b_1}\log(x+c)$$

$$\Rightarrow y = y_0\left(1+\frac{x}{c}\right)^p e^{-px/c} \qquad \text{where } p=\frac{c}{b} \text{ and } -c<x<\infty$$

This is pearsonian curve of type III.

Case IV

Roots of the equation $b_0+b_1x+b_2x^2=0$ are real and of same sign.

In this case k>1

This will follow the case I.

Accordingly let the roots be α_1 and α_2 so that

$\alpha_1=-a_1$ and $\alpha_2=-a_2$

$$\text{then } \frac{m_1}{a_1}=-\frac{m_2}{a_2}$$

In this case equation of the curve is

$$y=y_0\left(1+\frac{x}{a_1}\right)^{m_1}\left(1+\frac{x}{a_2}\right)^{-m_2}$$

This is Pearsonian curve of type IV.

Correlation and Regression

In earlier chapter our discussions have been confined to a single variable and distribu.ion we have discussed is univariate distributions only. But in statistical work we have often to deal with problems involving more than one variable. First let us consider the case of two variables. So we shall extend our discussions to bivariate distributions. In particular we shall study the simultaneous, variation of two variables X and Y. These variables for example may be heights and weights of students of a class, marks secured by students in mathematics and physics, records of rainfall and yields of crops etc. So our interest lies in studying the relationship between two variables. The relationship may be of any type. But the easiest and of great interest in Statistics is that of linear type. So we shall confine our discussion to linear type only.

Bivariate distribution:

So far in our discussions we have confined to a single variable except in the case of normal bivariate distributions. We shall now study in simultaneous variation of two variables X and Y. The relation between the two variables may be of any type, but we shall confine our discussion to linear type.

We know that $E(X-\bar{x})$ and $E(Y-\bar{y})$ given a measure for the variation of X and Y respectively. Normally we expect $E[(X-\bar{x}(Y-\bar{y})]$ to give a measure for simultaneous variations of X and Y. But this will depend on units used for X and Y. To make it independent of units. We divide it by quantity having same dimension as $E[(X-\bar{x})(Y-\bar{y})]$.

The quantity $E[(X-\bar{x})(Y-\bar{y})]$. is called co-variance of X and Y and is denoted by cov (x,y) or σ_{xy}

$$\sigma_{xy}=\text{cov}(x,y)=E[(X-\bar{x})(Y-\bar{y})].$$
$$=E[(XY-\bar{x}y-\bar{y}x+\bar{x}\,\bar{y}]$$
$$E[(XY)-\bar{x}E(Y)-\bar{y}E(x)+E(\bar{x}\,\bar{y})$$
$$E[(XY)-\bar{x}\,\bar{y}-\bar{y}\,\bar{x}+\bar{x}\,\bar{y}$$
$$=E(XY)-\bar{x}\,\bar{y}$$

The measure of the linear relationship between two variables X and Y is denoted by γ_{XY} or γ and is called the product moment correlation coefficient.

$$\gamma_{XY}=\frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X),\text{var}(Y)}}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y.}$$

$$=\frac{\Sigma xy-\frac{(\Sigma x)(\Sigma y)}{n}}{\sqrt{\Sigma x^2-\frac{(\Sigma x)^2}{n}}\sqrt{\Sigma y^2-\frac{(\Sigma y)^2}{n}}}\quad \text{since cov}\,(x,y=\frac{1}{n}\left[\Sigma xy-\frac{\Sigma x\Sigma y}{n}\right].$$

obviously $\gamma_{XY}=\gamma_{YX}$.

The concept of correlation coefficient was formulated by Karl Pearson.

Change of origin and scale. Let $U=\frac{X-a}{h}$ and $V=\frac{y-b}{k}$ so that

$$E(x)=a+hE(U) \text{ and } E(y)=b+kE(v).$$

$$\gamma_{XY}=\frac{E[(X-\bar{x})(Y-\bar{y})]}{\sqrt{E(X-\bar{x})}\sqrt{E(Y-\bar{y})^2}}=\frac{hkE[(u-\bar{u})(v-\bar{v})]}{hk\sqrt{E(u-\bar{u})^2}\sqrt{E(v-\bar{V})^2}}=\gamma_{UV}$$

This value of correlation coefficient is independent of the origin of reference and the unit of measurement in scale. In other words γ is a real number and it has no dimensions.

Theorem: If X and Y are independent random variables they are un correlated.

X and Y are independent $\Rightarrow E(xy)=E(x)E(y)=\overline{XY}$

$$\text{Now } \sigma_{xy}=E(xy)-\bar{x}\,\bar{y}=\bar{x}\,\bar{y}-\bar{x}\,\bar{y}=0$$

$$\text{So } \gamma_{xy}=\frac{0}{\sigma_x\,\sigma y}$$

Hence X and Y are un correlated.

Before discussing different properties of correlation coefficient let us first establish Schwarz's in equality for random variable X and Y.

$$[E(XY)]^2 \leq E(X^2)E(Y^2)$$

Proof : For every real constant a we have.

$$E(aX-Y)^2 \geq 0$$

$$\Rightarrow a^2E(X^2)-2aE(XY)+E(Y^2) \geq 0$$

In particular this bold for

$$a = \frac{E(XY)}{E(X^2)}$$

Substituting this value of a we have,

$$\frac{[E(XY)]^2}{E(X^2)} - 2\frac{[E(XY)]^2}{E(X^2)} + E(Y^2) \geq 0$$

$$\Rightarrow -[E(XY)]^2 + E(X^2)E(Y^2) \geq 0$$

$$\Rightarrow [E(XY)]^2 \leq E(X^2)E(Y^2)$$

Limits of the coefficient of correlation:

If x', y' denote deviations of the variate x and y from respective mean then from Schwarz's in equality we have

$$[E(x'y')]^2 \leq E(x'^2)E(y'^2)]$$

Dividing both sides by $E(x'^2)\,E(y'^2)$ we get

$$\frac{[E(x'y')]^2}{E\left(x'^2\right)E\left(y'^2\right)} \leq 1$$

$$\Rightarrow \gamma^2_{xy} \leq 1 \quad \Rightarrow -1 \leq \gamma^2_{xy} \leq 1$$

Interpretation of the value of γ :

The value of γ is a measure of strength of the linear relationship between the set of values of x_i and y_i. A large value of γ indicates that there is a strong linear relationship between x_i and y_i in the points (x_i, y_i) the near a straight line. A large positive value of γ indicates that the straight line has a positive slope and a large negative value of γ indicates that the line has negative slope. The value γ=1 or γ=-1 indicates that all points lie in a straight line. A small value of γ indicates that these is no linear relationship between x_i and y_i i.e. the points (x_i, y_i) do not lie near a line. Thus if a relationship exists it is not linear.

Regression

From above discussion we have seen that a high value of γ imply that the points will lie near about a straight line. The method to study this line, how to find it and its usefulness is called regression method. The objectives of the regression analysis are the formulation and determination of the mathematical form of the relationship between the variables and use its for prediction purposes.

When the points (x_i, y_i) i=1,2, ...,n are plotted in the xy-plane of a rectangular coordinate system, th set of points is called a scatter diagram. From scatter diagram it is possible to visualize a smooth curve approximations the data. The curve is called approximating curve. The problem of finding equation of approximating curves that fit given sets of data is called curve fitting. In practice the type of equations often suggested from scatter diagram is either

$y = ax+b$	straight line
or $y = a + bx + cx^2$	quadratic curve

A curve is completely known when the values of the parameters are known. Thus fitting of a straight line to a set of points (x_i, y_i) i=1,2,------ n amounts to finding the values of a and b, with the help of x_i, y_i such that the line will fit the set of points wall. Similar the processes for fitting a quadratic curve.

Let us consider the case, in which the points are (x_i, y_i) i=1,2,.....,n. For a given value x_i of x there will be difference between the value y_i. Let the difference d_i.

Of all curves in a given family of curves approximating a set of n-data points a curve with $d_1^2+d_2^2+.........+d_n^2$ = minimum, is the best fitting curve in the family. A curve having this property is called least square regression curve.

Derivation of normal equations for the least square line :

Since the equation of the line is assumed to be y = ax + b the values of y and the least square line corresponding to $x_1, x_2, ..., x_n$ are $a + bx_1, a+ bx_2, ..., a+bx_n$. The corresponding deviations are

$d_i=a+bx_i-y_i$

The sum of the squares of the deviations is

$$S = \sum_1^n d^2 = \sum_1^n (a + bx \quad - y)^2$$

A necessary condition for this to be minimum is

$$\frac{\partial s}{\partial a} = 0 \Rightarrow \Sigma\, 2(a + bx - y) = 0$$

$$\frac{\partial s}{\partial b} = 0 \Rightarrow \Sigma\, 2\, x(a + bx - y) = 0\,.$$

so we obtain,

$$\left.\begin{aligned} \Sigma y &= an + b\Sigma x \\ \text{and} \quad \Sigma xy &= a\Sigma x + b\Sigma x^2 \end{aligned}\right\} \qquad (1)$$

(1) are called the normal equations for the least square line.

Solving (4) for values of a and b we have,

$$\left.\begin{aligned} a &= \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n\Sigma x^2 - (\Sigma x)} \\ b &= \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2} \\ &= \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} \end{aligned}\right\} \qquad (2)$$

Dividing first equation of (1) by n we get.

$$\bar{y} = a + b\bar{x} \qquad (3)$$

Eliminating (3) and from

least square line $y=a+bx$ (4)

$$Y - \bar{y} = b(x - \bar{x}) \qquad (5)$$

(5) shows that b is the slope of the line (4) and the line passes through $(\bar{x}, \bar{y})$.

The constant b is known as the regression coefficient of y on x and hence some times 'b' is written as b_{yx}.

From (2),

$$b_{y_x} = \frac{\mu_{11}}{\sigma_x^2}, \quad \text{where } \mu_{11} = \text{Cov}(x,y)$$

$$= \frac{\gamma\sigma_x\sigma_y}{\sigma_x{}^2} = \gamma\frac{\sigma_y}{\sigma_x} \qquad \sigma_x \neq 0.$$

$$(3) \qquad \Rightarrow a = \bar{y} - \gamma\frac{\sigma_y}{\sigma_x}\bar{x}$$

and (5) $\Rightarrow y-\bar{y}=\gamma\frac{\sigma_y}{\sigma_x}(x-\bar{x})$ (6)

Similarly the equation of the line of regression of x on y is

$$x-\bar{x}=\gamma\frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

The regression coefficient 'b' is independent of origin of coordinates.

Let us change the origin such that $x=x'+h$, $y=y'+k$ where h,k, are constants. Then

$$b=\frac{n\Sigma x'y'-(\Sigma x')(\Sigma y')}{n\Sigma x'^2-(\Sigma x')^2}=\frac{\Sigma(x'\bar{x}')(y'-\bar{y}')}{\Sigma(x'-\bar{x}')^2}$$

Proof $x=x'+h,\quad y=y'+k\Rightarrow\bar{x}=\bar{x}'+h,\quad\bar{y}=\bar{y}'+k.$

Then

$$b=\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}=\frac{\Sigma(x'-\bar{x}')(y'-\bar{y}')}{\Sigma(x'-\bar{x}')^2}$$

$$=\frac{n\Sigma x'y'-(\Sigma x')(\Sigma y')}{n\Sigma x'^2-(\Sigma x')^2}$$

where x,y have been replaced by x', y'. Thus it is invariant under change of origin and scale.

Rank- correlation :

If we have a group of individuals ranked according to different characters, it is natural to enquire whether the ranks can be made to give us measure of degree of relation between the two characters. Suppose we have n individuals, whose rank according to character A are $x_1,x_2,.......,x_n$ and according to character B are $y_1,y_2,.......,y_n$ where x's and y's are discrete n numbers from 1 to n.

$$\Sigma x_i=\frac{n(n+1)}{2}=\Sigma y_i\Rightarrow\bar{x}=\bar{y}=\frac{n+1}{2}$$

$$\Sigma(x_i-\bar{x})^2=\Sigma x_i^2-2\bar{x}\Sigma x_i+\Sigma\bar{x}^2$$

$$=\frac{n(n+1)(2n+1)}{6}-(n+1)\frac{n(n+1)}{2}+\frac{(n+1)^2}{4}.n$$

$$=\frac{n(n+1)}{12}[4n+2-6n-6+3n+3]$$

$$=\frac{n(n+1)}{12}(n-1)=\frac{n(n^2-1)}{12}=\Sigma(y-\bar{y})^2$$

$$var(x)=\frac{n^2-1}{12}=var(y)$$

If d_i stands for difference in ranks of ith individuals we have

$$d_i=(x_i-\bar{x})-(y_i-\bar{y})=x_i'-y_i'=x_i'-y_i' \quad \text{since } \bar{x}=\bar{y}$$

where x_i' and y_i' are deviations of x_i and y_i from $\bar{x}$ and $\bar{y}$

$$\sum_1^n d_i^2=\sum_1^n x_i'^2+\sum_1^n y_i'^2-2\sum_1^n x_i'y_i'$$

$$\Rightarrow \sum_1^n x_i' y_i'=\frac{1}{2}\left[\frac{n^3-n}{6}-\sum_1^n d_i^2\right]$$

coefficient of correlation between x and y will be

$$\gamma=\frac{\sum_i^n x_i'y_i'}{\Sigma x_i'^2 \Sigma y_i'^2}=\frac{\frac{1}{2}\left[\frac{n^3-n}{6}-\Sigma d_i^2\right]}{\frac{1}{12}\left(\frac{n^3-n}{6}\right)}$$

$$=1-\frac{\Sigma d_i^2}{n(n^2-1)} \qquad (7)$$

Multiple and Partial Correlation :

Let us suppose that we are given n sets of corresponding values of three variables, x_1, x_2, x_3 and they are measured from their respective means. The regression equation of x_1 on x_2 and x_3 can be written as

$$X_1=a+b_{12.3}x_2+b_{13.2}x_3 \qquad (8)$$

where the constants $a, b_{12.3}, b_{13.2}$ are such as to give on the average the cost estimate of x_1, corresponding to any assigned values of x_2 and x_3.

The sum of the squares of the residuals is

$$S=\Sigma(x_1-X_1)^2=\Sigma(x_1-a-b_{12.3}x_2-b_{13.2}x_3)^2=\Sigma x^2_{1.23}$$

where $x_{1.23}=x_1-a-b_{12.3}x_2-b_{13.2}x_3$.

Now we are to find $a, b_{12.3}$ and $b_{13.2}$ so that S is imnimum. Then we have the equations

$$\frac{\partial s}{\partial a}=0 \Rightarrow \Sigma(x_1-a-b_{12.3}x_2-b_{13.2}x_3)=0$$

$$\frac{\partial s}{\partial b_{12.3}}=0 \Rightarrow \Sigma x_2(x_1-a-b_{12.3}x_2-b_{13.2}x_3)=0$$

and $$\frac{\partial s}{\partial b_{13.2}}=0 \Rightarrow \Sigma x_3(x_1-a-b_{12.3}x_2-b_{13.2}x_3)=0$$

These equations can be written as

$$\left.\begin{array}{l}\Sigma x_{1.23}=0\\ \Sigma x_2 x_{1.23}=0\\ \Sigma x_3 x_{1.23}=0\end{array}\right\} \quad (9)$$

(9) are normal equations for determining $a, b_{12.3}$ and $b_{13.2}$. The first of these equations gives $a = 0$. The 2nd and 3rd can be written as

$$\Sigma x_1 x_2 - b_{12.3}\Sigma x_2^2 - b_{13.2}\Sigma x_2 x_3 = 0$$

$$\Rightarrow \gamma_{12}\sigma_1\sigma_2 = b_{12.3}\sigma_2^2 + b_{13.2}\gamma_{23}\sigma_2\sigma_3 = 0 \qquad \text{(Dividing by n)}$$

$$\left.\begin{array}{ll} & \Rightarrow \gamma_{12}\sigma_1 = b_{12.3}\sigma_2 + b_{13.2}\gamma_{23}\sigma_3\\ \text{Similarly} & \gamma_{13}\sigma_1 = b_{12.3}\gamma_{23}\sigma_2 + b_{13.2}\sigma_3\end{array}\right\} \quad (10)$$

where γ_{ij} is the coefficient of correlation between x_i and x_j and σ_j is the s.d. of x_j.

Let Δ_{ij} be the co-factor in the ith row and jth column of the determinant

$$\Delta = \begin{vmatrix} 1 & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & 1 & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & 1 \end{vmatrix}$$

Then solving (10) for $b_{12.3}$ and $b_{13.2}$ we have

$$b_{12.3} = \begin{vmatrix} \gamma_{12}\sigma_1 & \gamma_{23}\sigma_3 \\ \gamma_{13}\sigma_1 & \sigma_3 \end{vmatrix} \div \begin{vmatrix} \sigma_2 & \gamma_{23}\sigma_3 \\ \gamma_{23}\sigma_2 & \sigma_3 \end{vmatrix}$$

$$= \frac{\sigma_1\sigma_3 \begin{vmatrix} \gamma_{12} & \gamma_{23} \\ \gamma_{13} & 1 \end{vmatrix}}{\sigma_2\sigma_3 \begin{vmatrix} 1 & \gamma_{23} \\ \gamma_{23} & 1 \end{vmatrix}}$$

$$= -\frac{\sigma_1}{\sigma_2} \cdot \frac{\Delta_{12}}{\Delta_{11}}$$

Similarly $b_{13.2} = -\frac{\sigma_1}{\sigma_3} \cdot \frac{\Delta_{13}}{\Delta_{11}}$

$$(8) \qquad \Rightarrow X_1 = -\frac{\sigma_1}{\sigma_2} \cdot \frac{\Delta_{12}}{\Delta_{11}} x_2 - \frac{\sigma_1}{\sigma_3} \cdot \frac{\Delta_{13}}{\Delta_{11}} . x_3$$

$$= -\frac{\sigma_1}{\Delta_{11}} \left(\frac{\Delta_{12}}{\sigma_2} x_2 + \frac{\Delta_{13}}{\sigma_3} x_3 \right)$$

$b_{12.3}$ and $b_{13.2}$ are called partial regression coefficients.

Now coriting $x_{1.2} = x_1 - b_{12} x_2$ from normal equations.

We get

$$\Sigma x_{1.23} x_{1.2} = \Sigma x_{1.23} (x_1 \text{-} b_{12}) x_2 = \Sigma x_1 x_{1.23}$$

and $\quad \Sigma x_{1.23} x_{1.23} = \Sigma x_{1.23} (x_1 \text{-} b_{12.3} x_2 \text{-} b_{13.2} x_3) = \Sigma x_{1.23} x_1 \qquad (11)$

Also $\quad \Sigma x_{3.2} x_{1.23} = (x_3 \text{-} b32 x_2) x_{1.23} = 0$

and $\quad \Sigma_{2.3} x_{1.23} = (x_2 \text{-} b_{23} x_3) x_{1.23} = 0$

Denoting the variance of the residual $x_{1.23}$ by $\sigma^2_{1.23}$ we have,

$$n\sigma^2_{1.23} = \Sigma x^2_{1.23} = \Sigma x^2_{1.23} = \Sigma x^1 x_{1.23} \text{ by (11)}$$

$=x_1(x_1-b_{12.3}x_2-b_{13.2}x_3)$

$= n\ \sigma_1^2-nb_{12.3}\gamma_{12}\sigma_1\sigma_2-nb_{13.2}\gamma_{13}\sigma_1\sigma_3$

$$\Rightarrow\left(1-\frac{\sigma_{1.23}^2}{\sigma_1^2}\right)\sigma_1=b_{12.3}\gamma_{12}\sigma_2+b_{13.2}\gamma_{13}\sigma_3 \qquad (12)$$

Eliminating $b_{12.3}$ and $b_{13.2}$ from (10) and (12) we have

$$\begin{vmatrix} 1-\frac{\sigma_{1.23}^2}{\sigma_1^2} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & 1 & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & 1 \end{vmatrix}=0$$

$$\Rightarrow\Delta-\frac{\sigma_{1.23}^2}{\sigma_1^2}\Delta_{11}-0$$

$$\Rightarrow\sigma_{1.23}^2=\frac{\Delta}{\Delta_{11}}\sigma_1^2 \qquad (13)$$

where Δ_{11} is the co-factor γ_{11} in

$$\Delta=\begin{vmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{vmatrix}$$

Multiple co-relation coefficient.

Let us define the correlation between x_1(observed value) and X_1 (expected value) given by (8) as

$$R_{1(23)}=\frac{\Sigma x_1X_1}{\sqrt{\Sigma x_1^2\Sigma X_1^2}}$$

Again $\Sigma x_1X_1=\Sigma\{x_1(x_1-x_{1.23})\}$ since a=0

$$= \Sigma x_1^2 - \Sigma x_1 x_{1.23}$$

$$= \Sigma x_1^2 - \Sigma x_{1.23}^2 \qquad \text{by 2nd equation of (11)}$$

$$= n(\sigma_1^2 - \sigma_{1.23}^2)$$

$$\Sigma X_1^2 = \Sigma(x_1 - x_{1.23})^2$$

$$= \Sigma x_1^2 - 2\Sigma x_1 x_{1.23} + \Sigma x_{1.23}^2 = \Sigma x_1^2 - 2\Sigma x_{1.23}^2 + \Sigma x_{1.23}^2 = n(\sigma_1^2 - \sigma_{1.23}^2)$$

$$R_{1(23)} = \frac{(\sigma_1^2 - \sigma_{1.23}^2)}{\sigma_1 \sqrt{\sigma_1^2 - \sigma_{1.23}^2}}$$

$$= \frac{1}{\sigma_1}\left(\sigma_1^2 - \sigma_{1.23}^2\right)^{1/2}$$

$$= \left(1 - \frac{\sigma_{1.23}^2}{\sigma_1^2}\right)^{1/2}$$

$$\Rightarrow 1 - R_{1(23)}^2 = \frac{\sigma_{1.23}^2}{\sigma_1^2} = \frac{\Delta}{\Delta_{11}\sigma_1^2}.\sigma_1^2 = \frac{\Delta}{\Delta_{11}}$$

again

$$R_{1(23)}^2 = \frac{1}{\sigma_1^2}\left(\sigma_1^2 - \sigma_{1.23}\right)^2 \qquad (14)$$

also from (13), $\sigma_1^2 \geq \sigma_{1.23}{}^2$

(14) implies

$$R_{1(23)} \geq 0$$

$$\Rightarrow 0 \leq R_{1(23)} \leq 1$$

since $R_{1(23)}$ is non-negative.

From (14),

$$R_{1(23)}=1$$

$$\Rightarrow \sigma_{1.23}=0$$

Thus when residuals $x_{1.23}=0$ the observed and expected values of x_1 coincide and x_1 is a linear function of x_2 and x_3.

Partial correlation coefficient.

The correlation coefficient between $x_{1.3}$ and $x_{2.3}$ to between x_1 and x_2 after the linears effect of the third variable x_3 has been eliminated from both x_1 and x_2 is defined as partial correlation coefficient and is denoted by $r_{12.3}$

From (11),

$$\Sigma x_{2.3}\, x_{1.23}=0$$

$$\Rightarrow \Sigma x_{2.3}\,(x_1-b_{12.3}x_2-b_{13.2}x_3)=0$$

$$\Rightarrow \Sigma x_1 x_{2.3}-b_{12.3}\Sigma x_{2.3}-x_2-b_{13.2}\Sigma x_{2.3}x_3=0$$

$$\Rightarrow \Sigma x_{1.3}x_{2.3}-b_{12.3}\Sigma x^2_{2.3}=0$$

$$\Rightarrow b_{12.3}=\frac{\Sigma x_{1.3}x_{2.3}}{\Sigma x^2_{2.3}} \qquad (15)$$

Thus $b_{12.3}$ is the coefficient of regression of $x_{2.3}$ on $x_{2.3}$ Similarly $b_{21.3}$ is the coefficient of regression of $x_{2.3}$ on $x_{1.3}$ so that $r_{12.3}$ is given by

$$r^2_{12.3}=b_{12.3}b_{21.3}=\frac{\Delta_{12}}{\Delta_{11}}\cdot\frac{\Delta_{21}}{\Delta_{22}}=\frac{\Delta^2_{12}}{\Delta_{11}\Delta_{22}}$$

Since $r_{12.3}$ has the same sign as $b_{12.3}$ which has the sign of $-\Delta_{12}$.

$$r_{12.3}=-\frac{\Delta_{12}}{\sqrt{\Delta_{11}\Delta_{12}}}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{1-r^2_{13}}\,\sqrt{1-r^2_{23}}} \qquad (16)$$

Now,

$$r_{12.3}^2 \le 1$$

$$\Rightarrow (r_{12} - r_{13}r_{23})^2 \le (1-r_{13}^2)(1-r_{23}^2)$$

$$\Rightarrow r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}\, r_{13}\, r_{23} \le 1 \qquad \text{and } r_{31.2}$$

Illustrative examples.

Example 1 Show that the coefficient of correlation r between two variables x and y is given by

$$r = \frac{\left(\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2\right)}{2\sigma_x\sigma_y}$$

$$\text{Var } (x-y) = E[\{(x-y)-(\bar{x}-\bar{y})\}^2]$$

$$= E[\{(x-\bar{x})(y-\bar{y})\}^2]$$

$$= E(x-\bar{x})^2 - 2E(x-\bar{x})(y-\bar{y}) + E(y-\bar{y})^2$$

$$= \text{Var } x + \text{Var } y - 2 \text{ cov. } (x,y)$$

So, $\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$.

$$\Rightarrow 2r\sigma_x\sigma_y = \sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2$$

$$\Rightarrow r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y}$$

Example 2 : x_1 and x_2 are two variates with variances σ_1^2 and σ_2^2 respectively and r is the correlation coefficient between them. Determine the value of k such that $u = x_1 + kx_2$ and $v = x_1 + \frac{\sigma_1}{\sigma_2} x_2$ are uncorrelated.

Since u and v are uncorreleted Cov (u,v)=0

$$Cov(u,v)=0 \Rightarrow E\left[\left\{(x_1-\bar{x}_1)+k(x_2-\bar{x}_2)\right\}\left(x_1-\bar{x}_1+\frac{\sigma_1}{\sigma_2}(x_2-\bar{x}_2)\right)\right]=0$$

$$\Rightarrow E\left[(x_1-\bar{x}_1)^2+k\frac{\sigma_1}{\sigma_2}(x_2-\bar{x}_2)^2+\left(k+\frac{\sigma_1}{\sigma_2}\right)(x_1-\bar{x}_1)(x_2-\bar{x}_2)\right]=0$$

$$\Rightarrow \sigma_1^2+k\frac{\sigma_1}{\sigma_2}\sigma_2^2+\left(k+\frac{\sigma_1}{\sigma_2}\right)r\sigma_1\sigma_2=0$$

$$\Rightarrow \sigma_1+k\sigma_2+kr\sigma_2+r\sigma_1=0$$

$$\Rightarrow (1+r)\sigma_1+k(1+r)\sigma_2=0$$

$$\Rightarrow \sigma_1+k\sigma_2=0$$

$$\Rightarrow k=-\frac{\sigma_1}{\sigma_2}$$

Example 3 : Let us consider two regression lines $3x+2y=26$ and $6x+y=31$

Let the first line be regression of y on x. Then

$$2Y_e=-3x+26 \Rightarrow Y_e=-\frac{3}{2}x+13.$$

Hence $$b_{yx}=-\frac{3}{2}$$

Now the other line will be regression of X on Y.

$$6X_e=-y+31 \Rightarrow X_e=-\frac{1}{6}y+31/6$$

$$b_{xy}=-\frac{1}{6}$$

$$b_{xy}=\left(-\frac{3}{2}\right)\left(-\frac{1}{6}\right)=\frac{1}{4}$$

$$\Rightarrow r^2=\frac{1}{4} \Rightarrow r<1.$$

Let now consider that first line is regression of X on Y.

Then $3X_e = -2y + 26. \Rightarrow X_e = -\frac{2}{3}y + \frac{26}{3}.$

$$\therefore b_{xy} = -2/3$$

The other line will now be regression of Y on X.

$$Y_e = -6x + 31$$

$$\therefore b_{xy} = -6$$

$b_{xy} . b_{yx} = 4 \Rightarrow r > 1$ which is absurd.

Hence regression lines are not mutually reversible.

Example 4 Show that $R_{1(23)} \geq r_{12}$

We know, $\sigma^2_{1.23} = \Sigma x^2_{1.23}$

$= \Sigma x_{1.23} (x_1 - b_{12.3} x_2 b_{13.2} x_3)$

$= \Sigma x^2_{1.2} - b_{13.2} \Sigma x_{1.2} x_{3.2}$ using normal equations

$= n\sigma^2_{1.2} - b_{31.2} b_{31.2} n\sigma^2_{1.2}$

$= n\sigma^2_{1.2} (1 - b_{31.2} b_{31.2})$

$\Rightarrow \sigma^2_{1.23} = \sigma^2_1 (1 - r^2_{12})(1 - r^2_{13.2})$

From (14) we have,

$$1 - R^2_{1(23)} = (1 - r^2_{12})(1 - r^2_{13.2}) \leq 1 - r^2_{12}$$

$$\Rightarrow R_{1(23)} \geq r_{12}$$

Bivariate normal distribution :

A generalization of the normal distribution to two continuous random variate X and Y is called a bivariate normal distribution. A normal distribution of a single variable X is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{(x-\mu)}{2\sigma^2}} \quad -\infty < x < \infty$$

where μ and σ are mean and s.d respectively.

Generalising to two continuous random variables X and Y we get,

$$f(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ \frac{-\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}{2(1-\rho^2)} \right\},$$

$$-\infty < x < \infty, -\infty < y < \infty \qquad \text{(A)}$$

where μ_1, μ_2 are the means of X and Y; $\sigma_1 \sigma_2$ are the s.d. of X and Y and R is the correlation coefficient between X and Y.

Above expression gives the univariate normal distribution. If the correlation coeffieient ρ=0 then,

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{ -\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

$$= \frac{1}{\sqrt{2\pi\sigma_1}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}$$

Thus f(x,y) is the product of a function of x-alone and a function of y alone for all values of x and y. Hence X and Y are independent.

Conversely if X and Y are independent, f(x,y) given by (A) must for all values of x and y be the product of a function of x alone and a function of y- alone.

Thus in a bivariate normal distribution for random variables X and Y, the two variables will be independent if and only if their correlation coefficient is zero.

The m.g.f of the bivariate normal distribution $(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$ about the mean is given by

$$M_{XY}(t_1, t_2) = E[\exp(t_1(x-\mu_1) + t_2(y-\mu_2))]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{t_1(x-\mu_1) + t_2(y-\mu_2)\} f(x, y)dxdy$$

Let $X = x - \mu_1 - \sigma_1^2 t_1 - \rho\sigma_1\sigma_2 t_2$; $\quad Y = y - \mu_2 - \rho\sigma_1\sigma_2 t_2 - \sigma_2^2 t_2$. We find that

$$M_{XY}(t_1,t_2) = \exp\left[\frac{1}{2}\left(t_1^2\sigma_1^2 + 2t_1t_2\sigma_1\sigma_2\rho + t_2^2\sigma_2^2\right)\right] \times \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$X \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp[-\frac{1}{2(1-\rho^2)}\left\{\frac{\xi^2}{\sigma_1^2} - \frac{2\rho^2\xi}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right\} d\xi dy$$

$$= \exp\left[\frac{1}{2}\left(t^2\sigma_1^2 + 2t_1t_2\sigma_1\sigma_2\rho + t_2^2\sigma_2^2\right)\right]$$

Now μ_{rs} = coefficient of $\frac{t_1^r \, t_2^s}{\underline{|r}\underline{|s}}$ in the expansion of $M_{XY}(t_1,t_2)$.

For different values of r and s we set following results

$\mu_{20} = \sigma_1^2, \mu_{11} = \rho\sigma_1\sigma_2, \mu_{02} = \sigma_2^2$

$\mu_{30} = \mu_{21} = \mu_{12} = \mu_{02} = \sigma_2^2$

$\mu_{40} = 3\sigma_1^4, \mu_{31} = 3\rho\rho_1\sigma_2 = \mu_{04} = 3\sigma_2^4$

$\mu_{22} = (1+2\sigma^2)\sigma_1^2\sigma_2^2$

$\mu_{13} = 3\rho\sigma_1\sigma_2^3$

Cor : If $\rho=0$ in X and Y are uncorrelated then

$$M_{XY}(t_1,t_2) = \exp\left[\frac{1}{2}(t^2\sigma_1^2 + t_2^2\sigma_2^2)\right]$$

$$= M_X(t_1)M_Y(t_2)$$

$\Rightarrow$ X and Y are independent.

Conversely if X and Y are independent then $\rho = 0$

●●●

Unit 3

Principles of least squares of curve fitting.

troduction.

Let us suppose that we have m independent linear equations in n unknowns $x_1, x_2, \ldots\ldots, x_n$

$$a_{11}x_1 + a_{12}x_2 + \ldots\ldots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \ldots\ldots + a_{2n}x_n = b_2$$

$$\ldots\ldots\ \ldots\ldots\ \ldots\ldots$$

$$a_{m1}x_1 + a_{m2}x_2 + \ldots\ldots + a_{mn}x_n = b_n$$

: equivalently $\sum_{j=1}^{n} a_{ij} x_j = b_i$ $i = 1,2$——m——(1) where a's and b's are constant. If m=n then we can

ıd a unique solution of the given system of equations. If $m \neq n$, no such solution exists. We therefore try ı find those values of $x_1, x_2, \ldots\ldots, x_n$ which will satisfy (1) as nearly as possible. The method we adopt is ıe principle of least squares. The values that we obtain is the best or most possible values in the least ıuare such.

lethod of least squares.

rom (1) Let us form

$$S = \sum_{1}^{m} \left(a_{i1} x_1 + a_{i2} x_2 + \ldots\ldots + a_{in} x_n - b_i\right)^2$$

$$= \sum_{1}^{m} E_i^2 \quad \text{where } E_i = a_{i1} x_1 + a_{i2} x_2 + \ldots\ldots + a_{in} x_n - b_i$$

Again from differential calculus the extreme values of the function $f(x_1, x_2, \ldots\ldots, x_n) = f$ are given

y $$\frac{\partial f}{\partial x_i} = 0 \qquad i = 1,2, \ldots\ldots, n.$$

rovided partial derivatives exist.

Applying these results, S will have maximum of minimum for those values of x_j ($j = 1,2,\ldots\ldots,n$) which satisfy the following equations

$$\frac{\partial S}{\partial x_j} = 0, \qquad j = 1,2,3, \ldots\ldots, n.$$

$$\Rightarrow \sum_{i=1}^{m} a_{i1} E_i = 0, \quad \sum_{i=1}^{m} a_{i2} E_i = 0, \ \ldots\ldots, \ \sum_{i=1}^{m} a_{in} E_i = 0 \qquad (2)$$

Curve fitting:

Often in practice a relationship is found to exist between two variables. We want to express this relationship in mathematical form by determining an equation correcting the variables. A first step is the collection of data showing corresponding values of the variables x and y say.

A next step is to plot the points (x_1,y_1), (x_2,y_2), ...,(x_n,y_n) an a rectangular co-ordinate system. The resulting set of points is called scatter diagram. From scatter diagram it is possible to visualise a smooth curve approximating the data. Such curves are called approximate curves.

The general problem of finding equations of approximating curves to fit a given set of data is called curve fitting.

For a given value of x, say x_1, these will be a difference between the value y_1 and the corresponding value determined from the curve. C. This difference is denoted by d_1 and is called deviation, error or residuals which may be positive, negative or zero, similar results for other variables.

A measure of goodness of fit of the curve C to the sets of data is provided by $\sum_1^n d_i^2$. If this is small, the fit is good, if . is large fit is bad. Of all curves in a given family of curves approximating a set of n data points, a curve having the property $\sum_1^n d_i^2$ = minimum is called a best fitting curve of the family.

Least square line :

The least square line approximating the set of points (x_1,y_1),(x_2,y_2), ...,(x_n,y_n) has the equation

$$y = a+bx \qquad (1)$$

where a,b are constants to be determined.

The values of y on the least square line corresponding to x_1,x_2, ...,x_n are

$$a+bx_1,\ a+bx_2,\ ......,\ a+bx_n.$$

Therefore deviation from $Y_i's$ are given by

$$d_i=a+bx_i-y_i \qquad i=1,2,......,n$$

The sum of the squares of the deviations are given by

$$\sum_1^n d_i^2 = \sum_1^n (a+bx_i-y_i)^2$$

This is a function of a and b.

i.e. $f(a,b) = \sum_{1}^{n} (a + bx_i - y_i)^2$

For minimum value of f we have,

$$\frac{\partial f}{\partial a} = 0 = \frac{\partial f}{\partial b}$$

Now,

$$\frac{\partial f}{\partial a} = 0 \quad \Rightarrow \sum_{1}^{n} (a + bx_i - y_i) = 0$$

$$\Rightarrow an + b\,\Sigma x_i = \Sigma y_i$$

This can be written as

$$\left.\begin{aligned} \Sigma y &= na + b\Sigma x \\ \text{similarly} \quad \Sigma\, xy &= a\,\Sigma x + b\,\Sigma\, x^2 \end{aligned}\right\}.$$

These are the normal equations corresponding to equation (1).

Illustrative example

Example 1 . Construct a straight line that approximates the data given below

x	:	1	3	4	6	8	9	11	14
y	:	1	2	4	4	5	7	8	9

Let us construct the table

x	y	xy	x^2
1	1	1	1
3	2	6	9
4	4	16	16
6	4	24	36
8	5	40	64
9	7	63	81
11	8	88	121
14	9	126	196
$\Sigma x = 56$	$\Sigma y = 40$	$\Sigma xy = 364$	$\Sigma x^2 = 524$

The equation of the line is $y = a+b\,x$

Normal equations give.

$$\Sigma y = na+b\Sigma x \Rightarrow 40=8a+56b \Rightarrow 5=a+7b$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \Rightarrow 364=56a+524b \Rightarrow 91=14a+131b$$

Solving we get $a=\frac{6}{11}$, $b=\frac{7}{11}$.

So, the required equation is $y=\frac{6}{11}+\frac{7}{11}x$

Least square parabola :

(i) Let us extend the idea of least square line to second degree curve. Parabola is the simplest one. The least square parabola that fits a set of points isgiven by

$$y=a+bx+cx^2 \qquad (3)$$

where a,b,c are constants to be determined.

Let the sample points be $(x_1,y_1),(x_2,y_2), \ldots\ldots,(x_n,y_n)$. The values of y on the least squares parabola corresponding to $x_1,x_2, \ldots\ldots,x_n$ are $a+bx_1+cx_1^2$, $a+bx_2+cx_2^2$, ..., $a+bx_n+cx_n^2$.

Therefore deviation from Y_i's are given by

$$d_i=a+bx_i+cx_i^2-y_i \qquad i=1,2,\ldots\ldots,n$$

The sum of the squares of the deviations are given by

$$\sum_1^n d_i^2 = \sum_1^n (a+bx_i+cx_i^2-y_i)^2$$

This is a function of a,b,c, in f(a,b,c)

$$\text{i.e.} \quad f(a,b,c)=\sum_1^n (a+bx_i+(cx_i^2-y_i)^2$$

For minimum value of f we have,

$$\frac{\partial f}{\partial a}=0=\frac{\partial f}{\partial b}=\frac{\partial f}{\partial c}$$

Now,

$$\frac{\partial f}{\partial a}=0 \quad \Rightarrow \sum_1^n (a+bx_i+cx_i^2-y_i)=0$$

$$\Rightarrow an + b\Sigma x_i + c\Sigma x_i^2 = \Sigma y_i$$

This can be written as

$$\left.\begin{aligned} &\Sigma y = na + b\Sigma x + c\Sigma x^2 \\ \text{similarly}\quad &\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \text{and}\quad &\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \end{aligned}\right\} \qquad (4)$$

(4) are the normal equations corresponding to equation (3).

Example 2. Fit a least squares parabola to the data given below

x	: 1.2	1.8	3.1	4.9	5.7	7.1	8.6	9.8
y	: 4.5	5.9	7.0	7.8	7.2	6.8	4.5	2.7

Let the equation of the parabola be $y = a+bx+cx^2$

Normal equations are

$$\Sigma y = na + b\Sigma x + c\Sigma x^2$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$

$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

x	y	x^2	x^3	x^4	xy	x^2y
1.2	4.5	1.44	1.73	2.08	5.40	6.48
1.8	5.9	3.24	5.83	10.49	10.62	19.12
3.1	7.0	9.61	29.79	92.35	21.70	67.27
4.9	7.8	24.01	11765	576.48	38.22	187.28
5.7	7.2	32.49	185.19	1055.58	41.04	233.93
7.1	6.8	50.41	357.91	2541.16	48.28	342.79
8.6	4.5	73.96	636.06	5470.12	38.70	332.82
9.8	2.7	96.04	941.19	9223.66	26.46	259.31
$\Sigma x = 42.2$	$\Sigma y = 46.4$	$\Sigma x^2 = 291.20$	$\Sigma x^3 = 2275.33$	$\Sigma x^4 = 18971.92$	$\Sigma xy = 230.42$	$\Sigma x^2 y = 1449$

Since n = 8, the normal equation are

$$8a+42.2b+291.20c=46.4$$

$$42.2a+291.20b+2275.33c=230.42$$

$$291.20a+2275.33b+18971.92c=1449$$

Solving with the help of calculation we have.

$$a=2.59, \quad b=2.07, \quad c=-0.21$$

Hence the required parabola is

$$y=2.59+2.07-0.21x^2$$

(ii) Let us fit a general parabola.

The equation of such curve will be of the type

$$Y = a_0+a_1x + a_2x^2 + ...+ a_px^p \qquad (5)$$

where $a_0, a_1, a_2, ..., a_p \neq 0$ are constants to be determined.

Let the given points be (x_i, y_i) $\quad i=1,2,.........,n$.

Equation (5) is called a parabola of degree p of best fit. Substituting values of x_i in (5) we get,

$$Y_i = a_0+a_1x_i+a_2x_i^2+------+a_px_i^p$$

Y_i is called expected value of y corresponding to $x = x_i$ and y_i are called observed value.

The difference y_i-Y_i is error of estimate or residual.

$$\text{Let } F(a_i)=\sum_1^n (y_i-Y_i)^2=\sum_1^n (y_i-a_0-a_1x_i-a_2x_i^2 -----------a_px_i^p)^2$$

According to the principle of least square we choose the constants a_i($i=0,1,2, \ldots, n$) so that F has minimum value. Accordingly we get.

$$\left.\begin{aligned}
\Sigma y&=na_0+a_1\Sigma x+a_2\Sigma x^2+.........+a_p\Sigma x^p\\
\Sigma xy&=a_0\Sigma x+a_1\Sigma x^2+a_2\Sigma x^3+.........+a_p\Sigma x^{p+1}\\
\Sigma x^2y&=a_0\Sigma x^2+a_1\Sigma x^3+a_2\Sigma x^4+.........+a_p\Sigma x^{p+2}\\
\Sigma x^py&=a_0\Sigma x^p+a_1\Sigma x^{p+1}+a_2\Sigma x^{p+2}+.........+a_p\Sigma x^{2p}
\end{aligned}\right\} \qquad (6)$$

These are normal equations for fitting of a parabola of degree p.

●●●

Unit 4

Theroy of Sampling

Introduction

Often we are interested in drawing valid conclusions about a large group of individuals or objects. Instead of examining the entire group, which may be difficult to do, we examine only a small part. We do this with the aim of inferring certain facts about the group of individuals.

The entire group of individuals is known as populations. The word 'population' used in statistics has altogether a different meaning.

A population is the totality of any kind of units under consideration. It may be finite or infinite.

A sample is any portion of the population selected for study. The process of obtaining examples is called sampling.

The aim of theory of sampling is to get as many information's as possible about the population form which in sample has been drawn.

A characteristic of the population is called a parameter. For example mean μ and s.d σ are parameters.

The fundamental assumption underlying the theory of sampling is random sampling, which consists in selecting the individuals from the populations, in such a way that each individuals of the population has the same chance of being selected.

Two types of sampling will be discussed-(i) Random sampling and (2) simple sampling.

A 'Statistic' is a function of the sample. Examples are arithmetic mean $\bar{x}$ and sample variances σ^2.

Simple sampling:

Sampling of attributes may be regarded as the drawing of samples from a population whose members possesses the attribute A or not (-A). The choosing of an individual in a sampling is called an event or trial. Possession of a specified attribute A by the individual is called a success.

Simple sampling means random sampling in which each events has the same probability p of success and the probability of success is independent of the success or failure of the events in the preceding trials. Thus simple sampling is a special case of random sampling in which trials are independent and probability of success is constant.

Let us consider an example. In an urn there are two white and 3 black balls. Probability of drawing a white ball is $\frac{2}{5}$. If a white ball is drawn on the first trial, probability of drawing a white ball will be $\frac{2}{5}$. In the second trial if the ball is not replaced the probability of drawing a white ball will be $\frac{1}{4}$. Hence the sampling though random is not simple.

A random sampling from an infinite population is always a simple, since drawing of one individual cannot effect the probability. But a sampling from finite population may or may not be simple.

Let us take a simple sample of n members which will be identical with a services of n independent trials with probability of success p (constant). The probabilities of 0, 1, 2, ..., n successes are the terms of binomial expansion of $(p + q)^n$. The distribution so obtained is called sampling distribution of number of successes in the sample. The mean value and s.d. (standard error s.e) are np and $\sqrt{npq}$ respectively.

Now proportion of successes in a simple is the number successes divided by the number of members in the sample. Hence

Mean of proportion of successes = p.

s.e of proportion of successes $= \sqrt{\frac{pq}{n}}$.

Sampling with and without replacement.

Sampling where each member of the population may be chosen more than once is called a sample with replacement, while sampling where each member cannot be chosen more than once is a sample without replacement.

If we draw an object from an urn, we have the choice of replacing or not replacing the object into the urn before we draw again. In the first case a particular object can come up again and again, where as in the second case it can come up only once.

Random sampling

If in a sampling each member of the population has the same chance of being selected then it is called a random sampling.

A random sample of size n taken with replacement from a population with distribution $p(x) = P(X=x)$ is a collection of random variables $X_1, X_2, \ldots, X_n$ each having the same distribution as X and $x_1, x_2, \ldots, x_n$ are independents

Now $E(X)=\mu$ and var $(X)=\sigma^2$.

Then $E(X_1)=E(X_2)=\ldots=E(X_n)=\mu$ and $V(X_1)=V(X_2)=\ldots=Var(X_n)=\sigma^2$

Theorem 1: If μ and σ^2 are mean and variance of a population and if a random samples of size n is taken, then the sampling distribution of sample mean $\bar{x}$ has mean μ and variance σ^2

$$\text{Now } \bar{x} = \frac{x_1 + x_2 + - - - - + x_n}{n}$$

$$E(\bar{x}) = \frac{\sum_1^n E(\bar{x}_i)}{n} = \frac{n\mu}{n} = \mu,$$

$$Var(\bar{x}) = var\left(\frac{x_1 + x_2 + \ldots\ldots + x_n}{n}\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Theorem 2 : The sample mean $\overline{X}$ from a random sample of size n from a normal population with mean μ and variance σ^2 has also normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.

A random sample of size n is a collection of n independent random variables $x_1, x_2, ..., x_n$ each having same distribution as X.

The sample mean

$$\overline{X} = \frac{x_1 + x_2 + ... + x_n}{n}$$ is the sum of n independent random variables

$$\frac{x}{n}, \frac{x_2}{n}, ..., \frac{x_n}{n},$$ each of which is again normal.

Sum of a number of independent random variables is normal

$$\Rightarrow \overline{x} \text{ is normal.}$$

Hence from theorem 1, we get, $E(\overline{x}) = \mu$ and $Var(\overline{x}) = \frac{\sigma^2}{n}$.

Theorem 3. If the population is of size N, and sample is of size n and if it is without replacement then the variance of the sampling distribution of the mean is $\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

The factor $\frac{N-n}{N-1}$ is called correction factor. When N is large enough compared to n then $\frac{N-n}{N-1} \rightarrow 1$ and correction factor is 1. Variance of the sampling distribution becomes $\frac{\sigma^2}{n}$ as in the case of sampling with replacement.

Large Samples:

Suppose that a large number n of independent Bernoullian trials is performed and x successes are obtained. The hypothesis we want to test is that the probability of success in each trial is p. Assuming the hypothesis to be correct, the mean and variance of the sampling distribution of the number of success with be np and npq.

Again we know for large n,

$$z = \frac{x - np}{\sqrt{npq}}$$ is distributed as a standard normal variate.

From table $P(|z| > 3) = 0.0027$. Therefore we conclude that the hypothesis is improbable in the difference between the observed number of successes x and the expected number of successes np is highly significant.

If $|z|<1.96$ the data is consistent with the hypothesis. If $|z|>1.96$ the distance is significant.

Let the observed proportion of successes in a sample of n observation be p then $E\left(\frac{x}{n}\right)=p$

and $Var\left(\frac{x}{n}\right)=\frac{npq}{n^2}=\frac{pq}{n}$.

Thus, $$z=\frac{\frac{x}{n}-p}{\sqrt{pq/n}} \sim N(0,1)$$

This test is valid for large n only.

We shall state central limit theorem without proof.

Central limit theorem:

Let $x_1, x_2, \ldots, x_n$ be independent random variables with same mean $E(x_i)=\mu$ and same variance σ^2. Then when n is large the distribution of the sum $S_n=x_1+x_2+\ldots+x_n$ approximately normal with mean $=n\mu$ and variance $=n\sigma^2$.

Exact sampling Distributions: So far we are discussing problems of testing a number of parametric hypothesis, where the population distributions are assumed to follow certain standard forms and the hypothesis tested are on the parameters of these basic distributions. Now we shall discuss some tests which are not of these nature.

Here we shall discuss three exact sampling distributions.

The Chi-square distribution: This is a distribution of the sum of squares of independent standard normal variates.

Let $X_1, X_2 \text{------} X_n$ be n independent normal variates, each of which distributed normally with mean 'zero' and variance unity.

$$X_1^2+X_2^2+\ldots\ldots+X_n^2=\sum_{i=1}^{n}X_i^2$$

This is denoted by the Greek letter χ^2.

The probability that the random value of the variable X_i will fall in the interval dx_i is

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2} dx_i.$$

Since x_i s are independent, the probability that the values of the variates will fall simultaneously in the respective intervals dx_i $i=1,2,\ldots\ldots,n$ is

$$dp = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}x^2} dx_1 dx_2 \ldots\ldots dx_n$$

If we represent the sample by the point $E(x_1, x_2 ---- x_n)$ in n-dimensional Euclidian space then dp is the probability that E will fall in the volume element $dx_1 dx_2 \ldots\ldots dx_n$.

If X is constant, then $\sum_1^n x_i^2 = X^2$ will be a hypersphere with centre at origin and radius X. As $\sqrt{\Sigma x_i^2}$ lines between X and X+dx the point E will the between two can centril hyperspheres of radii X and X+dX and the volume of the annulus will be proportional to $d(x^n)$ in to $X^{n-1}dx$. Hence the probability that value of X from the sample will fall in the interval dx is proportional to

$$e^{-\frac{1}{2}\chi^2} \chi^{n-1} dx \qquad \text{in to} \qquad e^{-\frac{1}{2}\chi^2} \left(\frac{1}{2}\chi^2\right)^{\frac{n-2}{2}} d(\chi^2)$$

Since χ^2 lines between 0 and ∞ and integral of the Probability density function over the whole range must be unity, the probability dp is adjusted as

$$dp = \frac{1}{\sqrt{\left(\frac{n}{2}\right)}} e^{-\frac{\chi^2}{2}} \left(\frac{1}{2}\chi^2\right)^{\frac{n-2}{2}} d(\chi^2)$$

$$\Rightarrow dp = \frac{1}{2^{\frac{n}{2}} \sqrt{\frac{n}{2}}} e^{-\frac{1}{2}\chi^2} (\chi^2)^{\frac{n-2}{2}} d\chi^2 \qquad 0 < \chi^2 < \infty \qquad (1)$$

(1) is called χ^2 distribution with n degrees of freedom (d.f). The number of independent variates is the number of d.f.

Cor 1. If x_is are n independent normal variates with mean μ_i, s.d. s_i (i=1,2, ...,n) then

$$Z_i = \frac{x_i - \mu_i}{\sigma_i}$$

is a standard normal variate and all Z_i s the independent variates. Hence

$$\sum_{i=1}^{n}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 = \sum_{1}^{n} z_i^2$$

is a χ^2 distribution with n d.f.

Cor 2. Since for a normal sampling from a normal population with mean μ and s.d s, $\bar{x}$ is distributed normally about mean μ and s.d. $\frac{\sigma}{\sqrt{n}}$,

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$ is a standard normal variate.

Hence $\sqrt{n}\left(\frac{\bar{x} - \mu}{\sigma}\right)^2$ is a χ^2 variate with 1. d.f.

Properties of χ^2 distribution

1. The m.g.f of χ^2 distribution w.r.t origin is given by

$$M_{\chi^2}(t) = E[e^{t\chi^2}] = \int_0^\infty e^{t\chi^2} f(\chi^2) d\chi^2$$

$$= \frac{1}{2^{\frac{n}{2}} \left|\overline{n/2}\right.} \int_0^\infty e^{t\chi^2} e^{-\frac{1}{2}\chi^2} (\chi^2)^{\frac{n-2}{2}} d\chi^2$$

$$= \frac{1}{2^{\frac{n}{2}} \left|\overline{n/2}\right.} \int_0^\infty e^{\frac{\chi^2}{2}(1-2t)} (\chi^2)^{\frac{n-2}{2}} d\chi^2$$

$$= \frac{2^{\frac{n}{2}} \left|\frac{n}{2}\right.}{2^{\frac{n}{2}} \left|\frac{n}{2}\right. (1-2t)^{\frac{n}{2}}}$$

$$= (1-2t)^{-\frac{n}{2}} \quad \text{for} \quad |2t| < 1$$

$$= 1 + \frac{n}{2}(2t) + \frac{\left(\frac{n}{2}\right)\left(\frac{n}{2}+1\right)}{\left|2\right.}(2t)^2 + \ldots\ldots\ldots$$

μ'_r = coefficient of $\frac{t^r}{\left|r\right.}$ in the expansion of $M_{\chi^2}(t)$

Thus mean $= \mu'_1 = n$

Variance of $\chi^2 = \mu'_2 - \mu'^2_1 = n(n+2) - n^2 = 2n$

Thus $\frac{\chi^2 - n}{\sqrt{2n}}$ is a standard variate.

Additive property :

If χ_1^2 and χ_2^2 are independent χ^2 variates with d.f n_1 and n_2 respectively then $\chi_1^2 + \chi_2^2$ is a χ^2 variate with d.f $(n_1 + n_2)$.

m.g.f of $(\chi_1^2 + \chi_2^2)$ = m.g.f $\chi_1^2 \times$ m.g.f χ_2^2

$$= (1-2t)^{-n_1/2} . (1-2t)^{-n_2/2} = (1-2t)^{-\frac{n_1+n_2}{2}}$$

Hence $\chi_1^2 + \chi_2^2$ is a χ^2 variate with $(n_1 + n_2)$ d.f.

3. χ^2 Distribution tends to a normal distribution as $n \rightarrow \infty$.

$$M(t) = e^{-\frac{nt}{\sqrt{2n}}} M_0\left(\frac{t}{\sqrt{2n}}\right) = e^{-\frac{nt}{\sqrt{2n}}}\left(-\frac{2t}{\sqrt{2n}}\right)^{-n/2}$$

$$\Rightarrow \log M(t) = -\frac{nt}{\sqrt{2n}} t \log\left(1-\frac{2t}{\sqrt{2n}}\right)^{-n/2}$$

$$= -\frac{nt}{\sqrt{2n}} + \frac{n}{2}\left[+\frac{2t}{\sqrt{2n}}) + \frac{1}{2}\left(\frac{2t}{\sqrt{2n}}\right)^2 + \ldots\ldots\right] = \frac{1}{2}t^2 + 0\left(\frac{1}{n}\right)$$

As $n \rightarrow \infty$, $\log M(t) \rightarrow \frac{1}{2}t^2$

$$\Rightarrow M(t) \rightarrow e^{\frac{1}{2}t^2}$$ which is the m.g.f of standard normal variate.

Students t- distribution:

We know for large samples the statistic $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ is approximately normal with mean and s.d. σ when σ is unknown, let s be the estimate of σ. For large sample use of s as an estimate of σ does not effect the distribution of Z. But it makes a difference in case of small samples. Hence let use another statistic for small sample defined by

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{(\bar{x}-\mu)\sqrt{n}}{s} \qquad (2)$$

The distribution of this statisstic is known as 'Student's t- distribution'.

Differential form of t- distribution: We know $S^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2$

$$\Rightarrow (n-1)S^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2 = ns^2$$

$$\Rightarrow \frac{S^n}{n} = \frac{s^2}{n-1}.$$

Now

$$t = \frac{(\bar{x}-\mu)\sqrt{n}}{S}$$

$$\Rightarrow t^2 = \frac{(\bar{x}-\mu)^2(n-1)}{s^2}$$

$$\Rightarrow \frac{t^2}{n-1} = \frac{(\bar{x}-\mu)^2}{s^2} = \frac{(\bar{x}-\mu)^2 \frac{n}{\sigma^2}}{s^2 \frac{n}{\sigma^2}}$$

Now $\frac{n(\bar{x}-\mu)^2}{\sigma^2}$ is distributed as a χ^2- variate with 1 d.f and $\frac{ns^2}{\sigma^2}$ is a χ^2-variate with (n-1) d.f.

Thus $\frac{t^2}{n-1}$ is the ratio of two χ^2-variates distributed with 1 and (n-1) d.f.

Let us assume without proof

$$dp = \frac{\left(\frac{t^2}{n-1}\right)^{-1/2} d\left(\frac{t^2}{n-1}\right)}{B\left(\frac{1}{2},\frac{n-1}{2}\right)\left(1+\frac{t^2}{n-1}\right)^{\frac{n}{2}}} \qquad -\infty < t < \infty$$

$$\Rightarrow dp = \frac{2dt}{\sqrt{n-1}.B\left(\frac{1}{2},\frac{n-1}{2}\right)\left(1+\frac{t^2}{n-1}\right)^{n/2}} \qquad (3)$$

This is t-distribution with (n-1) . d.f.

Properties of t-distribution:

1.(a) Let X_i ($i=1,2,\ldots,n_1$) and x'_j ($j=1,2,\ldots,n_2$) be the values of two random samples of size n_1 and n_2 respectively from the same normal population $N(\mu,\sigma)$.

$$\overline{X}_1 = \frac{\Sigma x_i}{n_1},\ \overline{X}_2 = \frac{\Sigma x'_j}{n_2} \text{ and } S^2 = \frac{\sum_{1}^{n} (X_i - \overline{X}_1)^2 + \sum_{1}^{n_2} (x'_j - \overline{x}_2)^2}{n_1 + n_2 - 2}$$

Then the statistic

$$t = \frac{\overline{x}_1 - \overline{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ has = t-distribution with (n_1+n_2-2) d.f.

(b) Next let us suppose that x_i $(i=1,2,\ldots\ldots, n_1)$, x'_j $(j=1,2,\ldots., n_2)$ be values of two random samples from two different normal populations $N(\mu_1,\sigma)$ and $N(\mu_2,\sigma)$ where μ_1 and μ_2 are means and σ is s.d. Then the statistic

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ has a t- distribution with (n_1+n_2-2) d.f.

2. All odd moments about the origin is zero. For convenience let us replace n-1 by n in (3).

$$(3) \Rightarrow \quad dp = \frac{2dt}{\sqrt{n}\,\beta\left(\frac{1}{2},\frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

This is symmetrical about t=0 and hence all odd order moments about the origin is zero.

In particular, $\mu'_1 = 0 \Rightarrow$ mean=0. Hence all odd order central moments $\mu_{2r+1} = 0$

$$\mu_{2r} = \mu'_{2r} = 2\int_0^\infty \frac{t^{2r}\,dt}{\frac{1}{\sqrt{n}}\beta\left(\frac{1}{2},\frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

Let

$$1 + \frac{t^2}{n} = \frac{1}{y} \Rightarrow \frac{t^2}{n} = \frac{1-y}{y} \text{ and } \frac{2t\,dt}{n} = -\frac{2}{y^2}dy$$

$$\mu_{2r}=\int_0^\infty \frac{t^{2r}\left(\frac{t^2}{n}\right)^{-y_2} d\left(\frac{t^2}{n}\right)}{\beta\left(\frac{1}{2},\frac{n}{2}\right)\left(1+t^2/_n\right)^{\frac{n+1}{2}}}=\frac{n^r}{\beta\left(\frac{1}{2},\frac{n}{2}\right)}\int_0^\infty \frac{\left(\frac{t^2}{n}\right)^{r+\frac{1}{2}-1}}{\left(1+t^2/_n\right)^{\frac{n+1}{2}}} d\left(\frac{t^2}{n}\right)$$

$$\frac{dy}{dt}=-\frac{\frac{2t}{n}}{\left(1+t^2/_n\right)^2}$$

$$\mu_{2r}=2.\frac{1}{B\left(\frac{n}{2},\frac{1}{2}\right)}.\frac{1}{\sqrt{n}}\int_0^1 y^{\frac{n+1}{2}}.\frac{1}{y^2}\frac{2y^{1/2}}{2\sqrt{n}\,(1-y)^{1/2}}\frac{n^r(1-y)^2}{y^r}dy$$

$$=\frac{n^r}{B\left(\frac{n}{2},1/2\right)}\int_0^1 y^{\frac{n-2r-2}{2}}(1-y)^{r-1/2}dy$$

$$=\frac{n^r}{B\left(\frac{n}{2},1/2\right)}.B\left(\frac{n-2r}{2},r+1/2\right).$$

In particular if r =1,

$$\mu_2=n.\frac{B(\frac{n}{2}-1,\frac{3}{2})}{B\left(\frac{n}{2},\frac{1}{2}\right)}=n.\frac{\Gamma\left(\frac{n}{2}-1\right)\Gamma\left(3/2\right)}{\Gamma\left(\frac{n+1}{2}\right)}.\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(n/2\right)\Gamma\left(1/2\right)}$$

$$=n\frac{1}{2}.\frac{\Gamma\left(n/_2-1\right)\Gamma\left(1/2\right)}{\Gamma\left(1/2\right)\left(\frac{2}{2}-1\right)\Gamma\left(\frac{n}{2}-1\right)}=\frac{n}{2}.\frac{2}{n-2}=\frac{n}{n-2}$$

Similarly, $\mu_4 = \frac{3x^2}{(n-2)(n-4)}$

Hence $\beta_1 = 0,$ $\beta_2 = \frac{\mu_4}{\mu_2^2} = 3\left(\frac{n-2}{n-4}\right) \to 3$ as $n \to \infty$

Thus for large degrees of freedom t-distribution tends to normal distribution.

F - distribution

Let x_{1i} ($i=1,2,......,n_1$) and x_{2j} ($j=1,2,......,n_2$) be the values of two independent random samples drawn from the same normal population with variance σ_1^2 and σ_2^2.

Let $\bar{x}_1$ and $\bar{x}_2$ be sample means of the two samples.

$$\text{Let } S_1^2 = \frac{1}{n_1-1}\sum_1^{n_1}(x_{1i}-\bar{x}_1)^2 \quad \text{and} \quad S_2^2 = \frac{1}{n_2-1}\sum_1^{n_2}(x_{2j}-\bar{x}_2)^2$$

Then we define the statistic F by the relation

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} (S_1^2 > S_2^2) = \frac{n_2-1}{n_1-1} \cdot \frac{\frac{1}{\sigma_1^2}\left[\sum_1^{n_1}(x_{1i}-\bar{x}_1)^2\right]}{\frac{1}{\sigma_2^2}\left[\sum_1^{n_2}(x_{1j}-\bar{x}_2)^2\right]} \tag{4}$$

This is known as F-distribution with n_1-1, n_2-1 d.f.

If we assume that the two samples are drawn from the normal population with equal variance then we define

$$F = \frac{S_1^2}{S_2^2}$$

$$\frac{\upsilon_1 F}{\upsilon_2} = \frac{\frac{1}{\sigma^2}\left[(n_1-1)S_1^2\right]}{\frac{1}{\sigma^2}\left[(n_2-1)S_2^2\right]} \text{ where } \upsilon_1 = n_1-1, \upsilon_2 = n_2-1$$

The numerator and denominator of he 2nd number are independent χ^2-variates with υ_1 and υ_2 d.f

Hence $\frac{\upsilon_1 F}{\upsilon_2}$ is a $\beta\left(\frac{\upsilon_1}{2}, \frac{\upsilon_2}{2}\right)$ variate so that the probability that a random value of F will in the interval dF is

$$dp = \frac{\upsilon_1^{\upsilon_1/2}\, \upsilon_2^{-\upsilon_2/2}\, F^{\frac{\upsilon_1-2}{2}}}{B\left(\frac{\upsilon_1}{2}, \frac{\upsilon_2}{2}\right)(\upsilon_1 F + \upsilon_2)^{\frac{1}{2}(\upsilon_1+\upsilon_2)}}\, dF \qquad (5)$$

This distribution is called the distribution of the variance ratio F with υ_1 and υ_2 d.f.

Fisher's Z- distribution.

Putting $F=e^{2z}$ in (5) we get.

$$dp = \frac{2\upsilon_1^{\upsilon_1/2}\, \upsilon_2^{\upsilon_2/2}\, e^{\upsilon_1 z}}{B\left(\frac{\upsilon_1}{2}, \frac{\upsilon_2}{2}\right)(\upsilon_1 e^{2z} + \upsilon_2)^{\frac{1}{2}(\upsilon_1+\upsilon_2)}}\, dz$$

This distribution is known as Fisher's z-distribution and is more nearly symmetrical than F-distribution.

Test of Significance based on t, F and χ^2 distributions

Application of t-distribution:

(a) For a given random sample from a normal population we are to test the hypothesis "mean of the population is μ". We assume in hypothesis to be correct and then see whether the given data is consistent with hypothesis or not.

To carry out the test we calculate in statistic 't'.

$$t = \frac{(\bar{x}-\mu)\sqrt{n}}{S} \text{ where } s^2 = \frac{1}{n-1}\sum_1^n (x_i-\bar{x})^2$$

For different values of d.f.(υ)the table gives the value of $t_{0.05}$ and $t_{0.01}$ which are defined by

$P[\,|\,t\,| \geq t_{0.05}]=.05$ and $P[\,|\,t\,| \geq t_{0.01}]=.01$

If calculated value of t exceed s $t_{0.05}$ the difference between $\bar{x}$ and μ is significant and if it exceeds $t_{0.01}$ the difference is highly significant.

(b) Given two independent random samples $x_{1i}(i=1,2,......,n_1)$ and $x_{2j}(i=1,2,......,n_2)$ with means $\bar{x}_1$ and $\bar{x}_2$ and s.d. σ_1 and σ_2 from normal populations with same variance we are going to lest the hypothesis "the population means are same".

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad \text{where} \quad S^2 \; \frac{1}{n_1+n_2-2}[(n_1-1)\sigma_1^2+(n_2-1)\sigma_2^2]$$

$$=\frac{1}{n_1+n_2-2}\left[\sum_{i=1}^{n_1}(x_{1i}-\bar{x}_1)^2+\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2\right]$$

This statistic follows t-distribution with n_1+n_2-2.d.f.If the calculated value of $|t|> t_{0.05}$ ($t_{0.01}$) the difference between the sample means is said to be significant (highly significant); otherwise they are said to be consistent with the hypothesis.

Example1. A census of retail stores in a particular year game the mean annual turn over of Rs. 1,50,000. A random sample of 10 stores in the following year game the mean turnover of Rs. 1,57,000 and a s.d. of Rs. 11,180. Has the mean tarn over changed in the following year?

Hypothesis : The mean turn over has changed in the following year.

m=1,50,000 $\quad \bar{x}$=1,57,000 $\quad$ S=1,118 as n =10

$$t=\frac{(\bar{x}-\mu)\sqrt{n}}{S}=\frac{(7000)\sqrt{10}}{1,118}=1.97$$

From t-table for (10-1)=9 d.f. $t_{0.05}$ = 2,262. The calculated value of t is less than table value. Hence there is no evidence that the mean turn over is changed in the following year 1.

Example 2. For a random sample of 10 pigs fed on diet A, the increase in weight in gun for a certain period were 10,6,16,17,13,8,14,15,9 gun

For another random sample of 12 pigs fed on diet B the increase in weight in the same period were 7,13,22,15,12,14,18,8,21,23,10,17, gun.

Test whether diets A and B differ significantly as regard the effect on increases in weight.

Hypothesis : Diet A and B differ significantly.

$$t=\frac{\bar{x}_1-\bar{x}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad \text{where} \quad S^2=\frac{1}{n_1+n_2-2}[\Sigma\Sigma(_{1i}-\bar{x}_1)^2+(x_{2j}-\bar{x}_2)^2]$$

Diet to A, Gain in wt = x_{1i}, Diet to B Gain in wt = x_{2j}

Diet A			Diet B.		
x_{1i}	$x_{11}-\bar{x}_1$	$(x_{1i}-\bar{x}_1)^2$	x_{2j}	$x_{2j}-\bar{x}_2$	$(x_{2j}-\bar{x}_2)^2$
10	-2	4	7	-8	64
6	-6	36	13	-2	4
16	4	16	22	7	49
17	5	25	15	0	0
13	1	1	12	-3	9
12	0	0	14	-1	1
8	-4	16	18	3	9
14	2	4	8	-7	49
15	3	9	21	6	36
9	-3	9	23	8	64
			10	-5	25
			17	2	4

$$\bar{x}_1=\frac{120}{10}=12,\ \bar{x}_2=\frac{180}{12}=15 \quad S=\sqrt{\frac{120+314}{10+12-2}}=\sqrt{\frac{434}{20}}=\sqrt{21.7}=4.65$$

$$t=\frac{15-12}{4.65\sqrt{\frac{1}{10}+\frac{1}{12}}}=\frac{3}{4.65\sqrt{\frac{11}{60}}}=1.6$$

From table value of $t_{0.05}$ for (10+12-2=20 d.f is 2.09. Hence difference of the sample means are not significant.

Application of F- distribution in testing hypothesis.

Given two independent random samples from normal populations we are to test the hypothesis "the population variances are same".

$$F = \frac{s_1^2}{s_2^2} \qquad S_1^2 = \frac{1}{n_1 - 1} \sum_1^{n_1} (x_{1i} - \bar{x}_1)^2 \text{ and } S_2^2 = \frac{1}{n_2 - 1} \sum_1^{n_2} (x_{2j} - \bar{x}_2)^2.$$

and $S_1^2 > S_2^2$ confirms to the F distribution with $n_1 - 1$ and $n_2 - 1$ d.f.

Example 3. In two groups of 10 children each the increases in weight due to two different diets in the same period were in gun.

x_{1i}:	8	5	7	8	3	2	7	6	5	7
x_{2j}:	3	7	5	6	5	4	4	5	3	6

Find whether the variances are significantly different.

Hypothesis : The variances in the mean are same.

$\bar{x}_1 = 5.8 \qquad \bar{x}_2 = 4.8 \qquad n_1 = n_2 = 10$

Deiet 1			Diet 2.		
x_{1i}	$x_{1i} - \bar{x}_1$	$(x_{1i} - \bar{x}_1)^2$	x_{2j}	$x_{2j} - \bar{x}_2$	$(x_{2j} - \bar{x}_2)^2$
8	2.2	4.84	3	1.8	3.24
5	-.8	.64	7	2.2	4.84
7	1.2	1.44	5	.2	.04
8	2.2	4.84	6	1.2	1.44
3	-2.8	7.84	5	.2	.04
2	-3.8	14.44	4	-.8	.64
7	1.2	.04	4	-.8	.64
6	.2	.04	5	.2	.04
5	-.8	.64	3	-1.8	3.24
7	1.2	1.44	6	1.2	1.44
58		36.2	48		15.6

$S_1^2 = 4.02 \qquad S_2^2 = 1.73 \qquad F = 2.32$

also $F_{0.05}$ for 9 and 9 d.f is less than 3.07. Hence the calculate value is not significant.

Application of χ^2 in testing hypothesis

Let X_i, i=1,2,----n be a random sample from a normal population $N(\mu,\sigma)$ and $\bar{x} = \frac{\Sigma xi}{n}$ be sample mean and $s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$ be sample variance.

$$\text{Then } \chi^2 = \frac{\Sigma(x-\bar{x})^2}{\sigma^2}$$

Example 4: In 200 tosses of a coin, 115 heads and 85 tails are observed. Test the hypothesis that the coin is fair. x_1=115 and x_2=85

If the coin is fair the expected frequencies of heads and tails are up_1=100 and up_2=100.

$$\chi^2 = \frac{\Sigma(x-\bar{x})}{np_1} = \frac{(115-100)^2}{100} = \frac{(85-100)^2}{100} = 2.25+2.25=4.5$$

The table value of χ^2 at 5% level for 1 d.f is 3.84 since 4.5 >3.84 we reject the hypothesis. Thus coin is biased.

Example-5 A die is thrown 90 times and the number of faces shown are as indicated beloe.

Face	:	1	2	3	4	5	6
Frequency	:	18	14	13	15	14	16

Is the die a fair one.

Hypothesis : Dic is fair.

Probability of showing one face = $\frac{1}{6}$.

$$\bar{x} = np_i = 90.\frac{1}{6} = 15$$

$$\chi^2 = \frac{(18-15)^2}{15} + \frac{(14-15)^2}{15} + \frac{(13-15)^2}{15} + \frac{(15-15)^2}{15} + \frac{(14-15)^2}{15} + \frac{(16-15)^2}{15}$$

$$= \frac{9}{15} + \frac{1}{15} + \frac{4}{15} + \frac{1}{15} + \frac{1}{15} = 16/15 = 1.07$$

Table value of χ^2 at 5% level for 5 d.f is 1.14. Hence dic is fair.

●●●

Unit 5
Estimation

Introduction

The principal objectives of statistical analysis is to draw inference about the population. The inference to be drawn relates to some parameters of the population. In absence of complete information about the population, it would not be possible to determine the true value of the parameter. So we try to obtain from the sample data an estimate of the exact value of the parameter or an interval of values in which parameter lies.

Given a random sample $x_1, x_2, \ldots\ldots, x_n$ of size u from a population with probability density function (pdf) of known from $f(x,\theta)$ with unknown parameter θ our problem is to find an estimate of θ in terms of sample values.

Fishers Criteria for best Estimation

Any statistic used to estimate a parameter is called an estimator. Its specific value is an estimate.

According to Fisher an estimator is said to be the best if it is (i) unbiased (ii) consistent (iii) efficient and (iv) sufficient.

(i) unbiased : An estimator $t_n = t(x_1, x_2, \ldots\ldots, x_n)$ drawn from a sample of size n is said to be unbiased estimator of a population parameter θ if $E(t_n) = \theta$

If a random sample $(x_1, x_2, \ldots\ldots, x_n)$ of size n is drawn from a normal population with mean μ and s.d σ then $E(\overline{x}) = \mu$ and $\text{var}(S^2) = \sigma^2$ where $S^2 = \frac{1}{n-1}\sum_1^n (x_i - \overline{x})^2$

Thus $\overline{x}$ and S^2 are unbiased estimates of the population mean μ and s.d σ^2 . $E(s^2) \neq \sigma^2$, the sample variance is biased estimate. The quantity $E(t_n)-\theta$ is called the bias of the estimate θ.

(ii) Consistent : An estimator $t_n = (x_1, x_2 \ldots\ldots, x_n)$ drawn from a random sample of n values is said to be consistent estimator for a population parameter θ, if it converges in probability to θ as as $n \to \infty$. Thus for given $\varepsilon > 0, y > 0$ we can find N depending ε, y on such that

$P[\,|t_n - \theta| < \varepsilon\,] > 1 - y$ for all $n > N$.

$$\Rightarrow p \lim_{n\to\infty} t_n = \theta$$

(iii) Efficient Estimates: Let for large samples two consistent estimators t_n and t'_n be both distributed asymptotically normally about the true value of the parameter θ with variance σ^2/n and $\frac{\sigma'^2}{n}$ respectively. This will always be the case due to central limit theorem. Then t_n is said to be more efficient than t'_n if $\sigma^2 < \sigma'^2$.

If we can find a consistent estimate t_n whose variance is less than that of all other consistent estimator for all n then t_n will be said to be most efficient. The efficiency E is defined as the ratio of the variance of the most efficient estimator to the variance of the given estimator.

(iv) Sufficient estimator : An estimator t_n is said to be sufficient for estimating a population parameter θ, if it contains all the information in the samples about the parameter.

If f(x,θ) is density functions for a population, then the likely-hood function $L(x_1,x_2 \text{-----} x_n;\theta)$ for random sample $x_1,x_2, \ldots, x_n$ is defined by

$$L(x_1,x_2,\ldots,x_n;\theta) = \prod_{i=1}^{n} f(x_i,\theta) = f(x_1,\theta) = f(x_2,\theta) = \ldots = f(x_n,\theta)$$ If it is possible to write

$$L(x_1,x_2,\ldots,x_n;\theta) = L_1(t_n,\theta)L_2(x_1,x_2,\ldots,x_n)$$ then t_n will be said to be sufficient for θ.

There are two types of estimation : 1. Point estimation 2. Interval estimation.

Point estimation

An estimate of a population parameter given by a single number is called point estimate of the parameter.

Let $x_1,x_2,\ldots,x_n$ be a sample drawn from a population and the unknown parameter be θ (if may be $\bar{x}, \sigma^2$ etc). The point estimation of θ will be based on the sample observations $X_i(i,1,2,\ldots,n)$. It will be a function of these observations and a statistic. The statistic to be used for point estimation of θ is called point estimator and is denoted by $\hat{\theta}$. When an actual set of sample values is given we can compute a numerical value which is the point estimate of $\hat{\theta}$.

Since X_i s are identically and independently distributed random variable having the same distribution as population, the point estimator $\hat{\theta}$ will also be a random variable and hence it will have a probability distribution.

Interval Estimation :

Let mean and variance of the population be μ and σ^2 respectively where μ is unknown. We are to estimate μ.

$$E(\overline{x}) = \mu, \qquad \sigma_{\overline{x}} = s.d(\overline{x}) = \frac{\sigma}{\sqrt{n}}$$

If the population is normal, $\overline{x}$ will be approximately normal. We have from table,

$$P\left\{\mu - \frac{2\sigma}{\sqrt{n}} < \overline{x} < \mu + \frac{2\sigma}{\sqrt{n}}\right\} = 0.954 \Rightarrow P\left\{\mu - \frac{2\sigma}{\sqrt{n}} < \overline{x} < \mu + \frac{2\sigma}{\sqrt{n}}\right\} = 0.954$$

Thus we can find the probability of difference of $\overline{x} - \mu$.

An estimation of a population parameter given by two numbers between which the parameter may be considered to lie is called an interval estimate of the parameter.

Let μ_S and σ_S be mean and s.d of the sampling distribution of a statistic S. If the sampling distribution is approximately normal we can expect to find s lying in the intervals $\mu_S - k\sigma_S$ to $\mu_s + k\sigma_S$ where k depends on particular level of confidence.

For large samples the distribution of the sample mean $\overline{x}$ is approximately normal with mean μ_S and s.d $\frac{\sigma_S}{\sqrt{n}}$. Then $P\left\{-\frac{2\sigma_s}{\sqrt{n}} < \overline{x} - \mu_s < \frac{2\sigma_s}{\sqrt{n}}\right\} = 0.93$ at 5% level.

The inequality $\overline{X} - \mu_s < \frac{2\sigma_s}{\sqrt{n}}$ is equivalent to $\mu_s > \overline{X} - \frac{2\sigma_s}{\sqrt{n}}$. and the inequality

$$-\frac{2\sigma_s}{\sqrt{n}} < \overline{X} - \mu_s \text{ is equivalent to } \mu_s < \overline{x} + \frac{2\sigma_s}{\sqrt{n}}.$$

The true value of μ_s these in the interval $\left(\overline{X} - \frac{2\sigma_s}{\sqrt{n}}, \overline{X} + \frac{2\sigma_s}{\sqrt{n}}\right)$ with probability 0.95.

Maximum Likelihood Estimates.

Although confidence limits are valuable for estimating a population parameter, it is often connivent to have a point estimate. To obtain a best such estimated, Fisher employ a technique known as maximum likelihood method.

In this method it is assumed that the population has a density function that contains the population parameter θ which is to be estimated by a certain statistic. Let the density function be $f(x,\theta)$. If there are n observation $x_1, x_2, \ldots, x_n$ the joint density function will be $L = f(x_1,\theta)\ f(x_2,\theta) \ldots\ldots f(x_n,\theta)$, L is called the likelihood (a) Let mean μ be unknown and s.d σ be known.

$$\text{Since } f(x_n, \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_k - \mu)^2/2} \text{ (normal distribution)}$$

we have, $L=(2\pi\sigma^2)^{-n/2}\sum_{1}^{n} e^{-(x_k-\mu)^2/2\sigma^2}$

$$\Rightarrow \log L = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{1}^{n}(x_k-\mu)^2 \qquad (1)$$

Taking partial derivatives w.r.t μ, $\frac{1}{L}\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2}\Sigma(x_k-\mu)$

For maximum, $\frac{\partial L}{\partial \mu} = 0 \Rightarrow \sum_{1}^{n} x_k - n\mu = 0 \Rightarrow \mu = \frac{\Sigma x_k}{n}$

Thus maximum likelihood estimate is the sample mean.

(b) Let μ be known and σ^2 be unknown. Then differentiating (1) partially w.r.t. σ^2,

$$\frac{1}{L}\cdot\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{1}^{n}(x_k-\mu)^2$$

$$\frac{\partial L}{\partial \sigma^2} = 0 \Rightarrow \sigma^2 = \frac{\sum_{1}^{n}(x_k-\mu)^2}{n}$$

Cramer - Rao in equality

Let X be a continuous random variable with p.d.f $f(x;\theta)$ and Likelihood function L. Let t be an unbiased estimates of some function θ ($\tau(\theta)$). Then

$$\text{Var } t \geq \frac{\{\tau'(\theta)\}^2}{E\left[\left(\frac{\partial}{\partial\theta}\log L\right)^2\right]} \qquad \tau'(\theta) = \frac{d}{d\theta}\tau(\theta).$$

Hypothesis is $\tau(\theta)=E(t)=\int t\, L dx$. where $L = \prod_{1}^{n} f(x_i, \theta)$.

and $\int f(x_1,\theta) f(x_2,\theta) \ldots\ldots\ldots f(x_n,\theta) dx_1 dx_2 \ldots\ldots dx_n = \int dx$

$$\therefore \tau(\theta) = \int t\left(\frac{\partial}{\partial\theta.}\log L\right) L\,dx \;= \int\{t-\tau(\theta)\}\left(\frac{\partial}{\partial\theta}\log L\right) L\,dx.$$

By Cauchy-schwerz inequality,

$$[\tau'(\theta)]^2 \leq \int\{t-\tau(\theta)^2 L dx. \int\left(\frac{\partial}{\partial\theta}\log L)^2\right) L dx$$

$$\Rightarrow \text{Var } t = E\{t-\tau(\theta)\}^2 /[\tau[\tau(\theta)^2 \geq E\left[\frac{\partial}{\partial\theta}(\log L)]^2\right]$$

Var t is called maximum variance bound for estimation of $\tau(\theta)$

Cor If $t = 0$ then $\tau\theta = \tau(\theta)$ and $\tau'\theta = 1$

$$\text{Var}(t) \geq \frac{1}{E\left(\frac{\partial}{\partial\theta}\log L\right)}$$

Illustrative examples.

Example 1. Show that in a random sampling from a normal population the sample mean is a consistent estimator for the population mean.

Let $x_i(i=1,2\text{------}u)$ be the sample drawn from a normal population with mean μ and s.d.σ.

Then the statistic $z = \frac{(\bar{x}-\mu)\sqrt{n}}{\sigma}$ is a standard normal variate.

$$\therefore P[\,|\bar{x}-\mu| < \varepsilon] = P\left[|z| < \frac{\varepsilon\sqrt{n}}{\sigma}\right] = \int_{-\frac{\sqrt{n}\varepsilon}{\sigma}}^{\frac{\sqrt{n}\varepsilon}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\,dz$$

Hence for given y>0 we can choose n so that the area under the standard normal curve between $-\frac{\sqrt{n}\varepsilon}{\sigma}$ and $\frac{\sqrt{n}\varepsilon}{\sigma}$ becomes greater that 1-y.

Example 2. Measurement of the diameters of a random sample of 200 balls bearing made by certain machine during one week showed mean of 0.824 and s.d of 0.042. Found 95% confidence limit for the mean diameter of all the ball bearings.

Since n=200 is large we can assume that $\bar{x}$ is approximately normal.

the 95 % confidence limits are

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = 0.824 \pm 1.96\frac{0.042}{\sqrt{200}}$$

$=0.824 \pm 0.0058 = 0.824 \pm 0.006$. Hence the interval is (0816,0.830)

Example 3 A diary produces a certain milk product which is marketed in tin containers. It is noted that means of the contents of tins very from batch to batch bur s.d is constant at σ=0.10gm. A sample of 25 tins from a batch is taken to find an accurate estimate of the mean of the batch. How accurate is the sample mean as a point estimate of the batea mean μ?

As the population is normal, the sample mean $\bar{x}$ will be notmal.and s.d of $\bar{x}$ will be Normal and s.d of $\bar{x}$ will be

$$s.d(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{0.10}{\sqrt{25}} = 0.02 \text{ gm}$$

Again $P\{-2 \times 0.02 < \bar{x} - \mu < 2 \times 0.02\} = 0.95$

$\Rightarrow P\{-0.04 < \bar{x} - \mu < 0.04\} = 0.95$

Thus the probability is 0.95, that the difference $\bar{x}-\mu$ between sample mean and population mean will not exceed 0.04.

Hence the manager can be confident that 95% of cases the sample mean will be within 0.04 gm of the population mean μ.

Example 4. A random sample $x_i(i=1,2,$----$u)$ is drawn from the exponential population density function

$$f(x_o \propto, p) = y_o e^{-\beta(x-\propto)} \qquad \propto \leq n \leq \infty, \beta > 0 \qquad y_o \text{ being constant.}$$

Find maximum likelihood estimator for $\propto$ and β.

To calculate y_o we have

$$y_0 \int_{\propto}^{\infty} e^{-\beta(x-\propto)} dx = 1 \Rightarrow y_0 \int_0^{\infty} e^{-\beta z} dz = 1 \qquad x - \propto = z$$

$$\Rightarrow -y_0 \frac{e^{-\beta z}}{\beta}]_0^{\infty} = 1 \Rightarrow y_0 = \beta$$

$$\text{Now} \quad L = \prod_1^n f(x; \propto, \beta) = \beta^n e^{-\beta \sum_1^n (x_i - \propto)} \quad \text{----(2)}$$

$$\Rightarrow \log L = u \log \beta - u\beta(\bar{x} - \propto)$$

Differentiating w.r.t $\propto$ and β we get,

$$0=\frac{\partial}{\partial\propto}\log L=u\beta.\text{ and }0=\frac{\partial}{\partial\propto}\log L=\frac{u}{\beta}-u(\bar{x}-\propto)$$

$u\beta=0\Rightarrow\beta=0$ which is impossibal since $\beta>0$

$$\frac{u}{\beta}-u(\bar{x}-\propto)=0\Rightarrow\beta=\frac{1}{\bar{x}-\propto}.$$ This fails to determine finite x.

From (2) for all $\beta>0$, L will be maximum when $\propto$ tanes its largest possible value. From definition, since population range is from $\propto$ to ∞, $\propto$ must be less than or equal to every member of the sample. Then

$$\hat{\propto}=x_1 \text{ and } \hat{\beta}=\frac{1}{\bar{x}-x_1}$$

Example 5 Find maximum likelihood estimator of the parameters of a $N(\mu,\sigma^2)$ population based on the random sample $x_i(i=1,2,\text{-----}u)$ (i) σ^2 is given (ii) μ is given.

$$L(\mu\sigma^2)=\left(\frac{1}{\sqrt{2\Pi}\,\sigma}\right)e^{-\frac{2}{2\sigma^2}\sum_1^n(x_i-\mu)^2}\Rightarrow\log L=\text{constent}-\frac{1}{2\sigma^2}\sum_1^n(x_i-\mu)^2$$

(i) $$\frac{\partial}{\partial\mu}\log L=0\Rightarrow\frac{1}{\sigma^2}\sum_1^n(x_i-\mu)=0\Rightarrow\hat{\mu}=\frac{\Sigma x_i}{n}=\bar{x}$$

Thus $\bar{x}$ is the maximum likelihood estimator of μ.

(ii) $$\frac{\partial}{\partial\sigma^2}=0\Rightarrow-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_1^n(x_i-\mu)^2=0\Rightarrow\sigma^2=\frac{1}{n}\sum_1^n(x_i-\mu)^2.$$

which is the maximum likely-hood estimator.

●●●