ALAGAPPA UNIVERSITY

[Accredited with 'A+' Grade by NAAC (CGPA:3.64) in the Third Cycle
and Graded as Category–I University by MHRD-UGC]

(A State University Established by the Government of Tamil Nadu)

KARAIKUDI – 630 003

Directorate of Distance Education

M.Sc. [Mathematics]

IV - Semester

311 44

PROBABILITY AND STATISTICS

Authors:
J.S. Chandan, Prof. Medgar Evers College, City University of New York, New York
Units (1.2.4-1.2.6, 1.3.1, 1.4.1-1.9, 5.6, 8.3, 9.4-9.10, 13.3-13.8)
V K Khanna, Formerly Associate Professors, Department of Mathematics, Kirori Mal College, University of Delhi
S K Bhambri, Formerly Associate Professors, Department of Mathematics, Kirori Mal College, University of Delhi
Units (1.2.7-1.2.8, 1.3, 1.3.3, 1.4)
G.S Monga, Dean, Invertis Institute of Management Studies, Bareilly
Units (2, 3.3, 4.4-4.9, 5.5.3, 11.0-11.2)
Late C.R. Kothari, Former Principal, University Commerce College, Jaipur, and Associate Professor in the Department of Economic Administration and Financial Management (EAFM), University of Rajasthan, Jaipur
Units (3.0-3.2, 4.3, 5.0-5.5.2, 5.7-5.11, 6.0-6.4, 6.6-6.10, 7.3-7.4, 8.0-8.2, 9.0-9.3, 10.3, 11.3)
Kalika Patrai, Senior Lecturer in Galgotia Institute of Management and Technology, Greater Noida
Units (3.4-3.9, 6.5, 7.5-7.10, 13.0-13.2, 14)
S Kalavathy, Dean, Department of Science and Humanities, K Ramakrishnan College of Technology, Samayapuram, Trichy
Units (8.4-8.9)
Vikas® Publishing House: Units (1.0-1.2.3, 1.3.2, 4.0-4.2, 7.0-7.2, 10.0-10.2, 10.4-10.8, 11.4-11.8, 12)

VIKAS®

Vikas® is the registered trademark of Vikas® Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: A-27, 2nd Floor, Mohan Co-operative Industrial Estate, New Delhi 1100 44
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

Work Order No. AU/DDE/DE-12-02/Preparation and Printing of Course Materials/2020 Dated 30.01.2020 Copies - 1000

SYLLABI-BOOK MAPPING TABLE

Probability and Statistics

Syllabi	Mapping in Book
BLOCK - I: PROBABILITY AND MULTIVARIATE DISTRIBUTIONS **UNIT - I Probability and Distribution:** Introduction - Set Theory - The Probability Set Function - Conditional Probability and Independence. **UNIT - II** Random Variables of the Discrete Type - Random Variables of the Continuous Type. **UNIT - III** Properties of the Distribution Function - Expectation of Random Variable - Some Special Expectations - Chebyshev's Inequality. **UNIT - IV Multivariate Distributions:** Distributions of Two Random Variables -Conditional Distributions and Expectations.	**Unit 1:** Probability and Set Theory **(Pages 1-60);** **Unit 2:** Random Variables of Discrete and Continuous Type **(Pages 61-74);** **Unit 3:** Expectation of Random Variables **(Pages 75-94);** **Unit 4:** Distribution of Random Variables **(Pages 95-114)**
BLOCK - II: CORRELATION COEFFICIENT, SPECIAL DISTRIBUTIONS AND DISTRIBUTION **UNIT - V** The Correlation Coefficient - Independent Random Variables - Extension to Several Random Variables. **UNIT - VI Some Special Distributions:** The Binomial and Related Distributions - The Poisson Distribution. **UNIT - VII** The Gamma and Chi-Square Distributions - The Normal Distribution - The Bivariate Normal Distribution. **UNIT-VIII Distributions of Functions of Random Variables:** Sampling Theory - Transformations of Variables of the Discrete Type.	**Unit 5:** Correlation **(Pages 115-132);** **Unit 6:** Distributions **(Pages 133-157);** **Unit 7:** Special Distributions **(Pages 158-191);** **Unit 8:** Distribution of Functions of Random Variable and Sampling **(Pages 192-225)**
BLOCK - III: BETA, t, F, X AND ns^2/σ^2 DISTRIBUTIONS **UNIT-IX** Transformations of Variables of the Continuous Type - the Beta, t and F Distributions - Extensions of the Change - of - Variable Technique. **UNIT-X** Distributions of Order Statistics - The Moment Generating - Function, Techniques. **UNIT-XI** The Distributions of X and ns^2/σ^2 - Expectations of Functions of Random Variables.	**Unit 9:** Transformation of Variable of the Continuous Type **(Pages 226-256);** **Unit 10:** Distributions of Order Statistics **(Pages 257-267);** **Unit 11:** Distribution of X and ns^2/σ^2 **(Pages 268-276)**
BLOCK - IV: LIMITING DISTRIBUTIONS **UNIT-XII Limiting Distributions:** Convergence in Distribution - Convergence in Probability. **UNIT-XIII** Limiting Moment Generating Functions - The Central Limit Theorem. **UNIT-XIV** Some Theorems on Limiting Distributions.	**Unit 12:** Limiting Distributions and Convergence **(Pages 277-287);** **Unit 13:** Limiting Moment Generating Function and Central Limit Theorem **(Pages 288-296);** **Unit 14:** Some Theorems on Limiting Distributions **(Pages 297-302)**

CONTENTS

INTRODUCTION

NOTES

Probability is the branch of mathematics concerning numerical descriptions of how likely an event is to occur or how likely it is that a proposition is true. Probability is simply how likely something is to happen. The probability of an event is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility of the event and 1 indicates certainty. The higher the probability of an event, the more likely it is that the event will occur. These concepts have been given an axiomatic mathematical formalization in probability theory, which is used widely in the areas of mathematics, statistics, finance, gambling, science (in particular physics), artificial intelligence/machine learning, computer science, game theory, and philosophy to draw inferences about the expected frequency of events. Probability theory is also used to describe the underlying mechanics and regularities of complex systems.

'Probability' and 'Statistics' are fundamentally related. The probability theory describes statistical phenomenon and analyses them to study correlation and regression, sampling methods, business decisions, statistical inferences, etc. Probability is considered as theory of chance whereas statistics is considered as a mathematical science pertaining to the collection, analysis, interpretation or explanation, and presentation of data. Statistical analysis is very important for taking decisions and is widely used by academic institutions, the natural and social sciences departments, the government and business organizations.

The word 'Statistics' is derived from the Latin word '*status*' which means a political state or government. It was originally applied in connection with kings and monarchs collecting data on their citizenry which pertained to state wealth, collection of taxes, study of population, and so on. The subject of statistics is primarily concerned with making decisions about various disciplines of market and employment, such as stock market trends, unemployment rates in various sectors of industry, demographic shifts, interest rate and inflation rate over the years, and so on. To a layman, it often refers to a column of figures, or perhaps tables, graphs and charts related to areas, such as population, national income, expenditures, production, consumption, supply, demand, sales, imports, exports, births, deaths and accidents.

Hence, the subject of statistics deals primarily with numerical data gathered from surveys or collected using various statistical methods. Its objective is to summarize such data, so that the summary gives us a good indication about certain characteristics of a population or phenomenon that we wish to study. To ensure that our conclusions are meaningful, it is necessary to subject our data to scientific analyses so that rational decisions can be made. Statistics is therefore concerned with proper collection of data, organization of this data into a manageable and presentable form, analysis and interpretation of the data into conclusions for useful purposes.

The book '*Probability and Statistics*' is divided into four blocks that are further divided into fourteen units which will help you understand how to solve the probability and multivariate distributions, set theory, random variables of the discrete type, random variables of the continuous type, some special expectations, Chebyshev's inequality, correlation coefficient, special distributions and distribution function of random variables, correlation coefficient, binomial and related distributions, Poisson distribution, Gamma and Chi-Square distributions, normal distribution, bivariate normal distribution, sampling theory, transformations of variables of the continuous type (Beta, *t* and F distributions), distributions of order statistics, the moment generating function techniques, distributions of X and ns^2/σ^2, expectations of functions of random variables, limiting distributions, convergence in distribution, convergence in probability, limiting moment generating functions and some theorems on limiting distributions.

The book follows the Self-Instruction Mode or the SIM format wherein each unit begins with an 'Introduction' to the topic followed by an outline of the 'Objectives'. The content is presented in a simple, organized and comprehensive form interspersed with 'Check Your Progress' questions and answers for better understanding of the topics covered. A list of 'Key Words' along with a 'Summary' and a set of 'Self Assessment Questions and Exercises' is provided at the end of the each unit for effective recapitulation.

NOTES

BLOCK - I
PROBABILITY AND MULTIVARIATE DISTRIBUTIONS

UNIT 1 PROBABILITY AND SET THEORY

Structure

1.0 INTRODUCTION

The subject of probability is in itself a vast one, hence only the basic concepts will be discussed in this unit. The word probability or *chance* is very commonly used in day-to-day conversation, and terms such as *possible* or *probable* or *likely,* all mean the same. *Probability* can be defined as *a measure of the likelihood that a particular event* will occur. It is a numerical measure with a value between 0 and 1 of such likelihood where the probability of *zero* indicates that the given event cannot occur and the probability of *one* assures certainty of such an occurrence. Probability theory helps a decision-maker to analyse a situation and decide accordingly. The following are a few examples of such situations:

- What is the *chance* that sales will increase if the price of the product is decreased?
- What is the *likelihood* that a new machine will increase productivity?
- How *likely* is it that a given project will be completed on time?
- What are the possibilities that a competitor will introduce a cheaper substitute in the market?

NOTES

In set theory, the members of a set are called elements. Sets having a definite number of elements are termed as finite sets. You will also learn about singleton set, equality of sets, subsets, empty set or null set and power set. Union, intersection and complement operations on sets are analogous to arithmetic operations such as addition, multiplication and subtraction of numbers, respectively. You will learn to draw Venn diagrams to show all the possible mathematical or logical relationships between sets. Mathematical logic is the analysis of language that helps you to identify valid arguments. You will learn about classes of sets, counting principle and duality.

In this unit, you will study about the probability, set theory, conditional probability and independence.

1.1 OBJECTIVES

After going through this unit, you will be able to:

- Explain about the probability, conditional probability and independence
- Describe the set theory

1.2 PROBABILITY

Probability theory is also called the theory of chance and can be mathematically derived using the standard formulas. A probability is expressed as a real number, $p \in [0, 1]$ and the probability number is expressed as a percentage (0 per cent to 100 per cent) and not as a decimal. For example, a probability of 0.55 is expressed as 55 per cent. When we say that the probability is 100 per cent, it means that the event is certain while the 0 per cent probability means that the event is impossible. We can also express probability of an outcome in the ratio format. For example, we have two probabilities, i.e., 'chance of winning' (1/4) and 'chance of not winning' (3/4), then using the mathematical formula of odds, we can say,

'chance of winning' : 'chance of not winning' = 1/4 : 3/4 = 1 : 3 or 1/3

We are using the probability in vague terms when we predict something for future. For example, we might say it will probably rain tomorrow or it will probably a holiday the day after. This is subjective probability to the person predicting, but implies that the person believes the probability is greater than 50 per cent.

Different types of probability theories are:

(*i*) Classical Theory of Probability

(*ii*) Axiomatic Probability Theory

(*iii*) Empirical Probability Theory

1.2.1 Classical Theory of Probability

NOTES

The classical theory of probability is the theory based on the number of *favourable outcomes* and the number of *total outcomes*. The probability is expressed as a ratio of these two numbers. The term 'favorable' is not the subjective value given to the outcomes, but is rather the classical terminology used to indicate that an outcome belongs to a given event of interest.

Classical Definition of Probability: If the number of outcomes belonging to an event E is N_E, and the total number of outcomes is N, then the probability of event E is defined as $p_E = \frac{N_E}{N}$

For example, a standard pack of cards (without jokers) has 52 cards. If we randomly draw a card from the pack, we can imagine about each card as a possible outcome. Therefore, there are 52 total outcomes. Calculating all the outcome events and their probabilities, we have the following possibilities:

- Out of the 52 cards, there are 13 clubs. Therefore, if the event of interest is drawing a club, there are 13 favourable outcomes, and the probability of this event becomes: $\frac{13}{52} = \frac{1}{4}$
- There are 4 kings (one of each suit). The probability of drawing a king is: $\frac{4}{52} = \frac{1}{13}$
- What is the probability of drawing a king or a club? This example is slightly more complicated. We cannot simply add together the number of outcomes for each event separately (4 + 13 = 17) as this inadvertently counts one of the outcomes twice (the king of clubs). The correct answer is: $\frac{16}{52}$ from $\frac{13}{52} + \frac{4}{52} - \frac{1}{52}$

  We have this from the probability equation, p(club) + p(king) – p(king of clubs).
- Classical probability has limitations, because this definition of probability implicitly defines all outcomes to be equiprobable and this can be only used for conditions such as drawing cards, rolling dice, or pulling balls from urns. We cannot calculate the probability where the outcomes are unequal probabilities.

It is not that the classical theory of probability is not useful because of the above described limitations. We can use this as an important guiding factor to calculate the probability of uncertain situations as mentioned above and to calculate the axiomatic approach to probability.

NOTES

Frequency of Occurrence

This approach to probability is widely used to a wide range of scientific disciplines. It is based on the idea that the underlying probability of an event can be measured by repeated trials.

Probability as a Measure of Frequency: Let n_A be the number of times event A occurs after n trials. We define the probability of event A as,

$$P_A = \lim_{n \to \infty} \frac{n_A}{n}$$

It is not possible to conduct an infinite number of trials. However, it usually suffices to conduct a large number of trials, where the standard of large depends on the probability being measured and how accurate a measurement we need.

Definition of Probability: The sequence $\frac{n_A}{n}$ in the limit that will converge to the same result every time, or that it will not converge at all. To understand this, let us consider an experiment consisting of flipping a coin an infinite number of times. We want that the probability of heads must come up. The result may appear as the following sequence:

HTHHTTHHHHTTTTHHHHHHHHTTTTTTTTHHHHHHHHHHHHHHHHTTTTTTTTTTTTTTTT...

This shows that each run of k heads and k tails are being followed by another run of the same probability. For this example, the sequence $\frac{n_A}{n}$ oscillates between, $\frac{1}{3}$ and $\frac{2}{3}$ which does not converge. These sequences may be unlikely, and can be right. The definition given above does not express convergence in the required way, but it shows some kind of convergence in probability. The problem of formulating exactly can be considered using axiomatic probability theory.

1.2.2 Axiomatic Probability Theory

The axiomatic probability theory is the most general approach to probability, and is used for more difficult problems in probability. We start with a set of axioms, which serve to define a probability space. These axioms are not immediately intuitive and are developed using the classical probability theory.

1.2.3 Empirical Probability Theory

The empirical approach to determining probabilities relies on data from actual experiments to determine approximate probabilities instead of the assumption of equal likeliness. Probabilities in these experiments are defined as the ratio of the frequency of the possibility of an event, $f(E)$, to the number of trials in the experiment, n, written symbolically as $P(E) = f(E)/n$. For example, while flipping a coin, the empirical probability of heads is the number of heads divided by the total number of flips.

The relationship between these empirical probabilities and the theoretical probabilities is suggested by the (Law of Large Numbers). The law states that as the number of trials of an experiment increases, the empirical probability approaches the theoretical probability. Hence, if we roll a die a number of times, each number would come up approximately 1/6 of the time. The study of empirical probabilities is known as *statistics*.

NOTES

1.2.4 Addition Rule

When two events are mutually exclusive, then the probability that either of the events will occur is the sum of their separate probabilities. For example, if you roll a single die then the probability that it will come up with a face 5 or face 6, where event A refers to face *5* and event B refers to face 6, both events being mutually exclusive events, is given by,

$$P[A \text{ or } B] = P[A] + P[B]$$

Or
$$P[5 \text{ or } 6] = P[5] + P[6]$$
$$= 1/6 + 1/6$$
$$= 2/6 = 1/3$$

$P\,[A \text{ or } B]$ is written as $P[A \cup B]$ and is known as $P\,[A \text{ union } B]$.

However, if events A and B are not mutually exclusive, then the probability of occurrence of either event A or event B or both is equal to the probability that event A occurs plus the probability that event B occurs minus the probability that events common to both A and B occur.

Symbolically, it can be written as,

$$P[A \cup B] = P[A] + P[B] - P[A \text{ and } B]$$

$P[A \text{ and } B]$ can also be written as $P[A \cap B]$, known as $P\,[A \text{ intersection } B]$ or simply $P[AB]$.

Events $[A \text{ and } B]$ consist of all those events which are contained in both A and B simultaneously. For example, in an experiment of taking cards out of a pack of 52 playing cards, assume that:

Event A = An ace is drawn.

Event B = A spade is drawn.

Event $[AB]$ = An ace of spade is drawn.

Hence, $P[A \cup B] = P[A] + P[B] - P[AB]$
$$= 4/52 + 13/52 - 1/52$$
$$= 16/52 = 4/13$$

This is because there are 4 aces, 13 cards of spades, including 1 ace of spades out of a total of 52 cards in the pack. The logic behind subtracting

NOTES

$P[AB]$ is that the ace of spades is counted twice—once in event A (4 aces) and once again in event B (13 cards of spade including the ace).

Another example for $P[A \cup B]$, where event A and event B are not mutually exclusive is as follows:

Suppose a survey of 100 persons revealed that 50 persons read *India Today* and 30 persons read *Time* magazine and 10 of these 100 persons read both *India Today* and *Time.* Then:

$$\text{Event } [A] = 50$$
$$\text{Event } [B] = 30$$
$$\text{Event } [AB] = 10$$

Since event $[AB]$ of 10 is included twice, both in event A as well as in event B, event $[AB]$ must be subtracted once in order to determine the event $[A \cup B]$ which means that a person reads *India Today* or *Time* or both. Hence,

$$\begin{aligned} P[A \cup B] &= P[A] + P[B] - P[AB] \\ &= 50/100 + 30/100 - 10/100 \\ &= 70/100 = 0.7 \end{aligned}$$

1.2.5 Multiplication Rule

Multiplication rule is applied when it is necessary to compute the probability if both events A and B will occur at the same time. The multiplication rule is different if the two events are independent as against the two events being not independent.

If events A and B are independent events, then the probability that they both will occur is the product of their separate probabilities. This is a strict condition so that events A and B are *independent if,* and only if,

$$\begin{aligned} P[AB] &= P[A] \times P[B] \text{ or} \\ &= P[A]P[B] \end{aligned}$$

For example, if we toss a coin twice, then the probability that the first toss results in a head and the second toss results in a tail is given by,

$$\begin{aligned} P[HT] &= P[H] \times P[T] \\ &= 1/2 \times 1/2 = 1/4 \end{aligned}$$

However, if events A and B are not independent, meaning that the probability of occurrence of an event is dependent or conditional upon the occurrence or non-occurrence of the other event, then the probability that they will both occur is given by,

$$P[AB] = P[A] \times P[B/\text{given the outcome of } A]$$

This relationship is written as:

$$P[AB] = P[A] \times P[B/A] = P[A]\ P[B/A]$$

NOTES

Where $P[B/A]$ means the probability of event B on the condition that event A has occurred. As an example, assume that a bowl has 6 black balls and 4 white balls. A ball is drawn at random from the bowl. Then a second ball is drawn without replacement of the first ball back in the bowl. The probability of the second ball being black or white would depend upon the result of the first draw as to whether the first ball was black or white. The probability that both these balls are black is given by,

P [two black balls] = P [black on 1st draw] × P [black on 2nd draw/ black on 1st draw]

$= 6/10 \times 5/9 = 30/90 = 1/3$

This is so because, first there are 6 black balls out of a total of 10, but if the first ball drawn is black then we are left with 5 black balls out of a total of 9 balls.

1.2.6 Bayes' Theorem

Reverend Thomas Bayes (1702–1761) introduced his theorem on probability which is concerned with a method for estimating the probability of causes which are responsible for the outcome of an observed effect. Being a religious preacher himself, as well as a mathematician, his motivation for the theorem came from his desire to prove the existence of God by looking at the evidence of the world that God created. He was interested in drawing conclusions about the causes by observing the consequences. The theorem contributes to the statistical decision theory in revising prior probabilities of outcomes of events based upon the observation and analysis of additional information.

Bayes' theorem makes use of conditional probability formula where the *condition* can be described in terms of the additional information which would result in the *revised probability* of the outcome of an event.

Suppose that there are 50 students in our statistics class out of which 20 are male students and 30 are female students. Out of the 30 females, 20 are Indian students and 10 are foreign students. Out of the 20 male students, 15 are Indians and 5 are foreigners, so that out of all the 50 students, 35 are Indians and 15 are foreigners. This data can be presented in a tabular form as follows:

	Indian	***Foreigner***	***Total***
Male	15	5	20
Female	20	10	30
Total	35	15	50

Based upon this information, the probability that a student picked up at random will be female is 30/50 or 0.6, since there are 30 females in the total class of 50 students. Now, suppose that we are given additional information that the person picked up at random is Indian, then what is the probability that this person

NOTES

is a female? This additional information will result in revised probability or *posterior probability* in the sense that it is assigned to the outcome of the event after this additional information is made available.

Since we are interested in the revised probability of picking a female student at random provided that we know that the student is Indian. Let A_1 be the event *female,* A_2 be the event *male* and B the event *Indian.* Then based upon our knowledge of conditional probability, Bayes' theorem can be stated as follows,

$$P(A_1/B) = \frac{P(A_1)P(B/A_1)}{P(A_1)P(B/A_1) + P(A_2)(P(B/A_2)}$$

In the example discussed here, there are 2 basic events which are A_1 (female) and A_2 (male). However, if there are n basic events, A_1, A_2, ...A_n, then Bayes' theorem can be generalized as,

$$P(A_1/B) = \frac{P(A_1)P(B/A_1)}{P(A_1)P(B/A_1) + P(A_2)(P(B/A_2) + ... + P(A_n)P(B/A_n)}$$

Solving the case of 2 events we have,

$$P(A_1/B) = \frac{(30/50)(20/30)}{(30/50)(20/30) + (20/50)(15/20)} = 20/35 = 4/7 = 0.57$$

This example shows that while the *prior probability* of picking up a female student is 0.6, the *posterior probability* becomes 0.57 after the additional information that the student is an American is incorporated in the problem.

Another example of application of Bayes' theorem is as follows:

Example 1. A businessman wants to construct a hotel in New Delhi. He generally builds three types of hotels. These are 50 rooms, 100 rooms and 150 rooms hotels, depending upon the demand for the rooms, which is a function of the area in which the hotel is located, and the traffic flow. The demand can be categorized as low, medium or high. Depending upon these various demands, the businessman has made some preliminary assessment of his net profits and possible losses (in thousands of dollars) for these various types of hotels. These pay-offs are shown in the following table.

States of Nature

		Demand for Rooms			
		Low (A_1)	*Medium* (A_2)	*High* (A_3)	
		0.2	0.5	0.3	Demand Probability
Number of Rooms	R_1 = (50)	25	35	50	
	R_2 = (100)	–10	40	70	
	R_3 = (150)	–30	20	100	

NOTES

Solution. The businessman has also assigned 'prior probabilities' to the demand structure or rooms. These probabilities reflect the initial judgement of the businessman based upon his intuition and his degree of belief regarding the outcomes of the states of nature.

Demand for rooms	*Probability of Demand*
Low (A_1)	0.2
Medium (A_2)	0.5
High (A_3)	0.3

Based upon these values, the expected pay-offs for various rooms can be computed as follows,

$$EV\,(50) = (\,25 \times 0.2) + (35 \times 0.5) + (50 \times 0.3) = 37.50$$

$$EV\,(100) = (-10 \times 0.2) + (40 \times 0.5) + (70 \times 0.3) = 39.00$$

$$EV\,(150) = (-30 \times 0.2) + (20 \times 0.5) + (100 \times 0.3) = 34.00$$

This gives us the maximum pay-off of $39,000 for building a 100 rooms hotel.

Now the hotelier must decide whether to gather additional information regarding the states of nature, so that these states can be predicted more accurately than the preliminary assessment. The basis of such a decision would be the cost of obtaining additional information. If this cost is less than the increase in maximum expected profit, then such additional information is justified.

Suppose that the businessman asks a consultant to study the market and predict the states of nature more accurately. This study is going to cost the businessman $10,000. This cost would be justified if the maximum expected profit with the new states of nature is at least $10,000 more than the expected pay-off with the prior probabilities. The consultant made some studies and came up with the estimates of low demand (X_1), medium demand (X_2), and high demand (X_3) with a degree of reliability in these estimates. This degree of reliability is expressed as conditional probability which is the probability that the consultant's estimate of low demand will be correct and the demand will be actually low. Similarly, there will be a conditional probability of the consultant's estimate of medium demand, when the demand is actually low, and so on. These conditional probabilities are expressed in Table 1.1.

Table 1.1 *Conditional Probabilities*

		X_1	X_2	X_3
States of	(A_1)	0.5	0.3	0.2
Nature	(A_2)	0.2	0.6	0.2
(Demand)	(A_3)	51	0.3	0.6

The values in the preceding table are conditional probabilities and are interpreted as follows:

The upper north-west value of 0.5 is the probability that the consultant's prediction will be for low demand (X_1) when the demand is actually low. Similarly,

NOTES

the probability is 0.3 that the consultant's estimate will be for medium demand (X_2) when in fact the demand is low, and so on. In other words, $P(X_1/A_1) = 0.5$ and $P(X_2/A_1) = 0.3$. Similarly, $P(X_1/A_2) = 0.2$ and $P(X_2/A_2) = 0.6$, and so on.

Our objective is to obtain posteriors which are computed by taking the additional information into consideration. One way to reach this objective is to first compute the joint probability which is the product of prior probability and conditional probability for each state of nature. Joint probabilities as computed is given as,

States of Nature	*Prior Probability*		*Joint Probabilities*	
		$P(A_1X_1)$	$P(A_1X_2)$	$P(A_1X_3)$
A_1	0.2	$0.2 \times 0.5 = 0.1$	$0.2 \times 0.3 = 0.06$	$0.2 \times 0.2 = 0.04$
A_2	0.5	$0.5 \times 0.2 = 0.1$	$0.5 \times 0.6 = 0.3$	$0.5 \times 0.2 = 0.1$
A_3	0.3	$0.3 \times 0.1 = 0.03$	$0.3 \times 0.3 = 0.09$	$0.3 \times 0.6 = 0.18$
Total marginal probabilities		$= 0.23$	$= 0.45$	$= 0.32$

Now, the posterior probabilities for each state of nature A_i are calculated as follows:

$$P(A_i / X_j) = \frac{\text{Joint probability of } A_i \text{ and } X_j}{\text{Marginal probability of } X_j}$$

By using this formula, the joint probabilities are converted into posterior probabilities and the computed table for these posterior probabilities is given as,

States of Nature	*Posterior Probabilities*		
	$P(A_1/X_1)$	$P(A_1/X_2)$	$P(A_1/X_3)$
A_1	$0.1/.023 = 0.435$	$0.06/0.45 = 0.133$	$0.04/0.32 = 0.125$
A_2	$0.1/.023 = 0.435$	$0.30/0.45 = 0.667$	$0.1/0.32 = 0.312$
A_3	$0.03/.023 = 0.130$	$0.09/0.45 = 0.200$	$0.18/0.32 = 0.563$
Total	$= 1.0$	$= 1.0$	$= 1.0$

Now, we have to compute the expected pay-offs for each course of action with the new posterior probabilities assigned to each state of nature. The net profits for each course of action for a given state of nature is the same as before and is restated as follows. These net profits are expressed in thousands of dollars.

		Low (A_1)	*Medium* (A_2)	*High* (A_3)
Number of Rooms	(R_1)	25	35	50
	(R_2)	–10	40	70
	(R_3)	–30	20	100

Let O_{ij} be the monetary outcome of course of action (i) when (j) is the corresponding state of nature, so that in the above case O_{i1} will be the outcome of course of action R_1 and state of nature A_1, which in our case is \$25,000. Similarly, O_{i2} will be the outcome of action R_2 and state of nature A_2, which in our case is – \$10,000, and so on. The expected value EV (in thousands of dollars) is calculated on the basis of actual state of nature that prevails as well as the estimate of the state of nature as provided by the consultant. These expected values are calculated as follows,

Course of action $= R_i$

Estimate of consultant $= X_i$

Actual state of nature $= A_i$

where, $i = 1, 2, 3$

Then

(A) Course of action $= R_1 =$ Build 50 rooms hotel

$$EV\left(\frac{R_1}{X_1}\right) = \Sigma P\left(\frac{A_i}{X_1}\right)O_{i1}$$

$$= 0.435(25) + 0.435\,(-10) + 0.130\,(-30)$$

$$= 10.875 - 4.35 - 3.9 = 2.625$$

$$EV\left(\frac{R_1}{X_2}\right) = \Sigma P\left(\frac{A_i}{X_2}\right)O_{i1}$$

$$= 0.133(25) + 0.667\,(-10) + 0.200\,(-30)$$

$$= 3.325 - 6.67 - 6.0 = -9.345$$

$$EV\left(\frac{R_1}{X_3}\right) = \Sigma P\left(\frac{A_i}{X_3}\right)O_{i1}$$

$$= 0.125(25) + 0.312(-10) + 0.563(-30)$$

$$= 3.125 - 3.12 - 16.89$$

$$= -16.885$$

(B) Course of action $= R_2 =$ Build 100 rooms hotel

$$EV\left(\frac{R_2}{X_1}\right) = \Sigma P\left(\frac{A_i}{X_1}\right)O_{i2}$$

$$= 0.435(35) + 0.435\,(40) + 0.130\,(20)$$

$$= 15.225 + 17.4 + 2.6 = 35.225$$

NOTES

$$EV\left(\frac{R_2}{X_2}\right) = \Sigma P\left(\frac{A_i}{X_1}\right)O_{i2}$$

$$= 0.133(35) + 0.667\ (40) + 0.200\ (20)$$

$$= 4.655 + 26.68 + 4.0 = 35.335$$

$$EV\left(\frac{R_2}{X_3}\right) = \Sigma P\left(\frac{A_i}{X_3}\right)O_{i2}$$

$$= 0.125(35) + 0.312(40) + 0.563(20)$$

$$= 4.375 + 12.48 + 11.26 = 28.115$$

(C) Course of action $= R_3 =$ Build 150 room-hotel

$$EV\left(\frac{R_3}{X_1}\right) = \Sigma P\left(\frac{A_i}{X_1}\right)O_{i3}$$

$$= 0.435(50) + 0.435(70) + 0.130\ (100)$$

$$= 21.75 + 30.45 + 13 = 65.2$$

$$EV\left(\frac{R_3}{X_2}\right) = \Sigma P\left(\frac{A_i}{X_2}\right)O_{i3}$$

$$= 0.133(50) + 0.667\ (70) + 0.200\ (100)$$

$$= 6.65 + 46.69 + 20 = 73.34$$

$$EV\left(\frac{R_3}{X_3}\right) = \Sigma P\left(\frac{A_i}{X_3}\right)O_{i3}$$

$$= 0.125(50) + 0.312(70) + 0.563(100)$$

$$= 6.25 + 21.84 + 56.3 = 84.39$$

The calculated expected values in thousands of dollars, are presented in a tabular form.

Expected Posterior Pay-Offs

Outcome	$EV(R_1/X_i)$	$EV(R_2/X_i)$	$EV(R_3/X_i)$
X_1	2.625	35.225	65.2
X_2	–9.345	35.335	73.34
X_3	–16.885	28.115	84.39

This table can now be analysed in the following manner.

If the outcome is X_1, it is desirable to build 150 rooms hotel, since the expected pay-off for this course of action is maximum of \$65,200. Similarly, if the

NOTES

outcome is X_2, the course of action should again be R_3 since the maximum pay-off is \$73,34. Finally, if the outcome is X_3, the maximum pay-off is \$84,390 for course of action R_3.

Accordingly, given these conditions and the pay-off, it would be advisable to build a hotel which has 150 rooms.

1.2.7 Sample Space and Events

Sample Space

A sample space is the collection of all possible events or outcomes of an experiment. For example, there are two possible outcomes of a toss of a fair coin, which are a head and a tail. Then, the sample space for this experiment denoted by S would be:

$$S = [H,\ T]$$

This makes the probability of the sample space equals 1 or,

$$P[S] = P[H,T] = 1$$

This is so because in the toss of the coin, either a head or a tail must occur. Similarly, when we roll a die, any of the six faces can come as a result of the roll, since there are a total of six faces. Hence, the sample space is $S = [1, 2, 3, 4, 5, 6]$, and $P[S] = 1$, since one of the six faces must occur.

Events

An event is an outcome or a set of outcomes of an activity or a result of a trial. For example, getting two heads in the trial of tossing three fair coins simultaneously would be an event. The following are the various types of events:

Elementary Event. An elementary event, also known as a simple event, is a single possible outcome of an experiment. For example, if we toss a fair coin, then the event of a head coming up is an elementary event. If the symbol for an elementary event is (E), then the probability of the event (E) is written as $P[E]$.

Joint Event. A joint event, also known as *compound event,* has two or more elementary events in it. For example, drawing a *black ace* from a pack of cards would be a joint event, since it contains two elementary events of black and ace.

Mutually Exclusive Events. Two events are said to be mutually exclusive if the occurrence of one precludes the occurrence of another.

All simple events are mutually exclusive. Compound events are mutually exclusive only when they contain no simple event in common. Thus, in drawing of a card from a normal pack, the events 'a red card' and 'a spade' are mutually exclusive as a card canot be red and a spade at the same time. But 'a heart' and 'a king' are not mutually exclusive as both the compound events include the elementary event 'King of Heart'.

NOTES

Complementary Event. Two mutually exclusive events are said to be complementary if they between themselves exhaust all possible outcomes. Thus, 'not Ace of Heart' and 'Ace of Heart' are complementary events. So also are 'no head' and 'at least one head' in repeated flipping of a coin. If one of them is denoted by A, the other (complementary) event is denoted by $\overline{A}$.

How is an Event Represented? It is a statement about one or more outcomes of an experiment. For example, 'a number greater than 4 appears' is an event for the experiment of throwing a dice.

Consider an experiment of drawing a card from a pack containing just four cards: Ace of Spades, Ace of Hearts, King of Spades and King of Hearts. We draw any one of these four cards. So there are four possible outcomes or events— drawing of *SA, HA, SK,* or *HK*. These events are called simple events because they cannot be decomposed further into two or more events.

Any set of *simple events* can be represented on diagram like Figure 1.1 The collection of all possible simple events in an experiment is called a *sample space* or a *possibility space*. Thus the sample space of drawing a card from the pack described earlier consists of four points.

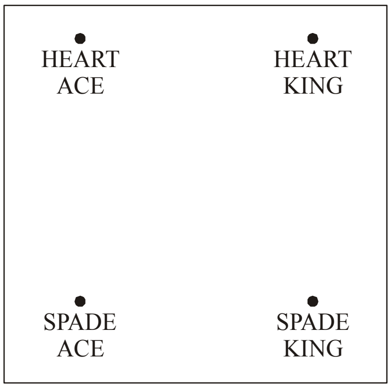


Fig. 1.1 *Simple Events*

An event is termed *compound* if it represents two or more simple events. Thus, the event 'Spade' is a compound event as it represents two simple events *SA* and *SK* (Refer Figure 1.2).

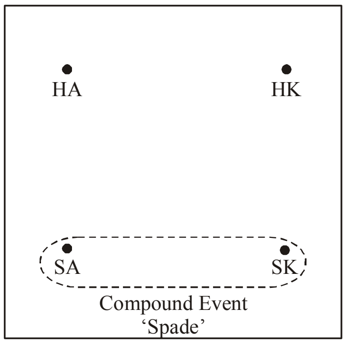


Fig. 1.2 *Compound Event*

NOTES

Similarly, 'not *HA*' is also a compound event made up of all event except *HA*, that is made up of *SA, SK* and *HK* (Refer Figure 1.3).

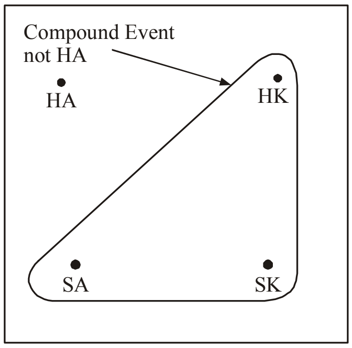


Fig. 1.3 *Compound Event*

The sample may be *discrete* or *continuous*. If we are dealing with discrete variable, the sample space is discrete and if we are dealing with continuous variables it is continuous. The sample space for rolling of two dices is a discrete one consisting of 36 points (Refer Figure 1.4) and that for the weights of individuals selected at random would be continuous.

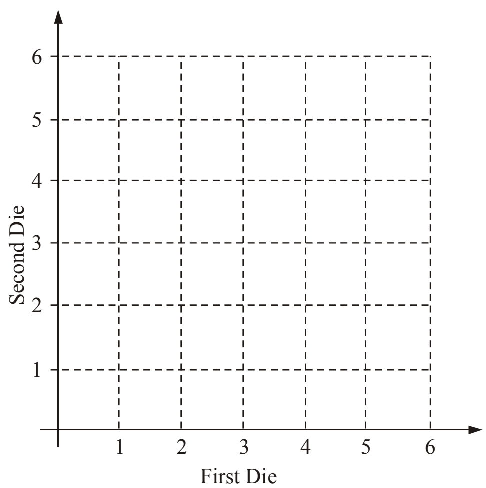


Fig. 1.4 *Discrete and Continuous Sample Space*

Sum of Events. Sum of two events A_1 and A_2 is the compound event 'either A_1 or A_2 or both', i.e., at least one of A_1 and A_2 occurs. This is denoted by $A_1 + A_2$. In general, $A_1 + A_2 + ... + A_n$ is the event which means the occurrence of at least one of A_i's.

Product of Events. Product of two events A_1A_2 is the compound event 'A_1 and A_2 both occur'. This is denoted by A_1A_2. Obviously, if A_1 and A_2 are two mutually exclusive events than A_1A_2 is impossible event.

Suppose a person is required to calculate the possibility of the occurence of one outcome (simple or compound) of an experiment. One method to do this is to try the experiment a large number of times under exactly similar circumstances. If an outcome occurs m times in n trials, m/n is called is relative frequency. It is conventional to use the term *success* whenever the event under consideration takes place and *failure* whenever it does not.

If the outcome of the experiment be represented by graph in which we have the total number of trials n on the horizontal axis and the proportion of successes m/n on the vertical axis, we note the following points:

1. When n is small, the ratio m/n fluctuates considerably.
2. When n becomes large, the ratio m/n becomes stable and tends to settle down to a certain value, say P.

From these, we conclude that when an experiment is repeated a large number of times, the proportion of times the event occurs would be practically equal to the number P.

We call the number P the probability of occurrence of the given event.

Thus, when we talk of the probability of an event, we simply refer to the proportion of times that event occurs in a *large* number of trials or *in a long run*. This is called the relative frequency approach of defining probability.

So the probability of getting a six in a single rolling of a die is the proportion of times a six would show up in a large number of rollings of a single die under exactly similar circumstances.

Note carefully that P and the proportion of success m/n are not the same things. The ratio m/n changes with n while P does not. It is a fixed number. However, when n is large and P is not known, m/n is taken as an estimate of P.

1.2.8 Finite Probability Spaces

A probability space is a measure of space such that the measure of the whole space is equal to 1. A simple finite probability space is an ordered pair (S, p) such that S is set and p is a function with domain S. The range is a subset of $[0,1]$ such that,

$$\sum_{s \in S} p(s) = 1$$

Suppose (S, p) be a simple finite probability space. Then,

$A = \{A : A \subset S\}$

Let, $P(A) = \sum_{s \in A} p(s) = 1$ for $A \in A$

It can be easily verified that, (S, A, P) is a probability space.

A simple and frequently used function p is obtained by letting $p(s)$ equal one over the number of elements of S for each $s \in S$.

Definition: A finite probability space is a finite set $\Omega \neq 0$ together with a function, $\Pr : \Omega \to \mathbf{R}^+$ such that, (*i*) $\forall \omega \in \Omega, \Pr(\omega) > 0$

(*ii*) $\sum_{\omega \in \Omega} \Pr(\omega) = 1$

NOTES

Here, the set Ω is the sample space and the function Pr is the probability distribution. The Elements $\omega \in \Omega$ are called atomic events or elementary events. An event is a subset of Ω. The uniform distribution over the sample space is defined by setting $\Pr(\omega) = 1/|\Omega|$ for every $\omega \in \Omega$. This distribution defines the uniform probability space over Ω. In a uniform space, calculation of probabilities amounts to counting: $\Pr(A) = |A| / |\Omega|$.

Check Your Progress

1. Define the terms simple probability and joint probability.
2. Explain the classical theory of probability.
3. What is the addition rule?
4. When is the law of multiplication applied?
5. What is Bayes' theorem?
6. What is a mutually exclusive event?

1.3 SET THEORY

Sets are one of the most fundamental concepts in mathematics. A set is a collection of distinct object considered as a whole. Thus we say,

'A *set* is any collection of objects such that given an object, it is possible to determine whether that object belongs to the given collection or not.'

The members of a set are called *elements.* We use capital letters to denote sets and small letters to denote elements. We always use { } brackets to denote a set.

Examples of Sets: (*i*) The set of all integers.
(*ii*) The set of all students of Delhi University.
(*iii*) The set of all letters of the alphabet.
(*iv*) The set of even integers 2, 4, 6, 8.

Example 2. Let M be the collection of all those men (only those men) in a village who do not shave themselves. Given that, (*i*) All men in the village must be clean shaven, (*ii*) The village barber shaves all those men who do not shave themselves.

Solution. Suppose b denotes the village barber. If $b \in M$, then b does not shave himself. Then as per given statement (*ii*), b shaves himself, is a contradiction.

If $b \notin M$, then b shaves himself. Then as per the given statement (*i*), b does not shave himself, again a contradiction.

Since we cannot answer 'Yes' or 'No' to the question, 'Is barber himself a member of M?' We conclude that M is not a set.

Elements

NOTES

The members of a set are called its *elements*. We use capital letters to denote sets and small letters to denote elements. If a is an element of the set A, we write it as, $a \in A$ (read as 'a belongs to A') and if a is not an element of the set A, we write it as, $a \notin A$ (read as 'a does not belong to A'). There are different ways of describing a set. For example, the set consisting of elements 1, 2, 3, 4, 5 could be written as $\{1, 2, 3, 4, 5\}$ or $\{1, 2, \ldots, 5\}$ or $\{x \mid x \in N, x \leq 5\}$,

Where, N = Set of natural numbers.

We always use { } brackets to denote a set. A set which has finite number of elements is called a *finite set*, else it is called an *infinite set.* For example, if A is the set of all integers, then A is an infinite set denoted by $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$ or $\{x \mid x \text{ is an integer}\}$.

Singleton

A set having only one element is called *singleton*. If a is the element of the singleton A, then A is denoted by $A = \{a\}$. Note that $\{a\}$ and a do not mean the same; $\{a\}$ stands for the set consisting of a single element a, while a is just the element of $\{a\}$. It is the simplest example of a nonempty set.

Equality of Sets

Two sets A and B are said to be *equal* if every member of A is a member of B and every member of B is a member of A. We express this by writing $A = B$, logically speaking $A = B$ means $(x \in A) \equiv (x \in B)$ or the biconditional statement $(x \in A) \Leftrightarrow (x \in B)$ is true for all x.

Notes: 1. The order of appearance of the elements of a set is of no consequence. For example, the set $\{1, 2, 3\}$ is same as $\{2, 3, 1\}$ or $\{3, 2, 1\}$, etc.

2. Each element of a set is written only once. For example, $\{2, 2, 3\}$ is not a proper way of writing a set and it should be written as $\{2, 3\}$.

Universal Set

Whenever we talk of a set, we shall assume it to be a subset of a fixed set U. This fixed set U is called the universal set.

Subsets

Let A and B be two sets. If every element of A is an element of B, then A is called a subset of B and we write $A \subseteq B$ or $B \supseteq A$ (read as 'A is contained in B' or 'B contains A').

Logically speaking, $A \subseteq B$ means $(xm \in A) \Rightarrow (x \in B)$ is true for every x.

Notes: 1. If $A \subseteq B$ and $A \neq B$, we write $A \subset B$ or $B \supset A$ (read as: A is a *proper subset* of B or B is a *proper superset of* A).

2. Every set is a subset and a superset of itself.
3. If A is not a subset of B, we write $A \not\subseteq B$.

Empty Set or Null Set

A set which has no element is called the *null set* or *empty set.* It is denoted by the symbol ϕ.

For example, each of the following is a null set:

(*i*) The set of all real numbers whose square is –1.

(*ii*) The set of all those integers that are both even and odd.

(*iii*) The set of all rational numbers whose square is 2.

(*iv*) The set of all those integers x that satisfy the equation $2x = 5$.

Example 3. The empty set ϕ is a subset of every set.

Solution. Suppose ϕ is not a subset of the set A. This means there exists $a \in \phi$ such that $a \notin A$. This is impossible as ϕ has no element. So, ϕ is a subset of every set.

Aliter. Logically speaking, this can be proved that the conditional statement $(x \in \phi) \Rightarrow (x \in A)$ is true for every x. Since ϕ has no element, the statement '$x \in \phi$' is false. Hence, the conditional statement $(x \in \phi) \Rightarrow (x \in A)$ is true, which proves the result.

Example 4. List the following sets (here N denotes the set of natural numbers and Z, the set of integers).

(*i*) $\{x \mid x \in N \text{ and } x < 10\}$

(*ii*) $\{x \mid x \in Z \text{ and } x < 6\}$

(*iii*) $\{x \mid x \in Z \text{ and } 2 < x < 10\}$

Solution. (*i*) We have to find the natural numbers which are less than 10. They are 1, 2, 3, 4, 5, 6, 7, 8, 9. The set can be described as $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

(*ii*) We have to find integers which are less than 6. They are all negative integers and the integers 0, 1, 2, 3, 4, 5. The set may be described as,

$\{\ldots, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$.

(*iii*) We have to find integers that are between 2 and 10. They are 3, 4, 5, 6, 7, 8, 9. The set may be described as $\{3, 4, 5, 6, 7, 8, 9\}$.

Example 5. Give the verbal translation of the following sets:

(*i*) $\{2, 4, 6, 8\}$

(*ii*) $\{1, 3, 5, 7, 9, \ldots\}$

(*iii*) $\{-1, 1\}$

Solution. (*i*) It consists of all positive even integers less than 10.

(*ii*) It consists of all positive odd integers.

(*iii*) It consists of those integers x which satisfy $x^2 - 1 = 0$.

NOTES

Example 6. If $a_1 \neq b_1$ and $\{a_1, b_1\} = \{a_2, b_2\}$ then show that $a_2 \neq b_2$.

Solution. Let $a_2 = b_2$. Then $a_1 \in \{a_1, b_1\}$ means $a_1 \in \{a_2, b_2\} = \{a_2\}$. So $a_1 = a_2$. Also $b_2 \in \{a_1, b_1\}$ means $b_1 \in (a_2, b_2) = \{a_2\}$. So, $b_1 = a_2$. Therefore, $a_1 = b_1$, which is wrong. Thus $a_2 \neq b_2$.

Example 7. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

Solution. Let $a \in A$ be any element of A. Then as $A \subseteq B$, $a \in B$.

Also $B \subseteq C \Rightarrow a \in C$.

Thus every element of A belongs to $C \Rightarrow A \subseteq C$.

Aliter. Logically speaking, we want to prove that,

$$[(x \in A) \Rightarrow (x \in B)] \wedge [(x \in B) \Rightarrow (x \in C)] \Rightarrow [(x \in A) \Rightarrow (x \in C)]$$

is true for every x. This follows the Transitive Law.

Example 8. If $A \subseteq B$ and $B \subseteq A$, then $A = B$.

Solution. Since $A \subseteq B$, every element of A is an element of B. Also $B \subseteq A$, means every element of B is also an element of A. This proves $A = B$.

Aliter. Logically speaking, this can be proved as,

$[(x \in A) \Rightarrow (x \in B)] \wedge [(x \in B) \Rightarrow (x \in A)] \Rightarrow [(x \in A) \Leftrightarrow (x \in B)]$ is true for every x. In other words, $[(p \Rightarrow q) \wedge (q \Rightarrow p)] \Rightarrow (p \Leftrightarrow q)$ is true. Since $p \Rightarrow q$ is true and $q \Rightarrow p$ is true, $(p \Rightarrow q) \wedge (q \Rightarrow p)$ is also true. This also means that $p \Leftrightarrow q$ is true. So, $[(p \Rightarrow q) \wedge (q \Rightarrow p)] \Rightarrow (p \Leftrightarrow q)$ is true. This proves the result.

Example 9. If $A \subset B$ and $B \subseteq C$, then $A \subset C$.

Solution. If $A = C$, then every element of B is also an element of A (as $B \subseteq A$). But $A \subset B$ means every element of A is also an element of B. Combining these facts, we get $A = B$ is a contradiction (as A is a proper subset of B). So, $A \neq C$. Clearly, every element of A is also an element of C. Therefore, A is a proper subset of C.

Aliter. If $A = C$, then $B \subseteq A$. This means $(x \in B) \Rightarrow (x \in A)$ is true for every x.

Also $A \subset B$ means $(x \in A) \Rightarrow (x \in B)$ is true for all x. Therefore,

$(x \in A) \Leftrightarrow (x \in B)$ is true for every x. So, $A = B$, is not possible as A is proper subset of B. Hence, $A \neq C$. A is subset of C.

Example 10. Find all possible solutions for x and y for each of the following cases:

(*i*) $\{2x, y\} = \{4, 6\}$

(*ii*) $\{x, 2y\} = \{1, 2\}$

(*iii*) $\{2x\} = \{0\}$

Solution. (*i*) Let $A = \{2x, y\}$ and $B = \{4, 6\}$

Now $2x \in A$ means $2x \in B$. So, $2x = 4$ or $2x = 6$. If $2x = 4$, then $x = 2$. Also $y \in A$ means $y \in B$. So, $y = 4$ or $y = 6$. However, y cannot be equal to 4. Then A will have only element 4 while B will have elements 4 and 6. Therefore, one solution

NOTES

is $x = 2$ and $y = 6$. If $2x = 6$, then $x = 3$. But y cannot be 6. Then A will have only element 6. Therefore,y must be 4. Another solution is $x = 3$ and $y = 4$.

(ii) Let , $A = \{x, 2y\}$ and $B = \{1, 2\}$

$x \in A$ means $x \in B$

So, $x = 1$ or $x = 2$

If $x = 1$, then $2y = 2$. So, one solution is $x = 1$ and $y = 1$

If $x = 2$, then $2y = 1$. So, another solution is $x = 2$ and $y = \frac{1}{2}$

(iii) Let, $A = \{2x\}$, $B = \{0\}$

$2x \in A$ means $2x \in B$

So, $2x = 0$ which means $x = 0$

Therefore, the only solution is $x = 0$.

Example 11. Find at least one set A such that,

(i) $\{1, 2\} \subseteq A \subset \{1, 2, 3, 4\}$

(ii) $\{0, 1, 2\} \subset A \subset \{2, 3, 0, 1, 4\}$

Solution. (i) Since $\{1, 2\} \subset A$, it means A must have 1, 2 as its elements and also some other members. Again, $A \subset \{1, 2, 3, 4\}$ means that the extra member should be 3 or 4. So $A = \{1, 2, 3\}$ or $A = \{1, 2, 4\}$.

(ii) Considering the solution of above case (i), there are two possibilities. Either$A = \{0, 1, 2, 3\}$ or $A = \{0, 1, 2, 4\}$.

Venn Diagrams

Venn diagrams are used to illustrate various set operations. It is named after John Venn (1834–1883). You can represent the universal set by the points in and on a rectangle and subsets $A, B, C, \ldots$by points in and on the circles or ellipses drawn inside the rectangle. In Figure 1.5, the shaded portion represents $A \cap B$.

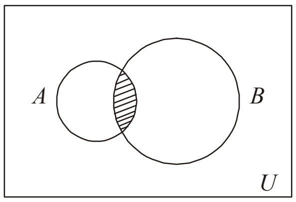


Fig. 1.5

In Figure 1.6, the shaded portion represents $A \cup B$.

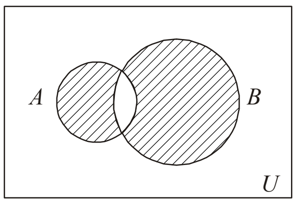


Fig. 1.6

NOTES

In Figure 1.7, the shaded portion represents A'.

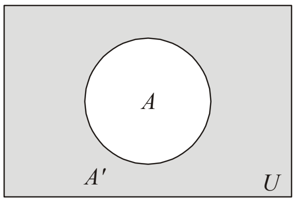


Fig. 1.7

In Figure 1.8, the shaded portion represents $A - B$.

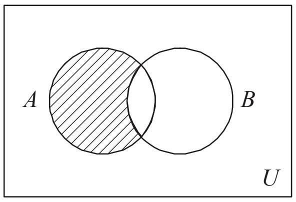


Fig. 1.8

In Figure 1.9, three sets A, B, C divide the universal set U into 8 parts. Eighth is part not numbered in the Venn diagram.

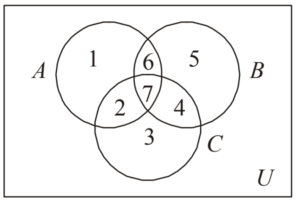


Fig. 1.9

Example 12. Prove that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ using Venn diagrams.

Solution. This can be proved with reference to Figure 1.9. $B \cap C$ is represented by areas 4 and 7, and A is represented by areas 1, 2, 6 and 7. So, $A \cup (B \cap C)$ is represented by areas 1, 2, 4, 6 and 7. Again, areas 1, 2, 4, 5, 6, 7 represent $A \cup B$ and areas 1, 2, 3, 4, 6, 7 represent $A \cup C$. So, areas 1, 2, 4, 6, 7 represents $(A \cup B) \cap (A \cup C)$. This proves our assertion.

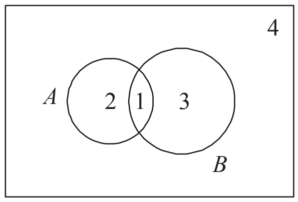
Example 13. Using Venn diagrams show that $A - (B \cup C) = (A - B) \cup (A - C)$.

Solution. To prove this see Figure 1.5, Areas 2, 3, 4, 5, 6, 7 represent $B \cup C$. Therefore area 1 represents $A - (B \cup C)$. Now areas, 1, 2 represent $A - B$ and areas 1, 6 represent $A - C$ and area 1 represents $(A - B) \cap (A - C)$. This proves the result.

Example 14. Using Venn diagrams show that for any two sets A and B,

$$(A \cap B)' = A' \cup B'$$

Solution. In Figure below, area 1 represents $A \cap B$ while areas 2, 3, 4 represent $(A \cap B)'$. Again areas 3, 4 represent A' and areas 2, 4 represent B'. Therefore areas 2, 3, 4 represent $A' \cup B'$.



NOTES

Example 15. Use Venn diagrams to show that for any sets A and B,

$$A \cup B = A \cup (B - A)$$

Solution. Refer Figure given in Example 14. Areas 1, 2, 3 represent $A \cup B$. Also areas 1, 2 represent A and area 3 represents $B - A$. So, areas 1, 2, 3 represent $A \cup (B - A)$. This proves the result.

Operations with Sets

The reader is familiar with the operations of addition and multiplication in Arithmetic. For any two given numbers, the operations of addition and multiplication associate another number which is called sum or product of two numbers respectively. In this section, we will define three operations for associating any two given sets as a third set. These three operations namely, union, intersection and complement, analogous to the operations of addition, multiplications and subtraction of numbers respectively.

Union

The union of any two sets A and B is the set of all those elements x such that x belongs to at least one of the two sets A and B. It is denoted by $A \cup B$. Logically speaking, if the biconditional statement $(x \in C) \Leftrightarrow (x \in A) \vee (x \in B)$ is true for all x, then $C = A \cup B$. In other words $(x \in A \cup B) \equiv (x \in A) \vee (x \in B)$.

Example 16. Prove that for any sets A and B (*i*) $A \subseteq A \cup B$, (*ii*) $B \subseteq A \cup B$.

Solution. (*i*) $x \in A$ means $x \in A \cup B$, by definition. Therefore, $A \subseteq A \cup B$.

(*ii*) $x \in B$ means $x \in A \cup B$, by definition. Therefore, $B \subseteq A \cup B$.

Aliter. (*i*) We want to prove that the conditional statement,

$$(x \in A) \Rightarrow (x \in A \cup B) \text{ is true}$$

But this statement is false if $(x \in A)$ is true and $(x \in A \cup B)$ is false. Such a situation cannot occur, therefore for $(x \in A)$ is true means that $(x \in A) \vee (x \in B)$ is true. Hence, $(x \in A) \vee (x \in B)$ is true and $(x \in A \cup B)$ is false. It means $(x \in A) \vee (x \in B) \Rightarrow (x \in A \cup B)$ is false. This is impossible by definition of $A \cup B$. Similarly, we can prove case (*ii*).

Example 17. If $A \subseteq B$, then $A \cup B = B$ and conversely, if $A \cup B = B$, then $A \subseteq B$.

Solution. Suppose $A \subseteq B$. Let $x \in A \cup B$. Then $x \in A$, or $x \in B$ or $x \in$ (A and B). If $x \in A$, then $x \in B$ (as $A \subseteq B$). In any case, $x \in A \cup B$ means $x \in B$. So, $A \cup B \subseteq B$. We have already proved $A \subseteq A \cup B$. Therefore, $A \cup B = B$. Conversely, let $A \cup B = B$. Let $x \in A$. Then $x \in A \cup B$, which means $x \in B$. Hence, $A \subseteq B$.

NOTES

Aliter. Suppose $A \subseteq B$. This can be proved that the biconditional statement $(x \in B) \Leftrightarrow (x \in A) \vee (x \in B)$ is true for every x. But this statement is false if and only if $(x \in B)$ is false and $(x \in A)$ is true. Such a situation cannot occur as $A \subseteq B$.

This proves $A \cup B = B$.

Conversely, if $A \cup B = B$, then we want to show that the conditional statement $(x \in A) \Rightarrow (x \in B)$ is true for every x. This is false if and only if $(x \in A)$ is true and $(x \in B)$ is false. Now $(x \in A)$ is true means $(x \in A) \vee (x \in B)$ is true. Therefore, $(x \in A) \vee (x \in B) \Rightarrow (x \in B)$ is false. This is impossible as $B = A \cup B$. This proves $A \subseteq B$.

Example 18. If $A \subseteq C$ and $B \subseteq C$, then $(A \cup B) \subseteq C$.

Solution. We want to show that $(x \in A \cup B) \Rightarrow (x \in C)$ is true for every x. This is equivalent to say that $(x \in A \cup B)$ is true and $(x \in C)$ is false cannot occur together. Suppose $(x \in A \cup B)$ is true. Then $(x \in A) \vee (x \in B)$ is true. This means $(x \in A)$ is true or $(x \in B)$ is true. If $(x \in A)$ is true then $(x \in C)$ is true as $A \subseteq C$. If $(x \in B)$ is true then $(x \in C)$ is true as $B \subseteq C$. In any case $(x \in C)$ is true. So, when $(x \in \mathrm{A} \cup B)$ is true, $(x \in C)$ should also be true. This proves our assertion.

Aliter. Let $x \in A \cup B$. This means $x \in A$ or $x \in B$ or $x \in$ $(A$ and $B)$. If $x \in A$, then $x \in C$ (as $A \subseteq C$). If $x \in B$, then $x \in C$ (as $B \subseteq C$). In any case, $x \in C$. So, $x \in A \cup B$ means $x \in C$.

This proves $A \cup B \subseteq C$.

Intersection

The intersection of two sets A and B is the set of all those elements x such that x belongs to both A and B and is denoted by $A \cap B$. If $A \cap B = \phi$, then A and B are said to be *disjoint*.

Logically speaking, if the biconditional statement $(x \in C) \Leftrightarrow (x \in A) \wedge (x \in B)$ is true for all x, then $C = A \cap B$. Hence,

$$(x \in A \cap B) \equiv (x \in A) \cap (x \in B)$$

Example 19. Show that for any sets A and B (*i*) $A \cap B \subseteq A$ (*ii*) $A \cap B \subseteq B$

Solution. Let $x \in A \cap B$. Then, by definition $x \in A$ and $x \in B$. Therefore, $A \cap B \subseteq A$ and $A \cap B \subseteq B$.

Aliter. (*i*) This can be proved that,

$$(x \in A \cap B) \Rightarrow (x \in A) \text{ is true for all } x.$$

Now, consider the case when $(x \in A \cap B)$ is true and $(x \in A)$ is false. Here, $(x \in A)$ is false means $(x \in A) \cap (x \in B)$ is false and so $(x \in A \cap B) \Rightarrow (x \in A) \cap (x \in B)$ is also false which is impossible by definition of $(A \cap B)$. This proves the result.

(*ii*) This can be proved that,

$$(x \in A \cap B) \Rightarrow (x \in B) \text{ is true for all } x.$$

The only doubtful case is when $(x \in A \cap B)$ is true and $(x \in B)$ is false. This is not possible according to definition of $A \cap B$. Hence proved.

NOTES

Example 20. If $A \subseteq B$ and $A \subseteq C$, then $A \subseteq (B \cap C)$

Solution. Let $x \in A$. Then $x \in B$ and $x \in C$ (as $A \subseteq B$ and $A \subseteq C$).
So, $x \in B \cap C$.

This proves that $A \subseteq B \cap C$.

Aliter. This can be proved that,

$$(x \in A) \Rightarrow (x \in B \cap C) \text{ is true for all } x.$$

The only doubtful case is when $(x \in A)$ is true and $(x \in B \cap C)$ is false.

Now $(x \in A)$ is true means $(x \in B)$ is also true (as $A \subseteq B$). Also $(x \in C)$ is true (as $A \subseteq C$). This means $(x \in B) \cap (x \in C)$ is true and therefore $(x \in B \cap C)$ is true. This proves the result.

Example 21. $A \cup B = A \cap B$ if and only if $A = B$.

Solution. Suppose $A \cup B = A \cap B$. Let $x \in A$. Then $x \in A \cup B$ and so $x \in A \cap B$. Therefore, $x \in B$. This proves that $A \subseteq B$. Similarly $B \subseteq A$ and hence $A = B$.

Aliter. Suppose $A \cup B = A \cap B$

According to Adsorption law,

$$\begin{aligned}
(x \in A) &\equiv (x \in A) \cap [(x \in A) \cup (x \in B)] \\
&\equiv (x \in A) \cap [x \in A \cup B] \\
&\equiv (x \in A) \cap [x \in A \cap B] \\
&\equiv [(x \in A) \cap (x \in A)] \cap (x \in B) \\
&\equiv (x \in A) \cap (x \in B) \\
&\equiv (x \in A \cap B) \\
&\equiv (x \in A \cup B) \\
&\equiv (x \in B) \cup (x \in A) \\
&\equiv (x \in B) \cup [(x \in A) \cup (x \in A)] \\
&\equiv (x \in B) \cup [x \in B \cup A] \\
&\equiv (x \in B) \cup [x \in A \cap B] \\
&\equiv (x \in B) \cup [(x \in A) \cap (x \in B)] \\
&\equiv [(x \in B) \cup (x \in A) \cap (x \in B) \\
&\equiv (x \in B) \text{ Adsorption law}
\end{aligned}$$

This proves that $A = B$.

Conversely, if $A = B$, then

$$\begin{aligned}
(x \in A \cup B) &\equiv (x \in A) \cup (x \in B) \\
&\equiv (x \in B) \cup (x \in B) \\
&\equiv (x \in B) \\
&\equiv (x \in B) \cap (x \in B) \\
&\equiv (x \in A) \cap (x \in B) \\
&\equiv (x \in A \cap B)
\end{aligned}$$

Note: Adsorption law in logic means that,

(*i*) $p \cap (p \cup r) \equiv p$ (*ii*) $p \cup (p \cap r) \equiv p$

NOTES

Complements

If A and B are two sets then complement of B relative to A is the set of all those elements $x \in A$ such that $x \notin B$ and is denoted by $A - B$. Logically speaking, if for a set C the biconditional statement $(x \in C) \Leftrightarrow (x \in A) \cap (x \notin B)$ is true for all x, then

$C = A - B$. In other words, if $(x \in C) \equiv (x \in A) \wedge (x \notin B)$ then C is called the complement of B relative to A.

Notes: 1. It is proved from the above definition that $A - B$ is a subset of A.

2. Whenever we say complement of B we mean complement of B relative to the universal set U. In such cases, we denote complement of B by B'.

So, $B' = U - B$.

Example 22. Show that $A - B = A \cap B'$.

Solution. Let $x \in A - B$. This means $x \in A$ and $x \notin B$. By definition of the universal set $A - B \subseteq U$. So, $x \in U$. Therefore $x \in U, x \notin B$, implies that $x \in B'$. This proves that $A - B \subseteq A \cap B'$. Again if $x \in A \cap B'$, then $x \in A$ and $x \in B'$. Now $x \in B'$ implies that $x \notin B$. So $x \in A - B$. This proves that $A \cap B' \subseteq A - B$.

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

Aliter. $(x \in A - B) \equiv (x \in A) \cap (x \notin B)$

$$\begin{aligned} &\equiv (x \in A \cap U) \cap (x \notin B) \text{ as } A \cap U = A \\ &\equiv [(x \in A) \cap (x \in U)] \cap (x \notin B) \\ &\equiv [(x \in A) \cap [(x \in U) \cap (x \notin B)] \\ &\equiv [(x \in A) \cap (x \in B')] \end{aligned}$$

This proves that $A - B = A \cap B'$.

Example 23. Prove that $A \subseteq B$ if and only if $B' \subseteq A'$.

Solution. Suppose $A \subseteq B$. Let $x \subseteq B'$. Then $x \in U$ and $x \notin B$. Now $x \notin B$ implies that $x \notin A$ (as $A \subseteq B$). Therefore $x \in U$ and $x \notin A$ implies that $x \in A'$. This proves that $B' \subseteq A'$. Conversely, let $B' \subseteq A'$. Let $B \in A$. Then $x \notin A'$. Now $x \notin A'$ implies that$x \notin B'$ (as $B' \subseteq A'$). This means that $x \in B$. Hence, $A \subseteq B$.

Aliter. Now $(x \in A) \Rightarrow (x \in B)$

$$\begin{aligned} &\equiv \sim (x \in B) \Rightarrow \sim (x \in A) \qquad \text{(by Contrapositive law in logic)} \\ &\equiv (x \notin B) \Rightarrow (x \notin A) \\ &\equiv (x \in B') \Rightarrow (x \in A') \end{aligned}$$

Suppose $A \subseteq B$. Then $(x \in A) \Rightarrow (x \in B)$ is true for all x. This is proved that, $(x \in B') \Rightarrow (x \in A')$ is true for all x. This means $B' \subseteq A'$. Conversely, suppose $B' \subseteq A'$. Then $(x \in B') \Rightarrow (x \in A')$ is true for all x. This is proved that $(x \in A) \Rightarrow (x \in B)$ is true for all x. This implies that $A \subseteq B$. Hence proved.

NOTES

Algebra of Sets

The following are some of the important laws of sets.

1. **Law of Idempotence.** For any set A,

$$A \cup A = A \text{ and } A \cap A = A$$

2. **Commutative Law.** For any sets A and B,

$$A \cup B = B \cup A, A \cap B = B \cap A$$

3. **Associative Law.** For any three sets A, B, C,

(*i*) $A \cup (B \cup C) = (A \cup B) \cup C$

(*ii*) $A \cap (B \cap C) = (A \cap B) \cap C$

Proof. (*i*) It can be proved that,

$$[x \in A \cup (B \cup C)] \Leftrightarrow [x \in (A \cup B) \subset C] \text{ is true for all } x.$$

Now by definition,

$$[x \in A \cup (B \cup C)] \equiv [(x \in A) \cup \{(x \in B) \cup (x \in C)\}]$$

and $\quad [x \in (A \cup B) \cup C] \equiv [\{(x \in A) \cup (x \in B)\} \cup (x \in C)]$

Hence, as per the associative law in logic, the result of case (*i*) follows.

Similarly, you can prove case (*ii*). The proof (*ii*) is left as an exercise.

Distributive Laws

For any three sets A, B, C,

(*i*) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(*ii*) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Proof. (*i*) Let $x \in A \cap (B \cup C)$. This implies that $x \in A$ and $x \in B \cup C$. Now, $x \in B \cup C$ implies that $x \in B$ or $x \in C$ or $x \in$ both B and C. If $x \in B$, then $x \in A \cap B$. If $x \in C$, then $x \in A \cap C$. In any case $x \in (A \cap B) \cup (A \cap C)$.

So, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Similarly $(A \cap B) \cup (A \cap C) = A \cap (B \cup C)$. This proves case (*i*).

Similarly, we can prove case (*ii*).

Aliter. $x \in [A \cap (B \cup C)] \equiv [(x \in A) \cap (x \in B \cup C)]$

$$\equiv [(x \in A) \cap \{(x \in B) \vee (x \in C)\}]$$

$$\equiv [(x \in A) \cap (x \in B)] \vee [(x \in A) \subset (x \in C)]$$

(by Distributive Law of logic)

$$\equiv [x \in (A \cap B)] \cup [x \in (A \cap C)]$$

$$\equiv [x \in (A \cap B) \cup (A \cap C)]$$

So, $\quad A \wedge (B \cup C) = (A \cap B) \cup (A \cap \text{C})$

Similarly, we can prove case (*ii*) by laws of logic.

NOTES

De Morgan's Laws

For any two sets A and B,

(i) $(A \cup B)' = A' \cap B'$

(ii) $(A \cap B)' = A' \cup B'$

Proof. (i) Let $x \in (A \cup B)'$. This implies that $x \notin A \cup B$ and $x \in U$. Now $x \notin A \cup B$ implies that $x \notin A$ and $x \notin B$. But $x \notin A$ and $x \in U$ implies that $x \in A'$ and $x \notin B$ and $x \in U$ implies that $x \in B'$. Therefore, $x \in A' \cap B'$ and so $(A \cup B)' = A' \cap B'$. Similarly $(A' \cap B') = (A \cup B)'$.

This proves that $(A \cup B)' = A' \cap B'$

Alternative proof of case (i) using logic:

$$\begin{aligned}
\text{Now } x \in (A \cup B)' &\equiv \sim[(x \in (A \cup B)] \\
&\equiv \sim [(x \in A) \cup (x \in B)] \\
&\equiv \sim (x \in A) \cap \sim (x \in B) \\
&\equiv (x \notin A) \cap (x \notin B) \\
&\equiv (x \in A') \cap (x \in B') \\
&\equiv (x \in A' \cap B')
\end{aligned}$$

Therefore, $(A \cup B)' = A' \cap B'$

The proof of case (ii) is left as an exercise.

Example 24. Let A, B, C be any three sets. Prove that,

$$A \cap (B - C) = (A \cap B) - (A \cap C)$$

Solution. $(A \cap B) - (A \cap C) = (A \cap B) \cap (A \cap C)'$

$$= (A \cap B) \cap (A' \cup C')$$

by De Morgan's law

$= [(A \cap B) \cap A'] \cup [(A \cap B) \cap C']$ by Distributive law

$\equiv [(A \cap A') \cap B] \cup [(A \cap B) \cap C']$ by Associative law

$\equiv [\phi \cap B] \cup [A \cap (B \cap C')]$

$\equiv \phi \cup [A \cap (B \cap C')]$

$$\equiv A \cap (B \cap C')$$

$$\equiv A \cap (B - C)$$

Example 25. For any sets A and B, show that,

$$(A - B) \cup (B - A) = (A \cup B) - (A \cap B)$$

Solution. $(A \cup B) - (A \cap B) = (A \cup B) \cap (A \cap B)'$

$$= (A \cup B) \cap (A' \cup B')$$

By De Morgan's law,

$$= [(A \cup B) \cap A'] \cup [(A \cup B) \cap B']$$

By Distributive law,

$$= [(A \cap A') \cup (B \cap A')] \cup [(A \cap B') \cup (B \cap B')]$$

$$= [\phi \cup (B \cap A')] \cup [(A \cap B') \cup \phi]$$

$$= (B \cap A') \cup (A \cap B')$$
$$= (B - A) \cup (A - B)$$
$$= (A - B) \cup (B - A)$$

By Commutative law,

NOTES

Finite Sets

If A is a finite set, then we shall denote the number of elements in A by $n(A)$. If A and B are two finite sets, then it is very clear from the Venn diagram of $A - B$ that,

$$n(A - B) = n(A) - n(B \cap A)$$

Suppose A and B are two finite sets such that $A \cap B = \phi$. Then, the number of elements in $A \cup B$ is the sum of number of elements in A and B the number of elements in B.

i.e., $\quad n(A \cup B) = n(A) + n(B)$ if $A \cap B = \phi$

For example, to find the number of elements in $A \cup B$, in case $A \cap B \neq \phi$, can be proved as follows:

For any two sets A and B,

$$A \cup B = A \cup (B - A)$$

Here, $\quad A \cap (B - A) = \phi$

Therefore, $\quad n(A \cup B) = n(A) + n(B - A)$

$$= n(A) + n(B) - n(A \cap B)$$

Note: According to the definition of empty set it follows that $n(\phi) = 0$.

Therefore, if A and B are two finite sets, then,

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

Similarly, if A, B, C are three finite sets, then,

$$n(A \cup B \cup C) = n(A \cup B) + n(C) - n[(A \cup B) \cap C]$$
$$= n(A) + n(B) - n(A \cap B) + n(C) - n[(A \cup B) \wedge C]$$
$$= n(A) + n(B) + n(C) - n(A \cap B) - n[(A \cap C) \cup (B \cap C)]$$
$$= n(A) + n(B) + n(C) - n(A \cap B) - [n(A \cap C)]$$
$$+ n(B \cap C) - n[(A \cap C) \wedge (B \cap C)]$$
$$= n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C)$$
$$+ n(A \cap B \cap C) \text{ as } A \cap C \cap B \cap C = A \cap B \cap C$$

These two results are used in the following problems.

Example 26. In a recent survey of 400 students in a school, 100 were listed as smokers and 150 as chewers of gum; 75 were listed as both smokers and gum chewers. Find out how many students are neither smokers nor gum chewers.

Solution. Let U be the set of students questioned. Let A be the set of smokers, and B the set of gum chewers.

Then, $n(U) = 400$, $n(A) = 100$, $n(B) = 150$, $n(A \cap B) = 75$

We have to find out $n(A' \cap B')$

NOTES

Now, $A' \cap B' = (A \cup B)' = U - (A \cup B)$

Therefore,

$$n(A' \cap B') = n\,[U - (A \cup B)]$$
$$= n(U) - n[(A \cup B) \cap U]$$
$$= n(U) - n(A \cup B)$$
$$= n(U) - n(A) - n(B) + n(A \cap B)$$
$$= 400 - 100 - 150 + 75$$
$$= 225$$

Example 27. Out of 500 car owners investigated, 400 owned Fiat cars and 200 owned Ambassador cars; 50 owned both Fiat and Ambassador cars. Is this data correct?

Solution. Let U be the set of car owners investigated. Let A be the set of those persons who own Fiat cars and B the set of persons who own Ambassador cars; then $A \cap B$ is the set of persons who own both Fiat and Ambassador cars.

$$n(U) = 500,\ n(A) = 400,\ n(B) = 200,\ n(A \cap B) = 50$$

Therefore, $n(A \cup B) = n(A) + n(B) - n(A \cap B)$
$$= 400 + 200 - 50 = 550$$

This exceeds the total number of car owners investigated.

So, the given data is not correct.

Example 28. A market research group conducted a survey of 1000 consumers and reported that 720 consumers liked product A and 450 consumers liked product B. What is the least number that must have liked both the products?

Solution. Let, $U =$ Set of consumers questioned

$S =$ Set of consumers who liked product A

$T =$ Set of consumers who liked product B

Then, $S \cap T =$ Set of consumers who liked both the products A and B

Now, $n(U) = 1000,\ n(S) = 720,\ n(T) = 450$

Therefore, $n(S \cup T) = n(S) + n(T) - n(S \cap T)$

$= 1170 - n(S \cap T)$

So, $n(S \cap T) = 1170 - n(S \cup T)$

Now, $n(S \cap T)$ is least when $n(S \cup T)$ is maximum. But $S \cup T \subseteq U$ implies that $n(S \cup T) \subseteq n(U)$.

This implies that maximum value of $n(S \cup T)$ is 1000.

So, least value of $n(S \cap T) = 170$

Hence, the least number of consumers who liked both the products A and B is 170.

Example 29. Out of 1000 students who appeared for C.A. Intermediate Examination, 750 failed in Maths, 600 failed in Accounts and 600 failed in Costing, 450 failed in both Maths and Accounts, 400 failed in both Maths and Costing,

NOTES

150 failed in both Accounts and Costing. The students who failed in all the three subjects were 75. Prove that the above data is not correct.

Solution.

Let, U = Set of students who appeared in the examination.

A = Set of students who failed in Maths.

B = Set of students who failed in Accounts.

C = Set of students who failed in Costing.

Then, $A \cap B$ = Set of students who failed in both Maths and Accounts.

$B \cap C$ = Set of students who failed in both Accounts and Costing.

$A \cap C$ = Set of students who failed in both Maths and Costing.

$A \cap B \cap C$ = Set of students who failed in all the three subjects.

Now, $n(U) = 1000$, $n(A) = 750$, $n(B) = 600$, $n(C) = 600$, $n(A \cap B) = 450$,

$n(B \cap C) = 150$, $n(A \cap C) = 400$, $n(A \cap B \cap C) = 75$

Therefore, $n(A \cup B \cup C) = 750 + 600 + 600 - 450 - 150 - 400 + 75$

$= 1025$

This exceeds the total number of students who appeared in the examination.

Hence, the given data is not correct.

Example 30. A factory inspector examined the defects in hardness, finish and dimensions of an item. After examining 100 items he gave the following report:

All three defects 5, defects in hardness and finish 10, defects in dimension and finish 8, defects in dimension and hardness 20. Defects in finish 30, in hardness 23, in dimension 50. The inspector was fined. Why?

Solution. Suppose H represents the set of items which have defect in hardness, F represents the set of items which have defect in finish and D represents the set of items which have defect in dimension.

Then, $n(H \cap F \cap D) = 5$, $n(H \cap F) = 10$, $n(D \cap F) = 8$

$n(D \cap H) = 20$, $n(F) = 30$, $n(H) = 23$, $n(D) = 50$

So, $n(H \cup F \cup D) = 30 + 23 + 50 - 20 - 10 - 8 + 5 = 70$

Now, $n(D \cup F) = n(D) + n(F) - n(D \cap F)$

$= 50 + 30 - 8 = 72$

$D \cup F \subseteq D \cup F \cup H$ implies that $n(D \cup F) \leq n(D \cup F \cup H)$, i.e., $72 \leq 70$

Hence, there is an error in the report and for this reason inspector was fined.

Example 31. In a survey of 100 families the numbers that read the most recent issues of various magazines were found to be as follows:

Readers Digest	28
Readers Digest and Science Today	8
Science Today	30
Readers Digest and Caravan	10

NOTES

Caravan	42
Science Today and Caravan	5
All the three Magazines	3

Using set theory, find

(*i*) How many read none of the three magazines?

(*ii*) How many read Caravan as their only magazine?

(*iii*) How many read Science Today if and only if they read Caravan?

Solution. Let, S = Set of those families who read Science Today

R = Set of those families who read Readers Digest

C = Set of those families who read Caravan

(*i*) Find $n(S' \cap R' \cap C')$

Let U = Set of the families questioned.

Now, $S' \cap R' \cap C' = (S \cup R \cup C)'$

$= U - (S \cup R \cup C)$

Therefore, $n(S' \cap R' \cap C') = n(U) - n(S \cup R \cup C)$

$= 100 - n(S \cup R \cup C)$

Now, $n(S \cup R \cup C) = 30 + 28 + 42 - 8 - 10 - 5 + 3 = 80$

So, $n(S' \cap R' \cap C') = 100 - 80 = 20$

(*ii*) Find $n[C - (R \cup S)]$

Now, $n[C - (R \cup S)] = n(C) - n[C \cap (R \cup S)]$

$= n(C) - n[(C \cap R) \cup (C \cap S)]$

$= n(C) - n(C \cap R) - n(C \cap S) + n(C \cap R \cap S)$

$= 42 - 10 - 5 + 3$

$= 30$

(*iii*) Find $n[(S \cap C) - R]$

Now, $n[(S \cap C) - R] = n(S \cap C) - n(S \cap C \cap R)$

$= 5 - 3 = 2$

Example 32. In a survey conducted of women it was found that,

(*i*) There are more single than married women in South Delhi.

(*ii*) There are more married women who own cars than unmarried women.

(*iii*) There are fewer single women who own cars and homes than married women who are without cars but own homes.

Is the number of single women who own cars and do not own homes greater than number of married women who do not own cars but own homes?

Solution. Let, A = Set of married women

B = Set of women who own cars

C = Set of women who own homes

NOTES

Then, the given conditions are,

(i) $n(A') > n(A)$

(ii) $n(A \cap B) > n(A' \cap B')$

(iii) $n(A \cap B' \cap C') > n(A' \cap B \cap C)$

Find $n(A' \cap B \cap C')$ and $n(A \cap B' \cap C)$

Let, U = Set of all women questioned.

Now, $A' = A' \cap U = A' \cap (B \cup B') = (A' \cap B) \cup (A' \cap B')$

$A = A \cap U = A \cap (B \cup B') = (A \cap B) \cup (A \cap B')$

So, $n(A') = n(A' \cap B) + n(A' \cap B')$

$n(A) = n(A \cap B) + n(A \cap B')$

According to case (i), we have,

$$n(A' \cap B) + n(A' \cap B') > n(A \cap B) + n(A \cap B')$$

Also, by case (ii) we have,

$$n(A' \cap B) + n(A' \cap B') > n(A \cap B) + n(A \cap B') > n(A' \cap B') + n(A \cap B')$$

Therefore,

$$n(A' \cap B) > n(A \cap B')$$

Also, $A' \cap B = (A' \cap B) \cap (C \cup C')$

$= (A' \cap B \cap C) \cup (A' \cap B \cap C')$

And $A \cap B' = (A \cap B') \cap (C \cup C')$

$= (A \cap B' \cap C) \cup (A \cap B' \cap C')$

So, $n(A' \cap B) = n(A' \cap B \cap C) + n(A' \cap B \cap C')$

$n(A \cap B') = n(A \cap B' \cap C) + n(A \cap B' \cap C')$

Using case (iii) we get,

$$n(A' \cap B \cap C) + n(A' \cap B \cap C') > n(A \cap B' \cap C) + n(A' \cap B \cap C)$$

i.e., $n(A' \cap B \cap C') > n(A \cap B' \cap C)$

So, the number of single women who own cars and do not own a home is greater than the number of married women who do not own cars but own homes.

1.3.1 Counting Principles

The probability of a successful outcome is calculated as the number of successful outcomes divided by the total number of possible outcomes. When the number of total outcomes is comparatively small we can list them all and this constitutes the entire sample space. This task becomes cumbersome when the number of possible outcomes is large. For such situations some counting methods have been developed, and this makes it easier to calculate the number of all possible outcomes of all events. For example, if we roll a die, we know that there are 6 possible outcomes, namely, 1, 2, 3, 4, 5, 6. When a second die is rolled, the number of possible outcomes for both dice together increases to $6 \times 6 = 36$. If the die is rolled four times, the number of possible outcomes becomes $6 \times 6 \times 6 \times 6 = 1296$. To solve such probability problems, the counting rule can be stated as follows:

If an event A can occur in n_1 ways and after its occurrence, event B can occur in n_2 ways, then both events can occur in a total of $n_1 \times n_2$ different ways in a given order of occurrences.

For example, if we toss a coin 3 times, there are two possible outcomes for each toss. Hence the total number of possible outcomes in 2 × 2 × 2 = 8. This can be illustrated as follows:

	Outcome
1	HHH
2	HHT
3	HTH
4	HTT
5	TTT
6	THH
7	THT
8	TTH

Accordingly, the fundamental counting principle can be expanded as follows:

If there are k separate parts of an experiment, and the first part can he done in n_1 ways, second successive part in n_2 ways .... and kth successive part in n_k ways, then the total number of possible outcomes is given by the following product:

$$n_1 \times n_2 \times........ \times n_k$$

1.3.2 Classes of Sets

Sets are fundamental objects aimed at defining all other concepts in mathematics. Sets are taken as something self-understood. It is a kind of standard concept with formal axioms. In earlier days 'class and set' were not considered different, as it is done now.

In set theory, a collection of set or some mathematical objects having similarity in property among members is said to form a class. A *class* in modern set theory speaks of an arbitrary *collection* of elements of the *universe*. Thus all sets are classes since these are collections of elements of the universe, but not all classes are sets. A proper class is not a set.

Set theory has its own language that defines the concept of membership. In real world, we deal with different kinds of sets. These may be sets composed of numbers, points, functions, or other sets.

Concept of class became more important after advent of computer science and various programming languages. Object oriented programming languages use classes and objects while defining functions that perform a particular task. They give a set of inputs that goes under a set of processes and gives a set of outputs.

Sets fall into two classes, *basic sets* which are typically simple sets which form a base for the topology, and *denotable sets* which are unions of basic sets.

NOTES

All geometrical shapes represent classes of sets. Like, a circle denotes a set that contains points in a plane or in three dimensional spaces whose distance from a fixed point is constant. A parabola is also a set containing points in a plane or in three dimensional spaces in which its distance from a fixed point known as focus is same as its distance from a line known as directrix. Similarily, other geometrical shapes like, ellipse, hyperbola, sphere, ellipsoid, paraboloid, etc., are defined in the language of sets. But here the set is not finite as it is not definite as how many points lie on these plane curves or three dimensional bodies.

The fundamental geometric operations are – contains (Set, State), disjoint (Set,Set) and subset(Set,Set). *Fuzzy basic sets*, which are sets of sets, defined using interval data, to store intermediate results if these cannot be computed exactly. These sets can be converted to ordinary basic sets by over or under approximation. Alternatively, the fundamental binary predicates can be computed directly for the fuzzy set types. Denotable sets are never fuzzy.

Basic sets are so-called because they form a base for the topology of the space. Typically, basic sets are (a subclass of) convex polytopes or ellipsoids. Basic sets must support the fundamental geometric predicates, both within the class and with the Rectangle class. Additionally, basic sets may support the optional geometric operations, but only if the class of set under consideration is closed under that operation. The result may be exactly computable if it involves no arithmetic (for example, intersection of two rectangles) or may need to be represented by a fuzzy set (for example, Minkowski sum of two rectangles).

Denotable Sets: A denotable set implements a set as a union of basic sets. The specification of the denotable set concept is given by the class followed by a **::** and name of the set that it contains.

Predicates and Operations on Sets

Predicates on Sets

The fundamental geometric predicates are:

- contains(): test whether a set contains a point.
- disjoint(): test whether two sets are disjoint.
- subset(): test whether a set is a subset of another.
- intersects(): test whether two sets intersect.
- superset(): test whether the second set is a subset of the first.

All these predicates return a fuzzy logic value, since the result of the test may be impossible to determine at the given precision.

1.3.3 Power Sets

Power set is a class of sets which is a collection all the subsets that is formed by member of a set denoted by A or any other letter. Number of power sets that can be formed is given by $2n$ where n is the cardinality of the set, i.e., number of

NOTES

members of the set. For example if a set contains three members then number of subsets that it can form is 23 = 8 subsets. There are 8 subsets in a power set of set A when #(A), n(A) or |A| = 3.

For example, let there be a set A = {1,2,3} then the power set contains subsets {{1,2,3}, {3,1}, {2,1}, {1,3}, {1,2}, {1}, {2}, {3}, { } }. Here, null set in which there is no element and its universal set in which all the elements are present, both are present.

Fuzzy Sets: This kind of set has varying degree of membership. It was proposed in 1965 by L.A. Zadeh and it is an extension of the conventional notation of set. In conventional set theory an element is either in a set or not in a set, but in a fuzzy set there is a grade of association.

If there is a set $S = \{x_1, \ldots\ldots\ldots, x_n\}$, the fuzzy set ($S$, m) is shown as

$\{m(x_1)/x_1, \ldots\ldots\ldots, m(x_n)/x_n\}$.

If $m(x) = 0$ for any x then it means that the member is not in the set and if

$m(x) = 1$, shows full membership in the fuzzy set. Any value in-between 0 to 1 shows varying degree of association of members of the fuzzy set.

Operation on Fuzzy Sets

Union: Union of two fuzzy sets S_1 and S_2 having membership function (S_1) and (S_2) is given by max((S_1), (S_2)). This operation resembles OR operation in Boolean algebra.

Intersection: Intersection of two fuzzy sets S_1 and S_2 having membership function (S_1) and (S_2) is given by min((S_1), (S_2)). This operation resembles AND operation in Boolean algebra.

Complement: This operation is the negation of the specified membership function and shows the negation criterion. This operation is like NOT operation in Boolean algebra.

Rules which are common in classical set theory also apply to fuzzy set theory.

Important Terms associated with Fuzzy Set

Universe of Disclosure: It is the range of input values for a fuzzy logic.

Fuzzy Set: A set that allows degree of association from 0 to 1. Zero (0) shows no association and 1 shows full association.

Fuzzy Singleton: It is a fuzzy logic having single point with membership of 1.

Example 33. Classification of dwelling units.

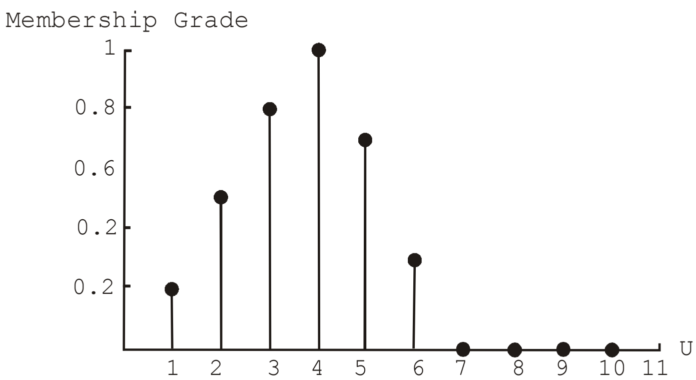
Problem Statement. A builder wants to classify the flats that he is building and intending to sell to home seekers. Level of comfort is given by number of bedrooms in a flat. If U represent the set of those available flats and is given as:

NOTES

$U = \{x \mid x \in [0, 1] \in I\}$. Flats are denoted by '*u*' number of rooms in a dwelling unit. Builder gives a comfort level for 'a family of four'.

Solution. A comfortable flat for a 'family of four' is described by a fuzzy set as given below:

```
    : FlatForFour
    = FuzzySet[{{1, 0.2}, {2, 0.5}, {3, 0.8}, {4, 11},
        {5, 0.7}, {6, 0.3}}, UniversalSpace → {1, 10}]
        Fuzzyplot[FlatFour, ShowDots→True]
```



Example 34. Problem on age. Range is from 0 to 100.

Problem Statement. Here fuzzy set is used for representing age that ranges from 1 to 100.

Solution. This can be discussed with the help of fuzzy sets and for that we set the universal space for age to have a range from 0 to 100.

```
SetOptions[FuzzySet, UniversalSpace→{0, 100}]
```

Example 35. Problem on youth age. Range in from 0 to 40.

Problem Statement. Represent concept of youth by a fuzzy set.

Solution. Defined as a fuzzy set.

```
Youth = FuzzyTrapezoid[0, 25, 40]
```

In a similar way, the property of 'being old' can also be given as a fuzzy set. This is as below:

```
Old = FuzzyTrapezoid[50, 65, 100, 100]
```

NOTES

Example 36. This example shows the operation 'intersection' on fuzzy sets.

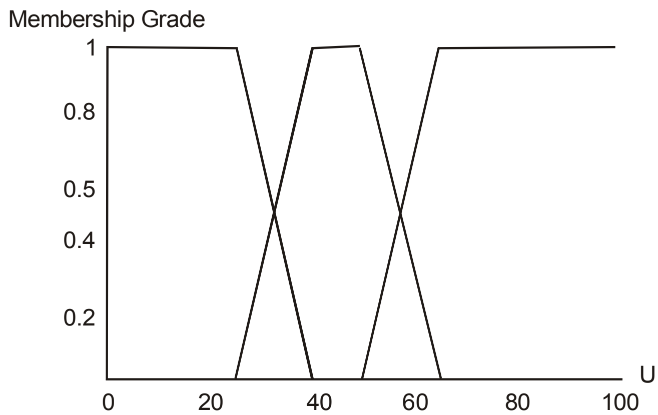
Problem Statement. Define the concept of 'middle-aged' using fuzzy set.

Solution. A middle age means 'not old' OR 'not young'. This requires use of operators NOT, OR. Thus a middle age is found by 'not old' (Complement of old) OR (Disjunction) 'not young' (Complement of youth) as in example 33(a) and 33(b).

```
Middle-Aged = Intersection[Complement[Young],
              Complement[Old]];
```

We can also define and operation FuzzyPlot to find a graphical presentation of age, performing an operation which named as FuzzyPlot.

```
FuzzyPlot[Young, Middle-Aged, Old,
        PlotJoined→True];
```



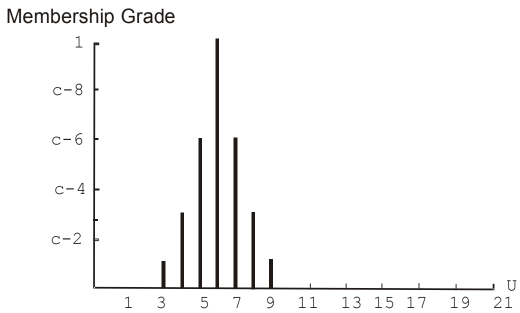
The graph shows that the intersection of 'not young' and 'not old' gives a reasonable definition for the concept of 'middle-aged.'

Example 37. Fuzzy set with natural numbers.

Problem Statement. To define a set of natural numbers in the neighbourhood of 6. This can be done in different ways.

Solution. Define a fuzzy set of number adjacent to 6.

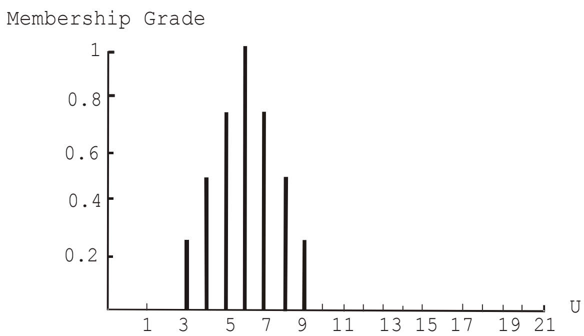
```
SetOptions[FuzzySet, UniversalSpace→{0, 20}];
Six1 =
FuzzySet[{{3, 0.1}, {4, 0.1 {4,0.3} {5, 0.6},
{6,1.0}, Number's closeness to 6.
{7, 0.6}, {8, 0.3}, {9, 0.1}}]
FuzzyPlot[Six1];
```



Solution. We use a function FuzzyTrapazoid and create a fuzzy set. If we think that triangular fuzzy set is perhaps a better choice then we have to set two parameters in the middle of the set as 6 which would tell about the closeness to number 6. We also define a set Six2 which shows a set of,

```
Six2=FuzzyTrapezoid[2, 6, 6, 10];
```

`FuzzyPlot[Six2];` This shows a set of points according this definition.

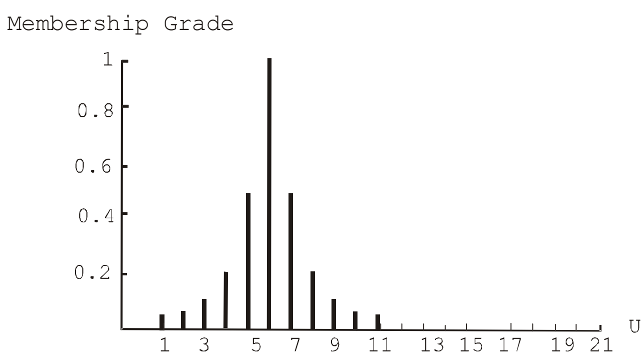


Solution. There may be a third solution which creates a fuzzy set defining nearness to 6.

$$\text{CloseTo[x]: } = \frac{1}{1+(1-x)^2}$$

We name it Six3.

```
Six3 = CreateFuzzySet[CloseTo[6]];
       FuzzyPlot[Six3];
```



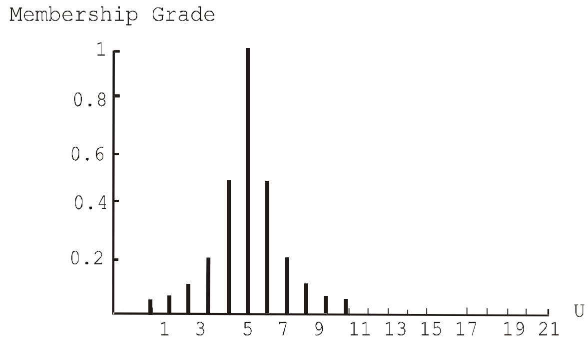
NOTES

Solution. There may be a fourth solution which uses a 'piece-wise defined function'.

```
NearSix[x] := which[x == 6, 1, x > 6 & &x < 12,
```

$$\frac{1}{(x-5)^2}, x < 6 \;\&\;\& > 0, \frac{1}{(7-x)^2}, \text{True}, \;0]$$

```
Six4 = CreateFuzzySet [NearSix];
FuzzyPlot [Six4];
```



Example 38. Problem on disjunctive sum.

Problem. To find,

RMat = {{.8, .3, .5, .2}, {.4, 0, .7, .3}, {.6, .2, .8, .6}}

This shows a one membership matrix.

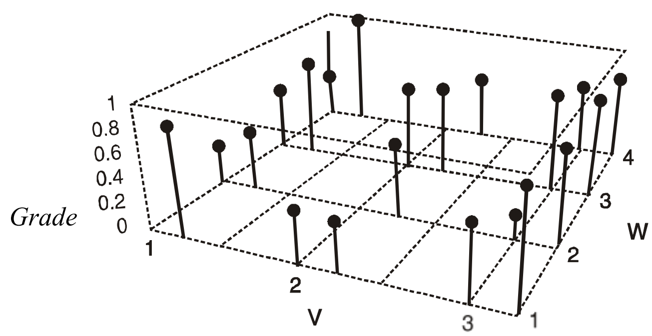
SMat = {{.9,.5,.8, 1}, {.4,.6,.7, .5}, {.7,.8,.8,.7}}

This shows another membership matrix.

R = FromMembershipMatrix[Rmat]; This defines relation in RMat

S = FromMembershipMatrix[SMat]; This defines relation in SMat.

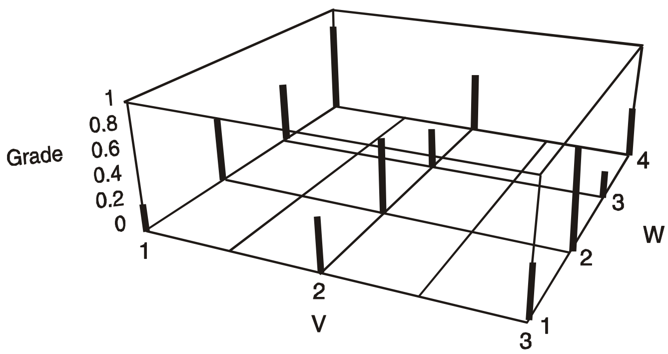
FuxxyPlot 3D[R,S, ShowDots→True] This shows location of point.



Let there be two fuzzy relationships R and S in the universal space V × W. The disjunctive sum is given by: DisSum = (R ∩ S') ∪ (R' ∩ S), which is a relation in V × W, with the following property:

NOTES

```
    ∀v, w ∈ V×W, and
  DisSum(v, w)= Maximum of (Min(R(v, w), 1 - S(v,
w)),
  Min(1 - R(v, w), S(v, w)))
  DisSum = Union[Intersection[R, Complement[s]],
       Intersection[Complement[R],S]];
  FuzzyPlot3D[DisSum];
```




```
ToMembershipMatrix[DisSum] // MatrixForm
```

$$\begin{pmatrix} 0.2 & 0.5 & 0.5 & 0.8 \\ 0.4 & 0.6 & 0.3 & 0.5 \\ 0.4 & 0.8 & 0.2 & 0.4 \end{pmatrix}$$

Example 39. Problem on distance.

Problem Statement. We take a fuzzy relation R on sets *X, Y*. We define these sets as: *X* = {Delhi, Moscow} and *Y* = {Dacca, Delhi, London). The relation R is representing the idea of being 'far'. The relation may be shown as R(*X,Y*) = 1.0/ Delhi, Dacca + 0/Delhi, Toronto + 0.6/Delhi, London + 0.9/Moscow, Delhi + Dacca + 0.9/London, Dacca + 0.7/London, Delhi + 0.3/London.

Solution. Representation of such a fuzzy relation is as below: A membership matrix is to be created to depict the relationship. For this we represent by number for each city in the set. For *X* we keep 1 for Delhi and 2 for Moscow. In *Y* we set 1 for Dacca, 2 for Delhi and 3 for London.

After this, we are in a position to create relation using a function by name from memebrship matrix and the relationship can be plotted using a function that is named FuzzyPlot3D.

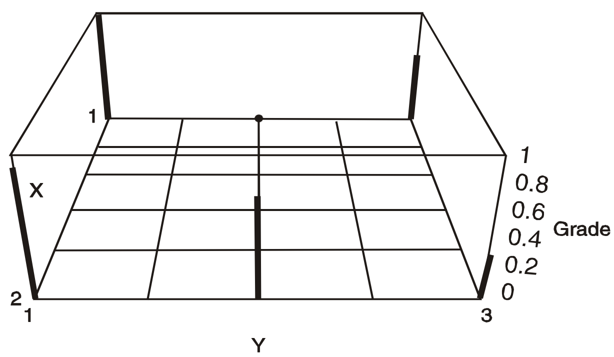
```
DistMat = {{1,0,0,6}, {0.9,0.7,0.3}}

{{1,0,0.6}, {0.9,0.7,0.3}}

{{1, 0, 0.6}, {0.9, 0.7, 0.3}}
```

NOTES

```
DistRel = FromMembershipMatrix[DistMat,
{{1,2}, {1,3}}]
FuzzyRelation[
{{{1,1},1},{{1,2},0}{{1,3},0.6
   {{2,1},0.9},{{2,2},0.7,{{2,3}0.3}},
UniversalSpace→{{1,2,1},{1,3,1}}]
```

This relation can be plotted using the FuzzyPlot3D function.

```
Fuzzy Plot3D[DistRel,
Axes Label→{" X", "Y", "Grade"],
View Point→{2, 0, 1},
Axxes Edge→{-1, -1}, {1, -1}, {1, 1}}];
```




```
ToMembershipMatrix[DisRel] // MatrixForm
```

$$\begin{pmatrix} 1 & 0 & 0.6 \\ 0.9 & 0.7 & 0.3 \end{pmatrix}$$

Example 40. To choose a job, Fuzzy sets can help to choose between four given jobs. Let this job be numbered as 1 to 4.

Problem Statement. We have to make selection such that the job provides best salary and at the same time it should be near to place of stay.

Solution. The first part of selection criteria is given by following definition of fuzzy sets. From this selection criterion Job 3 looks most attractive out of all these four jobs and Job 1 is least attractive. Similarly, a fuzzy set can be created for the second part of the selection criteria. We create a set named drive for the distance from the working place. Here membership has varying grade depending on distance. Here, least is desired.

```
Interest =
 FuzzySet[{{1, .4}, {2, .6}, {3, .8}, {4, .6}},
  UniversalSpace → {1, 4}]
```

NOTES

```
FuzzySet[{{1, 0.4}, {2, 0.6}, {3, 0.8}, {4, 0.6}},
 UniversalSpace → {1, 4, 1}]
```

From analysis we find that from the second criterion Job 4 is most attractive as it is nearest among the given four and Job 1 is the farthest.

As the goal is to get a good salary, we find that Job 1 has highest and Job 4 is the lowest.

```
Drive =
 FuzzySet[{{1, .1}, {2, .9}, {3, .7}, {4, 1}},
  UniversalSpace → {1, 4}]
```

```
FuzzySet[{{1, 0.1}, {2, 0.9}, {3, 0.7}, {4, 1}},
 UniversalSpace → {1, 4, 1}]
```

```
Salary =
 FuzzySet[{{1, .875}, {2, .7}, {3, .5}, {4, .2}},
  UniversalSpace → {1, 4}]
```

```
FuzzySet[
 {{1, 0.875}, {2, 0.7}, {3, 0.5}, {4, 0.2}},
 UniversalSpace → {1, 4, 1}]
```

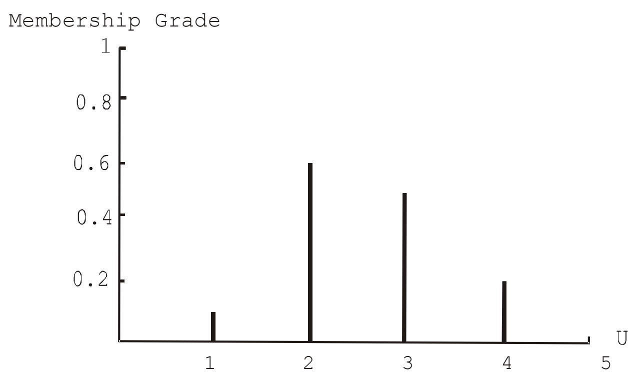
We have examined all criteria one-by-one using fuzzy sets, but have made anything to make a final decision. So a decision function is being defined using intersection. By applying intersection operation betwen the constraints and goals would give best decision. Plotting the fuzzy set for decision we can visualize the result graphically. Considering all these the decision function says that Job 2 is the best.

```
Decision = Intersection[Interest, Drive, Salary]
```

```
FuzzySet[{{1, 0.1}, {2, 0.6}, {3, 0.5}, {4, 0.2}},
 UniversalSpace → {1, 4, 1}]
```

We can plot the decision fuzzy set to see the results graphically.

```
FuzzyPlot[Decision];
```



NOTES

Power Set

The set of all subsets of a given set A is called the *power set* of A and is denoted by $P(A)$. The name power set is motivated by the fact that 'if A has n elements then its power set $P(A)$ contains exactly 2^n elements.'

Example 41. If $A = \{1, 2\}$, find $P(A)$.

Solution. Now ϕ is a subset of A. A is also a subset of A. $\{1\}$ and $\{2\}$ are also subsets of A. Therefore, these are all subsets of A. So, $P(A) = [\phi, \{1\}, \{2\}, A]$. Therefore $P(A)$ has $2^2 = 4$ elements.

Example 42. Let $A = \{1, 2, 3\}$. Find $P(A)$.

Solution. Now subsets consisting of one element only are $\{1\}, \{2\}, \{3\}$. Subsets consisting of two elements only are, $\{1, 2\}, \{2, 3\}, \{1, 3\}$. Also ϕ and A are subsets of A.

So $P(A) = [\phi, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, A]$ and the number of elements in $P(A)$ is $2^3 = 8$.

Example 43. Let B be a subset of A. Let $P(A : B) = \{X \in P(A) \mid B \subseteq X\}$. If $B = \{1,2\}$ and $A = \{1, 2, 3, 4, 5\}$, list all the elements of $P(A : B)$.

Solution. Clearly $B \subseteq \{1, 2\}$, $B \subseteq \{1, 2, 3\}$, $B \subseteq \{1, 2, 3, 4\}$, $B \subseteq \{1, 2, 3, 4, 5\}$, $B \subseteq \{1, 2, 4\}$, $B \subseteq \{1, 2, 5\}$, $B \subseteq \{1, 2, 3, 5\}$, $B \subseteq \{1, 2, 4, 5\}$. These give all the elements of $P(A : B)$.

Duality

Union $\cup$ and Intersection $\cap$ of sets are termed as dual operations.

If the validity of a law involving are of the two, $\cup$ or $\cap$ is established, then the dual obtained by replacing $\cup$ by $\cap$ and $\cap$ by $\cup$ is also established.

Partition of a Set

The partition of a set A is written as,

$$A = \{A_1, A_2, ..., A_n\}$$

Where $A_j \subseteq A \quad j = 1, 2, ..., n$ or A_j's are inclusive.

Thus, (*i*) $A_1, A_2, ..., A_n$ are subsets of A

(*ii*) $A_j \cap A_k = \phi \qquad j = 1, 2, ..., n, \qquad k = 1,2, ..., n$

i.e., Any A_j, A_k are disjoints.

(*iii*) $A_1 \cup A_2 \cup ... \cup A_n = A$, i.e., $A_1, A_2, ..., A_n$ are exhaustive.

Thus every elements of A is a member of one and only one of the subsets in the partition.

Any sample S can be written as,

$$S = \{A, \overline{A}\}$$

$$= \{A \cap B, A \cap \overline{B}, \overline{A} \cap B, \overline{A} \cap \overline{B}\}$$

In any exercise, if (*i*), (*ii*), (*iii*) are all satisfied for a set $A = \{A_1, A_2, ..., A_n\}$ then this represents the partition of A.

Check Your Progress

7. Define the terms set and element.
8. Differentiate between finite set and infinite set.
9. What is singleton set?
10. Define the term subset.
11. What do you mean by null set?
12. What are Venn diagrams? Why are they called so?
13. What is union of sets?
14. Define intersection of sets.
15. What does complement of sets mean?
16. What is a power set?

NOTES

1.4 CONDITIONAL PROBABILITY AND INDEPENDENCE

In many situations, a manager may know the outcome of an event that has already occurred and may want to know the chances of a second event occurring based upon the knowledge of the outcome of the earlier event. We are interested in finding out as to how additional information obtained as a result of the knowledge about the outcome of an event affects the probability of the occurrence of the second event. For example, let us assume that a new brand of toothpaste is being introduced in the market. Based on the study of competitive markets, the manufacturer has some idea about the chances of its success. Now, he introduces the product in a few selected stores in a few selected areas before marketing it nationally. A highly positive response from the test-market area will improve his confidence about the success of his brand nationally. Accordingly, the manufacturer's assessment of high probability of sales for his brand would be conditional upon the positive response from the test-market.

Let there be two events A and B. Then the probability of event A given the outcome of event B is given by:

$$P[A/B] = \frac{P[AB]}{P[B]}$$

Where $P[A/B]$ is interpreted as the probability of event A on the condition that event B has occurred and $P\,[AB]$ is the joint probability of event A and event B, and $P[B]$ is not equal to zero.

NOTES

As an example, let us suppose that we roll a die and we know that the number that came up is larger than 4. We want to find out the probability that the outcome is an even number given that it is larger than 4.

Let, event A = Even

And event B = Larger than 4

Then, $P[\text{Even / Larger than 4}] = \dfrac{\text{P[Even and larger than 4]}}{\text{P[Larger than 4]}}$

Or,
$$P[A/B] = \frac{P[AB]}{P[B]} = \frac{\left(\frac{1}{6}\right)}{\left(\frac{2}{6}\right)} = 1/2$$

But for independent events, $P[AB] = P[A]P[B]$. Thus substituting this relationship in the formula for conditional probability, we get:

$$P[A/B] = \frac{P[AB]}{P[B]} = \frac{P[A]P[B]}{P[B]} = P[A]$$

This means that $P[A]$ will remain the same no matter what the outcome of event B is. For example, if we want to find out the probability of a head on the second toss of a fair coin, given that the outcome of the first toss was a head, this probability would still be 1/2 because the two events are independent events and the outcome of the first toss does not affect the outcome of the second toss.

1.4.1 Independent and Dependent Events

Two events A and B are said to be independent events, if the occurrence of one event is not influenced at all by the occurrence of the other. For example, if two fair coins are tossed, then the result of one toss is totally independent of the result of the other toss. The probability that a head will be the outcome of any one toss will always be 1/2, irrespective of whatever the outcome is of the other toss. Hence, these two events are independent.

Let us assume that one fair coin is tossed 10 times and it happens that the first nine tosses resulted in heads. What is the probability that the outcome of the tenth toss will also be a head? There is always a psychological tendency to think that a tail would be more likely in the tenth toss since the first nine tosses resulted in heads. However, since the events of tossing a coin 10 times are all independent events, the earlier outcomes have no influence whatsoever on the result of the tenth toss. Hence the probability that the outcome will be a head on the tenth toss is still 1/2.

On the other hand, consider drawing two cards from a pack of 52 playing cards. The probability that the second card will be an ace would depend upon whether the first card was an ace or not. Hence these two events are not independent events.

NOTES

Independent Repeated Trials

Probability is a measure of relative frequency. According to definition, *the probability of an event E is equal to the number of equally likely ways E can occur divided by the total number of equally likely things which can occur.*

For studying independent repeated trials, combinatorics is considered important when we consider that an event is consisting of repeated trials. Tossing a fair coin many times is an example of it. Suppose that a fair coin is tossed 10 times. Now the probability of the coin landing on heads for each toss is 1/2, since there are two possible equally-likely outcomes (heads or tails) and just one way it can come up heads. Furthermore, each toss of the coin is ***independent*** of every other toss. Basically this defines that the coin has no memory and the probability of the coin landing on heads for any given toss will be always 1/2. It has no relation with the outcome history of the previous tosses. Suppose the coin had just landed on heads 8 times in a row. Is it possible that the coin is more likely to land on tails on the next toss? The probability of it landing on heads is always 1/2, since the coin has no memory. It is true that in the long run, 50 per cent of the tosses will be heads and 50 per cent tails, but this is not achieved by the coin making up for any deficit of heads or tails but rather by turning up heads roughly half the time in all future tosses.

The fraction of successes depends on the probability of success in each trial, as the number of trials increases in repeated independent trials with the same probability of success. This is also known as the (Law of Large Numbers). According to this law if a discrete random variable is observed repeatedly in independent experiments then the fraction of experiments for which the random variable equals any of its possible values has the probability that the random variable equals that value.

Check Your Progress

17. Explain the concept of independent events.
18. Define binomial distribution.
19. What is a random variable?

1.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The term simple probability refers to a phenomenon where only a simple or an elementary event occurs. For example, assume that event (E), the drawing of a *diamond* card from a pack of 52 cards, is a simple event. Since there are 13 diamond cards in the pack and each card is equally likely to be drawn, the probability of event (E) or $P[E]$ = 13/52 or 1/4.

NOTES

The term joint probability refers to the phenomenon of occurrence of two or more simple events. For example, assume that event (E) is a joint event (or compound event) of drawing a black ace from a pack of cards. There are two simple events involved in the compound event, which are: the card being black and the card being an ace. Hence, P[Black ace] or $P[E] = 2/52$ since there are two black aces in the pack.

2. The classical theory of probability is the theory based on the number of favourable outcomes and the number of total outcomes. The probability is expressed as a ratio of these two numbers. The term 'favourable' is not the subjective value given to the outcomes, but is rather the classical terminology used to indicate that an outcome belongs to a given event of interest.
3. The addition rule states that when two events are mutually exclusive, then the probability that either of the events will occur is the sum of their separate probabilities. For example, if you roll a single dice then the probability that it will come up with a face 5 or face 6, where event A refers to face 5 and event B refers to face 6, both events being mutually exclusive events, is given by,

$$\begin{aligned} P[A \text{ or } B] &= P[A] + P[B] \\ \text{Or, } P[5 \text{ or } 6] &= P[5] + P[6] \\ &= 1/6 + 1/6 \\ &= 2/6 = 1/3 \end{aligned}$$

4. Multiplication rule is applied when it is necessary to compute the probability in case two events occur at the same time.
5. Bayes' theorem on probability is concerned with a method for estimating the probability of causes which are responsible for the outcome of an observed effect. The theorem contributes to the statistical decision theory in revising prior probabilities of outcomes of events based upon the observation and analysis of additional information.
6. Two events are said to be mutually exclusive, if both events cannot occur at the same time as the outcome of a single experiment. For example, if we toss a coin, then either event head or event tail would occur, but not both. Hence, these are mutually exclusive events.
7. 'A *set* is any collection of objects such that given an object, it is possible to determine whether that object belongs to the given collection or not.'

   The members of a set are called elements. We use capital letters to denote sets and small letters to denote elements. We always use { } brackets to denote a set.
8. A set which has finite number of elements is called a finite *set*, else it is called an infinite set.

NOTES

9. A set having only one element is called Singleton. If a is the element of the singleton A, then A is denoted by $A = \{a\}$.
10. Let A and B be two sets. If every element of A is an element of B, then A is called a subset of B and we write $A \subseteq B$ or $B \supseteq A$ (read as 'A is contained in B' or 'B contains A').
11. A set which has no element is called the *null set* or *empty set.* It is denoted by the symbol ϕ.
12. Venn diagrams are used to illustrate various set operations. It is named after John Venn (1834–1883).
13. The union of any two sets A and B is the set of all those elements x such that x belongs to at least one of the two sets A and B. It is denoted by $A \cup B$. Logically speaking, if the biconditional statement $(x \in C) \Leftrightarrow (x \in A) \vee (x \in B)$ is true for all x, then $C = A \cup B$. In other words $(x \in A \cup B) \equiv (x \in A) \vee (x \in B)$.
14. The intersection of two sets A and B is the set of all those elements x such that x belongs to both A and B and is denoted by $A \cap B$. If $A \cap B = \phi$, then A and B are said to be disjoint.

    Logically speaking, if the biconditional statement $(x \in C) \Leftrightarrow (x \in A) \wedge (x \in B)$ is true for all x, then $C = A \cap B$. Hence,

$$(x \in A \cap B) \equiv (x \in A) \cap (x \in B)$$

15. If A and B are two sets then complement of B relative to A is the set of all those elements $x \in A$ such that $x \notin B$ and is denoted by $A - B$. Logically speaking, if for a set C the biconditional statement $(x \in C) \Leftrightarrow (x \in A) \cap (x \notin B)$ is true for all x, then $C = A - B$. In other words, if $(x \in C) \equiv (x \in A) \wedge (x \notin B)$ then C is called the complement of B relative to A.
16. The set of all subsets of a given set A is called the *power set* of A and is denoted by $P(A)$. The name power set is motivated by the fact that 'if A has n elements then its power set $P(A)$ contains exactly 2^n elements.'
17. Two events A and B are said to be independent events, if the occurrence of one event is not at all influenced by the occurrence of the other. For example, if two fair coins are tossed, then the result of one toss is totally independent of the result of the other toss. The probability that a head will be the outcome of any one toss will always be 1/2, irrespective of whatever the outcome is of the other toss. Hence, these two events are independent.
18. Binomial distribution is one of the simplest and most frequently used discrete probability distribution and is very useful in many practical situations involving *either/or* types of events.
19. A random variable is a phenomenon of interest in which the observed outcomes of an activity are entirely by chance, are absolutely unpredictable and may differ from response to response.

1.6 SUMMARY

NOTES

- A probability is expressed as a real number, $p \in [0, 1]$ and the probability number is expressed as a percentage (0 per cent to 100 per cent) and not as a decimal.
- The classical theory of probability is the theory based on the number of favourable outcomes and the number of total outcomes.
- If the number of outcomes belonging to an event E is N_E, and the total number of outcomes is N, then the probability of event E is defined as

  $$p_E = \frac{N_E}{N}$$
- We cannot calculate the probability where the outcomes are unequal probabilities.
- The sequence $\frac{n_A}{n}$ in the limit that will converge to the same result every time, or that it will not converge at all.
- The axiomatic probability theory is the most general approach to probability, and is used for more difficult problems in probability.
- The empirical approach to determining probabilities relies on data from actual experiments to determine approximate probabilities instead of the assumption of equal likeliness.
- The relationship between these empirical probabilities and the theoretical probabilities is suggested by the (Law of Large Numbers). The law states that as the number of trials of an experiment increases, the empirical probability approaches the theoretical probability.
- Multiplication rule is applied when it is necessary to compute the probability if both events A and B will occur at the same time.
- Bayes' theorem makes use of conditional probability formula where the condition can be described in terms of the additional information which would result in the revised probability of the outcome of an event.
- A sample space is the collection of all possible events or outcomes of an experiment.
- An event is an outcome or a set of outcomes of an activity or a result of a trial.
- Two mutually exclusive events are said to be complementary if they between themselves exhaust all possible outcomes.
- A probability space is a measure of space such that the measure of the whole space is equal to 1. A simple finite probability space is an ordered pair (S, p) such that S is set and p is a function with domain S.

NOTES

- The members of a set are called its elements. We use capital letters to denote sets and small letters to denote elements.
- A set having only one element is called singleton. If a is the element of the singleton A, then A is denoted by $A = \{a\}$.
- Whenever we talk of a set, we shall assume it to be a subset of a fixed set U. This fixed set U is called the universal set.
- A set which has no element is called the null set or empty set. It is denoted by the symbol *f*.
- The union of any two sets A and B is the set of all those elements x such that x belongs to at least one of the two sets A and B.
- The intersection of two sets A and B is the set of all those elements x such that x belongs to both A and B and is denoted by $A \cap B$. If $A \cap B = \phi$, then A and B are said to be disjoint.
- If A and B are two sets then complement of B relative to A is the set of all those elements $x \in A$ such that $x \notin B$ and is denoted by $A - B$.
- Two events A and B are said to be independent events, if the occurrence of one event is not influenced at all by the occurrence of the other.
- The fraction of successes depends on the probability of success in each trial, as the number of trials increases in repeated independent trials with the same probability of success.

1.7 KEY WORDS

- **Classical theory of probability:** It is the theory of probability based on the number of favourable outcomes and the number of total outcomes.
- **Event:** It is an outcome or a set of outcomes of an activity or the result of a trial.
- **Elementary event:** It is the single possible outcome of an experiment. It is also known as a simple event.
- **Joint event:** It is also known as compound event and has two or more elementary events in it.
- **Sample space:** It is the collection of all possible events or outcomes of an experiment.
- **Addition rule:** It states that when two events are mutually exclusive, then the probability that either of the events will occur is the sum of their separate probabilities.
- **Multiplication rule:** It is applied when it is necessary to compute the probability, if both events A and B occur at the same time. Different rules are applied for different conditions.

NOTES

- **Set:** It is a collection of objects, such that given an object, it is possible to determine whether that object belongs to the given collection or not.
- **Element:** The members of a set are called its elements.
- **Singleton:** It is a set having only one or single element.
- **Null set:** It is a set with no elements.

1.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Explain the concept of probability.
2. What are the different theories of probability? Explain briefly.
3. What do you understand by simple probability?
4. Explain the axiomatic approach to probability.
5. Explain the concept of multiplication rule.
6. What is Bayes' theorem? What is its importance in statistical calculations?
7. Define sample space.
8. Explain event and its types with the help of examples.
9. What is a mutually exclusive event?
10. Explain the terms, 'Sum of Events' and 'Product of Events'.
11. What is finite probability space?
12. Define set with the help of examples.
13. How will you define a universal set?
14. When are two sets termed equal?
15. Explain power set.
16. Describe union, intersection and complement set operations.
17. Explain the importance of Venn diagrams.
18. What do you understand by finite sets and counting principle?
19. Explain distributive laws and De Morgan's laws of set theory.
20. What are the important applications of set theory?
21. Explain duality.
22. What is mathematical induction?
23. Describe the importance of conditional probability.
24. When are independent repeated trials used?
25. Explain the properties of binomial distribution.

26. What is a random variable? Differentiate between qualitative and quantitative random variables.

Long-Answer Questions

NOTES

1. A family plans to have two children. What is the probability that both children will be boys? (List all the possibilities and then select the one which would be two boys.)
2. A family plans to have three children. List all the possible combinations and find the probability that all the three children will be boys.
3. A card is selected at random from an ordinary well-shuffled pack of 52 cards. What is the probability of getting:
   (*i*) A king
   (*ii*) A spade
   (*iii*) A king or an ace
   (*iv*) A picture card
4. A wheel of fortune has numbers 1 to 40 painted on it, each number being at equal distance from the other so that when the wheel is rotated, there is the same chance that the pointer will point at any of these numbers. Tickets have been issued to contestants numbering 1 to 40. The number at which the wheel stops after being rotated would be the winning number. What is the probability that:
   (*i*) Ticket number 29 wins.
   (*ii*) One person who bought 5 tickets numbered 18 to 22 inclusive wins the prize.
5. In a computer course, the probability that a student will get an A is 0.09. The probability that he will get a B grade is 0.15 and the probability that he will get a C grade is 0.45. What is the probability that the student will get either a D or an F grade?
6. The Dean of the School of Business has two secretaries, Mary and Jane. The probability that Mary will be absent on any given day is 0.08. The probability that Jane will be absent on any given day is 0.06. The probability that both the secretaries will be absent on any given day is 0.02. Find the probability that either one of them will be absent on any given day.
7. A fair die is rolled once. What is the probability of getting:
   (*i*) An odd number.
   (*ii*) A number greater than 3.
8. Two fair dice are rolled. What is the probability of getting:
   (*i*) A sum of 10 or more.
   (*ii*) A pair of which atleast one number is 3.
   (*iii*) A sum of 8, 9, or 10.
   (*iv*) One number less than 4.

NOTES

9. An urn contains 12 white balls and 8 red balls. Two balls are to be selected in succession, at random and without replacement. What is the probability that:
   (*i*) Both balls are white.
   (*ii*) The first ball is white and the second ball is red.
   (*iii*) One white ball and one red ball are selected.
   (*iv*) Would the probabilities change if the first ball after being identified is put back in the urn before the second ball is selected?
10. In a statistics class, the probability that a student picked up at random comes from a two parent family is 0.65, and the probability that he will fail the exam is 0.20. What is the probability that such a randomly selected student will be a low achiever given that he comes from a two parent family?
11. The following is a breakdown of faculty members in various ranks at the college.

Rank	*Number of Males*	*Number of Females*
Professor	20	12
Assoc. Professor	18	20
Asst. Professor	25	30

   What is the probability that a faculty member selected at random is:
   (*i*) A female.
   (*ii*) A female professor.
   (*iii*) A female given that the person is a professor.
   (*iv*) A female or a professor.
   (*v*) A professor or an assistant professor.
   (*vi*) Are the events of being a male and being an associate professor statistically independent events?
12. A movie house is filled with 700 people and 60 per cent of these people are females. 70 per cent of these people are seated in the no smoking area including 300 females. What is the probability that a person picked up at random in the movie house is:
   (*i*) A male.
   (*ii*) A female smoker.
   (*iii*) A male or a non-smoker.
   (*iv*) A smoker if we knew that the person is a male.
   (*v*) Are the events sex and smoking statistically independent?
13. A part-time student is taking two courses, namely, Statistics and Finance. The probability that the student will pass the Statistics course is 0.60 and the probability of passing the Finance course is 0.70. The probability that the student will pass both courses is 0.50. Find the probability that the student:

NOTES

(*i*) will pass at least one course.

(*ii*) will pass either or both courses.

(*iii*) will fail both courses.

14. 200 students from the college were surveyed to find out if they were taking any of the Management, Marketing or Finance courses. It was found that 80 of them were taking Management courses, 70 of them were taking Marketing courses and 50 of them were taking Finance courses. It was also found that 30 of them were taking Management and Marketing courses, 30 of them were taking Management and Finance courses and 25 of them were taking Marketing and Finance courses. It was further determined that 20 of these students were taking courses in all the three areas. What is the probability that a particular student is not taking any course in any of these areas?
15. Out of 20 students in a Statistics class, 3 students are failing in the course. If 4 students from the class are picked up at random, what is the probability that one of the failing students will be among them.
16. The New York Pick Five lottery drawing draws five numbers at random out of 39 numbers labelled 1 to 39. How many different outcomes are possible?
17. A company has 18 senior executives. Six of these executives are women including four blacks and two Indians. Six of these executives are to be selected at random for a Christmas cruise. What is the probability that the selection will include:

    (*i*) All the black and Indian women.

    (*ii*) At least one Indian woman.

    (*iii*) Not more than two women.

    (*iv*) Half men and half women.
18. The probability that a management trainee will remain with the company after the training programme is completed is 0.70. The records indicate that 60 per cent of all managers earn over $60,000 per year. The probability that an employee is a management trainee or who earns more than $60,000 per year is 0.80. What is the probability that an employee earns more than $60,000 per year, given that he is a management trainee who stayed with the company after completing the training programme.
19. A manufacturer of laptop computer monitors has determined that on an average, 3 per cent of screens produced are defective. A sample of one dozen monitors from a production lot was taken at random. What is the probability that in this sample fewer than 2 defectives will be found?
20. A fair coin is tossed 16 times. What is the probability of getting no more than 2 heads?
21. A student is given 4 True or False questions. The student does not know the answer to any of the questions. He tosses a fair coin. Each time he gets a head, he selects True. What is the probability that he will get:

NOTES

(*i*) Only one correct answer

(*ii*) At most 2 correct answers

(*iii*) At least 3 correct answers

(*iv*) All correct answers

22. A newly married couple plans to have 5 children. An astrologist tells them that based on astrological reading, they have an 80 per cent chance of having a baby boy on any particular birth. The couple would like to have 3 boys and 2 girls. Find the probability of this event.
23. An automatic machine makes paper clips from coils of wire. On an average, one in 400 paper clips is defective. If the paper clips are packed in small boxes of 100 clips each, what is the probability that any given box of clips contains.

    (*i*) No defectives

    (*ii*) One or more defectives

    (*iii*) Less than two defectives

    (*iv*) Two or less defectives
24. Because of recycling campaigns, a number of empty glass soda bottles are being returned for refilling It has been found that 10 per cent of the incoming bottles are chipped and hence are discarded. In the next batch of 20 bottles, what is the probability that:

    (*i*) None will be chipped.

    (*ii*) Two or fewer will be chipped.

    (*iii*) Three or more will be chipped.

    (*iv*) What is the expected number of chipped bottles in a batch of 20?
25. The customers arrive at a drive-in window of Apple bank at an average rate of one customer per minute.

    (*i*) What is the probability that exactly two customers arrive in a given minute?

    (*ii*) What is the probability of no arrivals in a particular minute?
26. Write the following sets by listing elements enclosed in brackets { }:

    (*i*) A is the set whose elements are first five letters of the alphabet.

    (*ii*) B is the set of all odd integers.

    (*iii*) X is the set of all two digit positive numbers which are divisible by 15.
27. Write the following sets using a statement to designate each:

    (*i*) $A = \{3, 6, 12, 15, 18\}$

    (*ii*) $B = \{s, t, u, v, w, x, y, z\}$

    (*iii*) $C = \{1, 3, 5, 7, \ldots, 2n - 1, \ldots\}$

    (*iv*) $D = \left\{\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots \frac{1}{n}, \ldots\right\}$

28. Indicate which of the following sets are finite:
    (*i*) $A = \{x \mid x \text{ is a positive integer}\}$
    (*ii*) $B = \{x \mid x \text{ is an even integer lying between 2 and 10}\}$
    (*iii*) $C = \{x \mid x \text{ is a letter of the alphabet}\}$
    (*iv*) $E = \{x \mid x \text{ is an integter less than 10}\}$
29. Let A be the set $\{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\}$. Now, list the following sets:
    (*i*) $\{x \mid x \text{ is an element of } A \text{ and } x + 1 \text{ is even}\}$
    (*ii*) $\{x \mid x \text{ is an element of } A \text{ and } 2x \text{ is an element of } A\}$
    (*iii*) $\{x \mid x \text{ is an element of } A \text{ and } 2x < 20\}$
    (*iv*) $\{x \mid x \text{ is not an element of } A \text{ and } 0 < x < 21\}$
30. Find all possible solutions for x and y in each of the following cases:
    (*i*) $\{2, 3\} = \{2x, y\}$
    (*ii*) $\{x, y\} = \{1, 2\}$
    (*iii*) $\{x, x^2\} = \{9, 3\}$
31. Show that if $a_1 \neq b_1$, $a_1 \neq a_2$ and $\{a_1, b_1\} = \{a_2, b_2\}$, then $b_1 = b_2$.
32. Show that $\{a\} = \{b, c\}$ if and only if $a = b = c$.
33. State the relation if any, between sets A and B in the following:
    (*i*) $A = \{1, 3, 5, 7, 9, \ldots\}$
    $B = \{3, 9, 15, 21, \ldots, 3(2n - 1), \ldots\}$
    (*ii*) $A = \{2, 4, 7, 12, 18, 24\}$
    $B = \{1, 3, 7, 11, 16, 22, 29\}$
    (*iii*) $A = \{x \mid x \text{ is an even natural number less than 20}\}$
    $B = \{x \mid x \text{ is natural number less than 20 which is divisible by 2}\}$
    (*iv*) $A = \{x \mid x \text{ is an even integer}\}$
    $B = \{x \mid x \text{ is an integer divisible by 3}\}$
34. Prove that $A \subseteq B$ and $B \subset C$ implies $A \subset C$.
35. Prove that $A \subseteq \phi$ implies $A = \phi$.
36. If a set A has 101 elements, find the number of subsets of A having odd number of elements.
37. What are the elements of the power set of the set $[1, \{2, 3\}]$?
38. If $A = \{1, 2, 3, 4, 5\}$, $B = \{2, 4, 6, 8, 10\}$, $C = \{3, 6, 9, 12, 15\}$. Find that,
    (*i*) $(A \cup B) \cap C$ (*ii*) $A \cup (B \cap C)$
    (*iii*) $(A \cap C) \cup B$
39. If $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$, $C = \{4, 5, 6, 7\}$. Find that,
    (*i*) $A - B$ (*ii*) $(A \cup B) - C$
    (*iii*) $A - (B \cap \text{C})$ (*iv*) $(A \cap B) - (B\ U\ C)$

40. If $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
    $A = \{1, 4, 7, 10\}$
    $B = \{2, 5, 8\}$
    Find (*i*) A' (*ii*) B' (*iii*) $A \cap B'$ (*iv*) $A' \cap B$ (*v*) $A' \cap B'$
41. Prove that $(A \cap B) \cup C = A \cap (B \cup C)$ if and only if $C \subseteq A$.
42. Prove that if $A \subseteq B$ then $P(A) \subseteq P(B)$.
43. For any sets A and B, prove or disprove that,
    $P(A) \cap P(B) = P(A \cap B)$
    $P(A) \cup P(B) = P(A \cup B)$
44. In a survey of 100 students, the numbers studying various languages were found to be: Spanish 28; German 30; French 42; Spanish and French 10; Spanish and German 8; German and French 5; all the three languages 3.
    (*i*) How many students were studying no language?
    (*ii*) How many students had French as their only language?
    (*iii*) How many students studied German if and only if they studied French?
45. In each of the following sentences, determine which is a statement (*S*), or not (*N*):
    (*i*) Every rectangle is a square.
    (*ii*) The sum of three angles of a triangle is 180°.
    (*iii*) How are you?
    (*iv*) $2 + 1 = 3$.
    (*v*) $\sqrt{2}$ is a rational number.
46. In a latter survey of the 100 students, the numbers studying the various languages were found to be:
    German only 18; German but not Spanish 23; German and French 8; German 26; French 48; French and Spanish 8; studying no language 24.
    (*i*) How many students took Spanish?
    (*ii*) How many took German and Spanish but not French?
    (*iii*) How many took French if and only if they did not take Spanish?
47. If A and B are two sets, prove that number of elements in $A \cap B'$ is equal to: (Number of elements in A – Number of elements in $A \cap B$).
48. The report of one survey of 100 students stated that the numbers studying the various languages were; all three languages 5; German and Spanish 10; French and Spanish 8; German and French 20; Spanish 30; German 23; French 50. The surveyor who turned in this report was fired. Why?
49. In a recent survey of 5000 people, it was found that 2800 read Indian Express and 2300 read Statesman while 400 read both the papers. How many read neither Indian Express nor Statesman?

NOTES

50. In a survey of 30 students, it was found that 19 take Mathematics, 17 take Music, 11 take History, 7 take Mathematics and History, 12 take Mathematics and Music, 5 take Music and History and 2 take all three courses. Find (*i*) The number of students that take Mathematics but do not take History (*ii*) The number that take exactly two of the three courses.
51. In a Chemistry class there are 20 students, and in a Psychology class there are 30 students. Find the number either in a Psychology class or Chemistry class if,
    (*i*) The two classes meet at the same hour.
    (*ii*) The two classes meet at different hours and 10 students are enrolled in both courses.
52. On an Air India flight, there are 9 boys, 5 Indian children, 9 men, 7 foreign boys, 14 Indian, 6 Indian males and 7 foreign females. What is the number of people in this plane?
53. A college awarded 38 medals in Football, 15 in Basket ball and 20 in Cricket. If these medals went to a total of 58 men and only three of these men got medals in all the three sports, how many men received medals in exactly two of the three sports?
54. Suppose that in survey concerning the reading habits of students it is found that:
    60 per cent read magazine *A*,
    50 per cent read magazine *B*,
    50 per cent read magazine *C*,
    30 per cent read magazines *A* and *B*,
    20 per cent read magazines *B* and *C*,
    30 per cent read magazines *A* and *C*,
    10 per cent read all three magazines.
    (*i*) What per cent read exactly two magazines?
    (*ii*) What per cent do not read any of the magazines?
55. In a survey of 500 consumers, it was found that 425 liked product *A* and 375 liked product *B*. What is the least number of consumers that must have liked both products assuming that there may be consumers of products different from *A* and *B*.
56. Explain how a class is different from a set.
57. How many power sets are formed from vowels of English language?
58. Give a brief description of basic sets and denotable sets.
59. Explain how a fuzzy set is different from a generally defined set.
60. Give your comments on ‘operations on fuzzy sets’. Explain three operations.
61. What are the areas of application of fuzzy set? Explain with three examples.

1.9 FURTHER READINGS

NOTES

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 2 RANDOM VARIABLES OF DISCRETE AND CONTINUOUS TYPE

Structure

2.0 INTRODUCTION

Randomness means each possible entity has the same chance of being considered. A random variable may be qualitative or quantitative in nature. You will study probability distribution, which means listing of all possible outcomes of an experiment together with their probabilities. It may be discrete or continuous.

In this unit, you will study about the random variables of the discrete type and random variables of the continuous type.

2.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the random variables of the discrete type
- Analyse the random variables of the continuous type

2.2 RANDOM VARIABLES OF THE DISCRETE TYPE AND RANDOM VARIABLES OF THE CONTINUOUS TYPE

Discrete Probability Distributions

When a random variable x takes discrete values $x_1, x_2,...., x_n$ with probabilities $p_1, p_2,...,p_n$, we have a discrete probability distribution of X.

The function $p(x)$ for which $X = x_1, x_2, ..., x_n$ takes values $p_1, p_2, ..., p_n$, is the probability function of X.

The variable is discrete because it does not assume all values. Its properties are:

$$p(x_i) = \text{Probability that } X \text{ assumes the value } x.$$
$$= \text{Prob } (x = x_i) = p_i$$
$$p(x) \geq 0, \; \Sigma p(x) = 1$$

For example, four coins are tossed and the number of heads X noted. X can take value 0, 1, 2, 3, 4 heads.

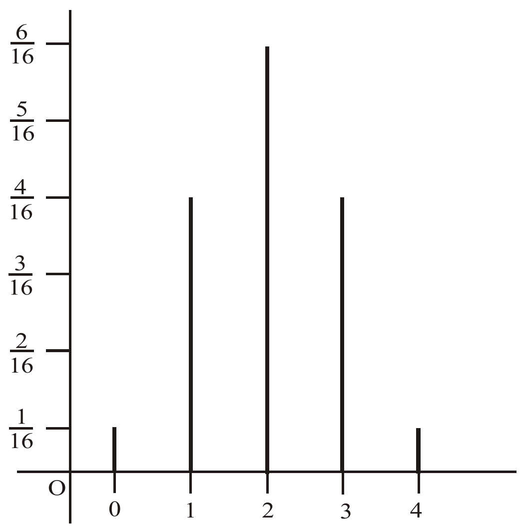
$$p(X = 0) = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

$$p(X = 1) = {}^4C_1\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 = \frac{4}{16}$$

$$p(X = 2) = {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{6}{16}$$

$$p(X = 3) = {}^4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) = \frac{4}{16}$$

$$p(X = 4) = {}^4C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = \frac{1}{16}$$



$$\sum_{x=2}^{4} p(x) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16} = 1$$

This is a discrete probability distribution.

Example 1: If a discrete variable X has the following probability function, then, find (*i*) a (*ii*) $p(X \leq 3)$ (*iii*) $p(X \geq 3)$

Solution:

x_1	$p(x_i)$
0	0
1	a
2	$2a$
3	$2a^2$
4	$4a^2$
5	$2a$

Since $\Sigma p(x) = 1$, $0 + a + 2a + 2a^2 + 4a^2 + 2a = 1$

$$\therefore \qquad 6a^2 + 5a - 1 = 0, \text{ so that } (6a - 1)(a + 1) = 0$$

$$a = \frac{1}{6} \text{ or } a = -1 \text{ (not admissible)}$$

For $a = \frac{1}{6}$, $p(X \leq 3) = 0 + a + 2a + 2a^2 = 2a^2 + 3a = \frac{5}{9}$

$$p(X \geq 3) = 4a^2 + 2a = \frac{4}{9}$$

Discrete Distributions

There are several discrete distributions. Some other discrete distributions are described below.

Uniform or Rectangular Distribution

Each possible value of the random variable x has the same probability in the uniform distribution. If x takes vaues $x_1, x_2 ...., x_k$, then

$$p(x_i, k) = \frac{1}{k}$$

The numbers on a die follow the uniform distribution,

$$p(x_i, 6) = \frac{1}{6} \text{ (Here } x = 1, 2, 3, 4, 5, 6)$$

Bernoulli Trials

In a Bernoulli experiment, an even E either happens or does not happen (E'). Examples are, getting a head on tossing a coin, getting a six on rolling a die and so on.

The Bernoulli random variable is written,

$$X = 1 \text{ if } E \text{ occurs}$$
$$= 0 \text{ if } E' \text{ occurs}$$

Since there are two possible value it is a case of a discrete variable where,

Probability of success = $p = p(E)$

Profitability of failure = $1 - p = q = p(E')$

We can write,

For $k = 1, f(k) = p$

For $k = 0, f(k) = q$

For $k = 0$ or $1, f(k) = p^k q^{1-k}$

Negative Binomial

In this distribution the variance is larger than the mean.

Suppose, the probability of success p in a series of independent Bernoulli trials remains constant.

Suppose the rth success occurs after x failures in $x + r$ trials.

1. The probability of the success of the last trial is p.
2. The number of remaining trials is $x + r - 1$ in which there should be $r - 1$ successes. The probability of $r - 1$ successes is given by,

$${}^{x+r-1}C_{r-1} p^{r-1} q^x$$

The combined pobability of cases (1) and (2) happening together is,

$$p(x) = px^{x+r-1}C_{r-1}p^{r-1}q^x \qquad x = 0, 1, 2,....$$

This is the Negative Binomial distribution. We can write it in an alternative form,

$$p(x) = {}^{-r}C_x p^r (-q)^x \quad x = 0, 1, 2,....$$

This can be summed up as follows:

In an infinite series of Bernoulli trials the probability that $x + r$ trials will be required to get r successes is the Negative Binomial,

$$p(x) = {}^{x+r-1}C_{r-1} p^{r-1} q^x \qquad r \geq 0$$

If $r = 1$, it becomes the Geometric distribution.

If $p \to 0, \to \infty, rp = m$ a constant, then the negative binomial tends to the Poisson distribution.

Geometric Distribution

Suppose the probability of success p in a series of independent trials remains constant.

Suppose, the first success occurs after x failures, i.e., there are x failures preceding the first success. The probability of this event will be given by $p(x) = q^x p$ $(x = 0, 1, 2,.....)$

This is the Geometric distribution and can be derived from the Negative Binomial. If we put $r = 1$ in the Negative Binomial distribution:

$$p(x) = {}^{x+r-1}C_{r-1}p^{r-1}\,q^{x}$$

We get the Geometric distribution,

$$p(x) = {}^{x}C_{0}\,p^{1}\,q^{x} = pq^{x}$$

$$\Sigma p(x) = \sum_{n=0}^{p} q^{x}\,p = \frac{p}{1-q} = 1$$

$$E(x) = \text{Mean} = \frac{p}{q}$$

$$\text{Variance} = \frac{p}{q^{2}}$$

$$\text{Mode} = \left(\frac{1}{2}\right)^{x}$$

Example 2: Find the expectation of the number of failures preceding the first success in an infinite series of independent trials with constant probability p of success.

Solution: The probability of success in,

1st trial $= p$ (Success at once)

2nd trial $= qp$ (One failure then success and so on)

3rd trial $= q^2p$ (Two failures then success and so on)

The expected number of failures preceding the success,

$$E(x) = 0\,.\,p + 1.\,pq + 2p^{2}p + \ldots\ldots\ldots\ldots$$

$$= pq(1 + 2q + 3q^{2} + \ldots\ldots\ldots)$$

$$= pq\frac{1}{(1-q)^{2}} = qp\frac{1}{p^{2}} = \frac{q}{p}$$

Since $p = 1 - q$.

Hypergeometic Distribution

From a finite population of size N, a sample of size n is drawn without replacement.

Let there be N_1 successes out of N.

The number of failures is $N_2 = N - N_1$

The disribution of the random variable X, which is the number of successes obtained in the above case, is called the Hypergeometic distribution.

NOTES

$$p(x) = \frac{{}^{N_1}C_x \, {}^{N}C_{n-x}}{{}^{N}C_n} \quad (X = 0, 1, 2, ..., n)$$

Here x is the number of successes in the sample and $n-x$ is the number of failures in the sample.

It can be shown that,

$$\text{Mean} : E(X) = n\frac{N_1}{N}$$

$$\text{Variance} : \text{Var (X)} = \frac{N-n}{N-1}\left(\frac{nN_1}{N} - \frac{nN_1^2}{N}\right)$$

Example 3: There are 20 lottery tickets with three prizes. Find the probability that out of 5 tickets purchased exactly two prizes are won.

Solution: We have $N_1 = 3$, $N_2 = N - N_1 = 17$, $x = 2$, $n = 5$.

$$p(2) = \frac{{}^{3}C_2 \, {}^{17}C_3}{{}^{20}C_5}$$

The probability fo no pize $p(0) = \dfrac{{}^{3}C_0 \, {}^{17}C_5}{{}^{20}C_5}$

The probability of exactly 1 prize $p(1) = \dfrac{{}^{3}C_1 \, {}^{17}C_4}{{}^{20}C_5}$

Example 4: Examine the nature of the distibution of r balls are drawn, one at a time without replacement, from a bag containing m white and n black balls.

Solution: It is the hypergeometric distribution. It corresponds to the probability that x balls will be white out of r balls so drawn and is given by,

$$p(x) = \frac{{}^{x}C_x \, {}^{n}C_{r-x}}{{}^{m+n}C_r}$$

Multinomial

There are k possible outcomes of trials, viz., $x_1, x_2, ..., x_k$ with probabilities $p_1, p_2, ..., p_k$, n independent trials are performed. The multinomial distibution gives the probability that out of these n trials, x_1, occurs n_1 times, x_2 occurs n_2 times and so on. This is given by $\dfrac{n!}{n_1!n_2!....n_k!} p_1^{n_1} p_2^{n_2} ....p_k^{n}$

Where, $\sum_{i=1}^{k} n_i = n$

Characteristic Features of the Binomial Distribution

The following are the characteristics of Binomial distribution:

1. It is a discrete distribution.
2. It gives the probability of x successes and $n-x$ failures in a specific order.
3. The experiment consists of n repeated trials.
4. Each trial results in a success or a failure.
5. The probability of success remains constant from trial to trial.
6. The trials are independent.
7. The success probability p of any outcome remains constant over time. This condition is usually not fully satisfied in situations involving management and economics, for example, the probability of response from successive informants is not the same. However, it may be assumed that the condition is reasonably well satisfied in many cases and that the outcome of one trial does not depend on the outcome of another. This condition too, may not be fully satisfied in many cases. An investigator may not approach a second informant with the same set-up of mind as used for the first informant.
8. The binomial distribution depends on two parameters n and p. Each set of different values of n, p has a different binomial distribution.
9. If $p = 0.5$, the distribution is *symmetrical.* For a symmetrical distribution, in n

   $$\text{Prob. } (X = 0) = \text{Prob } (X = n)$$

   i.e., the probabilities of 0 or n successes in n trials will be the same. Similarly,

   $$\text{Prob } (X = 1) = \text{Prob}(X = n - 1) \text{ and so on.}$$

   If $p > 0.5$, the distribution is not symmetrical. The probabilities on the right are larger than those on the left. The reverse case is when $p < 0.5$.

   When n becomes large the distribution becomes bell shaped. Even when n is not very large but $p \cong 0.5$, it is fairly bell shaped.
10. The binomial distribution can be approximated by the normal. As n becomes large and p is close to 0.5, the approximation becomes better.

Example 5: If the ratio *n/N,* i.e., sample size to population size is small, the result given by the Binomial may not be reliable. Comment.

Solution: When the distribution is binomial, each successive trial, being independent of other trials, has constant probability of success. If he sampling of n items is without replacement from a population of size N, the probability of success of any event depends upon what happened in the previous events. In this case the Bionomial cannot be used unless the ratio n/N is small. Even then there is no guarantee of getting accurate results.

The Binomial should be used only if the ratio $\dfrac{n}{N}$ is very small, say less that 0.05.

Example 6: Explain the concept of a discrete probability distribution.

Solution: If a random variable x assumes n discrete values $x_1, x_2, \ldots\ldots x_n$, with respective probabilities $p_1, p_2, \ldots\ldots\ldots p_n (p_1 + p_2 + \ldots\ldots + p_n = 1)$ then, the distribution of values x_i with probabilities p_i (= 1, 2,.....n), is called the discrete probability distribution of x.

The frequency function or frequency distribution of x is defined by $p(x)$ which for different values $x_1, x_2, \ldots\ldots x_n$ of x, gives the corresponding probabilities:

$$p(x_i) = p_i \text{ where, } p(x) \geq 0 \ \Sigma p(x) = 1$$

Example 7: For the following probability distribution, find $p(x > 4)$ and $p(x \geq 4)$:

x	0	1	2	3	4	5
$p(x)$	0	a	$a/2$	$a/2$	$a/4$	$a/4$

Solution:

Since, $$\Sigma p(x) = 1, 0 + a + \frac{a}{2} + \frac{a}{2} + \frac{a}{4} + \frac{a}{4} = 1$$

$$\therefore \quad \frac{5}{2}a = 1 \quad \text{or} \quad a = \frac{2}{5}$$

$$p(x > 4) = p(x = 5) = \frac{9}{4} = \frac{1}{10}$$

$$p(x \leq 4) = 0 + a + \frac{a}{2} + \frac{a}{2} + \frac{a}{4} + \frac{9a}{4} = \frac{9}{10}$$

Example 8: A fair coin is tossed 400 times. Find the mean number of heads and the corresponding standard deviation.

Solution: This is a case of Binomial distribution with $p = q = \frac{1}{2}, n = 400$

The mean number of heads is given by $\mu = np = 400 \times \frac{1}{2} = 200$

and S. D. $\sigma = \sqrt{npq} = \sqrt{400 \times \frac{1}{2} \times \frac{1}{2}} = 10$

Example 9: A manager has thought of 4 planning strategies each of which has an equal chance of being successful. What is the probability that at least one of his strategies will work if he tries them in 4 situations? Here $p = \frac{1}{4}, q = \frac{3}{4}$.

Solution: The probability that none of the strategies will work is given by,

$$p(0) = {}^4C_0 \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^4 = \left(\frac{3}{4}\right)^4$$

The probability that at least one will work is given by $1 - \left(\frac{3}{4}\right)^4 = \frac{175}{256}$

Example 10: Suppose the proportion of people preferring a car C is 0.5. Let X denote the number of people out of a set of 3 who prefer C. The probabilities of 0, 1, 2, 3 of them preferring C are,

Solution:

$$p(X=0) = {}^3C_0(0.5)^0\,(0.5)^3 = \frac{1}{8}$$

$$p(X=1) = {}^3C_1(0.5)^1\,(0.5)^2 = \frac{3}{8}$$

$$p(X=2) = {}^3C_2(0.5)^2\,(0.5)^1 = \frac{3}{8}$$

$$p(X=3) = {}^3C_3(0.5)^3\,(0.5)^0 = \frac{1}{8}$$

$$\mu = E(X) = \Sigma x_i p_i = 0\times\frac{1}{8}+1\times\frac{3}{8}+2^2\frac{3}{8}+2\times\frac{3}{8}+3\times\frac{1}{8}=1.5$$

$$\sigma^2 = E(X-\mu)^2 = E(X^2)-\mu^2 = \Sigma x_i^2 p_i - \mu^2$$

$$= 0^2\times\frac{1}{8}+1^2\times\frac{3}{8}+2^2\times\frac{3}{8}+3^2\times\frac{1}{8}-1.5^2$$

$$= 0.75$$

Example 11: For the Poisson distribution, write the probabilities of 0, 1, 2, ... successes.

Solution:

x	$p(x) = e^{-m}\dfrac{m^x}{x!}$
0	$p(0) = e^{-m}\, m^0/0!$
1	$p(1) = e^{-m}\dfrac{m}{1!} = p(0).m$
2	$e^{-m}\dfrac{m^2}{2!} = p(2) = p(1).\dfrac{m}{2}$
3	$e^{-m}\dfrac{m^3}{3!} = p(3) = p(2).\dfrac{m}{3}$
$\vdots$	

and so on.

Total of all probabilities $\Sigma p(x) = 1$

Example 12: What are the raw moments of Poisson distribution?

Solution: First raw moment $\mu'_1 = m$

Second raw moment $\mu'_2 = m^2 + m$

Third raw moment $\mu'_3 = m^3 + 3m^2 + m$

NOTES

Example 13: For a Poisson distribution $p(0) = p(1)$, find $p(x > 0)$.

Solution: We have $e^{-m}\, m^0/0! = e^{-m}\, m^0/1!$ so that $m = 1$

$$\therefore\ p(0) = e^{-1} = 1/2.718 = 0.37$$

$$p(x > 0) = 1 - p(0) = 0.63$$

Example 14: It is claimed that the TV branded M is demanded by 70% of customers. If X is the number of TVs demanded, find the probability distribution of X when there are four customers.

Solution: If all the four demand M then $p(4) = 0.7^4 = 0.2401$. The probability distribution is

X	4	3	2	1	0
p	0.2401	0.4116	0.2646	0.0756	0.0081

These values may be plotted.

Note: Poisson Approximation to the Binomial. When the number of trials n is large and the probability p of a success is small, the binomial distribution can be approximated by the Poisson. This approximation is useful is practice.

If $p = 0.002$, $n = 1000$, $m = np = 2$

The probability of 2 successes in 1000 trials is,

$$p(X = 2) = {}^{n}C_x p^x q^{n-x} = {}^{100}C_2 (0.002)^2\ (0.998)^{998}$$

Similarly, $$p(X = 3) = {}^{100}C_3 (0.002)^3\ (0.998)^{997}\text{, etc.}$$

These terms are difficult to calculate. If we employ the Poisson, we have a much easier task before us.

$$m = np = 100 \times 0.002 = 2$$

$$p(X = 2) = \frac{e^{-m} m^x}{x!} = \frac{e^{-2} 2^2}{2!} = 0.1353 \times 20.2706$$

$$p(X = 3) = \frac{e^{-m} m^x}{x!} = \frac{e^{-2} 2^3}{3!} = 0.1804$$

Example 15: One in every 100 items made on a machine is defective. Out of 25 items picked, find the probability of 1 item being defective.

$$p = 0.01,\ q = 0.99,\ n = 25,\ np = 0.25$$

$$\text{Binomial}: p(1) = {}^{25}C_1 (0.1)^1 (0.99)^{24} = 0.1964$$

$$\text{Poisson}: p(1) = \frac{e^{-25} (0.25)^1}{1!} = 0.1947$$

Continuous Probability Distributions

When a random variate can take any value in the given interval $a \le x \le b$, it is a continuous variate and its distribution is a *Continuous Probability Distribution.*

Theoretical distributions are often continuous. They are useful in practice because they are convenient to handle mathematically. They can serve as good approximations to discrete distributions.

The range of the variate may be finite or infinite.

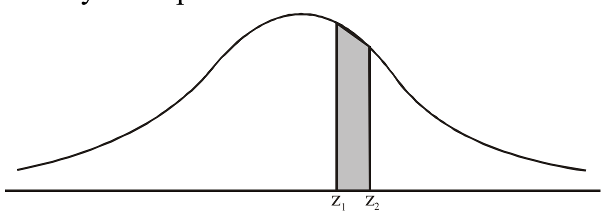
A continuous random variable can take all values in a given interval.

A continuous probability distribution is represented by a smooth curve.

The total area under the curve for a probability distribution is necessarily unity. The curve is always above the x axis because the area under the curve for any interval represents probability and probabilities cannot be negative.

If X is a continous variable, the probability of X falling in an interval with end points z_1, z_2 may be written $p(z_1 \leq X \leq z_2)$.

This probability corresponds to the shaded area under the curve.



A function is a probability density function if,

$\int_{-\infty}^{\infty} p(x)dx = 1, p(x) \geq 0, -\infty < x < \infty,$ i.e., the area under the curve $p(x)$ is '1' and the probability of x lying between two values a, b, i.e., $p(a < x < b)$ is positive. The most prominent example of a continuous probability function is the normal distribution.

Cumulative Probability Function (CPF)

The Cumulative probability function (CPF) shows the probability that x takes a value less than or equal to, say, z and corresponds to the area under the curve up to z:

$$p(x \leq z) = \int_{-\infty}^{z} p(x)dx$$

This is denoted by $F(x)$.

Check Your Progress

1. Explain the moment generating function with reference to probability theory.
2. Explain discrete probability distribution.
3. What is a hypergeometric distribution?
4. Explain continuous probability distribution.

NOTES

2.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. According to probability theory, moment generating function generates the moments for the probability distribution of a random variable *X,* and can be defined as:

   $M_X(t) = E\ (e^{tX}),\qquad t \in \overset{+}{\mathrm{R}}$

2. When a random variable x takes discrete values $x_1, x_2,...., x_n$ with probabilities $p_1, p_2,...,p_n$ we have a discrete probability distribution of X.
3. The disribution of the random variable X which is the number of successes obtained is called the hypergeometic distribution.
4. When a random variate can take any value in the given interval $a \le x \le b$, it is a continuous variate and its distribution is a Continuous Probability Distribution.

2.4 SUMMARY

- When a random variable x takes discrete values x_1, x_2,...., x_n with probabilities $p_1, p_2,...,p_n$, we have a discrete probability distribution of X.
- Each possible value of the random variable x has the same probability in the uniform distribution. If x takes vaues $x_1, x_2....,x_k$, then

  $$p(x_i,\ k) \ = \frac{1}{k}$$

- In a Bernoulli experiment, an even E either happens or does not happen (E').
- Suppose, the probability of success p in a series of independent Bernoulli trials remains constant.
- Suppose, the first success occurs after x failures, i.e., there are x failures preceding the first success. The probability of this event will be given by $p(x) = q^x p$ $(x = 0, 1, 2,.....)$
- The multinomial distibution gives the probability that out of these n trials, x_1, occurs n_1 times, x_2 occurs n_2 times and so on. This is given by

  $$\frac{n!}{n_1!n_2!....n_k!}\, p_1^{n_1}\, p_2^{n_2} ....p_k^{n}$$

- When a random variate can take any value in the given interval $a \le x \le b$, it is a continuous variate and its distribution is a Continuous Probability Distribution.

NOTES

- The Cumulative probability function (CPF) shows the probability that x takes a value less than or equal to, say, z and corresponds to the area under the curve up to z:

$$p(x \le z) = \int_{-\infty}^{z} p(x)dx$$

2.5 KEY WORDS

- **Hypergeometic distribution:** The disribution of the random variable X, which is the number of successes obtained in the above case, is called the Hypergeometic distribution.
- **Random variablele:** It is a variable which takes on different values as a result of the outcomes of a random experiment. It can be either discrete or continuous.

2.6 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Why is random variable considered important in statistics?
2. Explain the techniques of assigning probability.
3. What is moment generating function?
4. Explain briefly the probability distribution and its types.

Long-Answer Questions

1. Differentiate between a discrete and a continuous variable.
2. A continuous variable is an uninterrupted motion like the fall of a rain drop. Comment on it.
3. There are 3 tickets numbered 0, 2, 3. One ticket is selected and replaced; another ticket is selected and replaced. A third ticket is selected and replaced once again. If X stands for the sum of the 3 ticket numbers, construct the probability distribution of X and find m and σ^2.
4. A car insurance policy offers Rs. 1000 after an accident the probability of whose occurrence is 0.04. If the expected gain is to be zero what should be the premium?
5. For the probability distribution,

x	– 1	– 2	3	7.5	8
p	0.2	0.15	0.3	0.1	0.05

$E(x) = 2.15$ $\quad$ $V(x) = 9.5$

6. In a queue the probability of the number of people joining per minute given below.

Number of persons Joining the queue	0	1	2	3
Probability	0.4	0.3	0.2	0.1

7. From an urn containing 20 black and 30 red balls, 6 balls are drawn at random. Find the probability that no red ball is selected.
8. Examine the nature of the distribution of r balls are drawn, one at a time, without replacement, from a bag containing m white and n black balls.
9. 20 per cent of bolts in a factory are defective. Deduce the probability distribution of the number of defectives in a sample of 5.

2.7 FURTHER READINGS

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 3 EXPECTATION OF RANDOM VARIABLES

NOTES

Structure

3.0 INTRODUCTION

An expected value is the sum of each possible outcome and the probability of occurrence of outcome. Expectation may be conditional or iterated. You will study the moment generating function (MGF), which generates the moments for the probability distribution of a random variable. The subject of probability in itself is a cumbersome one, hence, only the basic concepts will be discussed here.

Since the outcomes of most decisions cannot be accurately predicted because of the impact of many uncontrollable and unpredictable variables, it is necessary that all the known risks be scientifically evaluated. Probability theory, sometimes referred to as the science of uncertainty, is very helpful in such evaluations. It helps the decision-maker with only limited information to analyse the risks and select the strategy of minimum risk.

In this unit, you will study about the properties of the distribution function, expectation of random variable, Chebyshev's inequality.

3.1 OBJECTIVES

After going through this unit, you will be able to:

- Describe the properties of the distribution function
- Analyse the expectation of random variable
- Understand about the Chebyshev's inequality

NOTES

3.2 PROPERTIES OF THE DISTRIBUTION FUNCTION

Distribution function can be related to any random variable which refers to the function that assigns a probability to each number in an organized and well-arranged method such that the value of the random variable is equal to or less than the given number.

The distribution function is also known as cumulative frequency distribution or cumulative distribution function. Fundamentally, it defines the probability that is the value related to the variable X tend to be equal to or less than number 'x'.

Probability distributions indicate the likelihood of an event or outcome, i.e., a probability distribution is a function that describes the likelihood of obtaining the possible values that a random variable can assume. In other words, the values of the variable vary based on the underlying probability distribution. This type of distribution is very useful when you want to know which outcomes are most likely, the spread of potential values, and the likelihood of different results. The sum of all probabilities for all possible values must equal 1. Furthermore, the probability for a particular value or range of values must be between 0 and 1. Probability distributions, therefore, describe the dispersion of the values of a random variable.

A random variable is a real valued function from the probability space. We can compute the probability that a random variable takes values in an interval by subtracting the distribution function evaluated at the endpoints of the intervals.

3.2.1 Function of a Random Variable

A *random variable is a variable that takes on different values as a result of the outcomes of a random experiment.* In other words, a function which assigns numerical values to each element of the set of events that may occur (i.e., every element in the sample space) is termed a random variable. The value of a random variable is the general outcome of the random experiment. One should always make a distinction between the random variable and the values that it can take on. All this can be illustrated by a few examples shown in the Table 3.1.

Table 3.1 *Random Variable*

Random Variable	*Values of the Random Variable*	*Description of the Values of the Random Variable*
X	0, 1, 2, 3, 4	Possible number of heads in four tosses of a fair coin
Y	1, 2, 3, 4, 5, 6	Possible outcomes in a single throw of a die
Z	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	Possible outcomes from throwing a pair of dice
M	0, 1, 2, 3, ... S	Possible sales of newspapers by a newspaper boy, S representing his stock

NOTES

All these above stated random variable assignments cover every possible outcome and each numerical value represents a unique set of outcomes. *A random variable can be either discrete or continuous*. If a random variable is allowed to take on only a limited number of values, it is a discrete random variable but if it is allowed to assume any value within a given range, it is a continuous random variable. Random variables presented in the above table are examples of discrete random variables. We can have continuous random variables if they can take on any value within a range of values, for example, within 2 and 5, in that case we write the values of a random variable x as follows:

$$2 \leq x \leq 5$$

Techniques of Assigning Probabilities

We can assign probability values to the random variables. Since the assignment of probabilities is not an easy task, we should observe certain rules in this context as given below:

(*i*) A probability cannot be less than zero or greater than one, i.e., $0 \leq pr \leq 1$, where pr represents probability.

(*ii*) The sum of all the probabilities assigned to each value of the random variable must be exactly one.

There are three techniques of assignment of probabilities to the values of the random variable:

(a) **Subjective Probability Assignment.** It is the technique of assigning probabilities on the basis of personal judgement. Such assignment may differ from individual to individual and depends upon the expertise of the person assigning the probabilities. It cannot be termed as a rational way of assigning probabilities but is used when the objective methods cannot be used for one reason or the other.

(b) **A-Priori Probability Assignment.** It is the technique under which the probability is assigned by calculating the ratio of the number of ways in which a given outcome can occur to the total number of possible outcomes. The basic underlying assumption in using this procedure is that every possible outcome is likely to occur equally. But at times the use of this technique gives ridiculous conclusions. For example, we have to assign probability to the event that a person of age 35 will live upto age 36. There are two possible outcomes, he lives or he dies. If the probability assigned in accordance with a-priori probability assignment is half then the same may not represent reality. In such a situation, probability can be assigned by some other techniques.

(c) **Empirical Probability Assignment.** It is an objective method of assigning probabilities and is used by the decision-makers. Using this technique the probability is assigned by calculating the relative frequency of occurrence

NOTES

of a given event over an infinite number of occurrences. However, in practice only a finite (perhaps very large) number of cases are observed and relative frequency of the event is calculated. The probability assignment through this technique may as well be unrealistic, if future conditions do not happen to be a reflection of the past.

Thus, what constitutes the 'best' method of probability assignment can only be judged in the light of what seems best to depict reality. It depends upon the nature of the problem and also on the circumstances under which the problem is being studied.

Variance and Standard Deviation of Random Variable

The mean or the expected value of random variable may not be adequate enough at times to study the problem as to how random variable actually behaves and we may as well be interested in knowing something about how the values of random variable are dispersed about the mean. In other words, we want to measure the dispersion of random variable (X) about its expected value, i.e., E(X). The variance and the standard deviation provide measures of this dispersion.

The variance of random variable is defined as the sum of the squared deviations of the values of random variable from the expected value weighted by their probability. Mathematically, we can write it as follows:

$$\text{Var}(X) = \sigma_X^2 = \sum_{i=1}^{n}\left[X_i - E(X)\right]^2 .pr.(X_i)$$

Alternatively, it can also be written as,

$$\text{Var}(X) = \sigma_X^2 = \Sigma X_i^{\,2}\, pr.(X_i) - \left[E(X)\right]^2$$

Where, E (X) is the expected value of random variable.

X_i is the ith value of random variable.

$pr.$ (X_i) is the probability of the ith value.

The standard deviation of random variable is the square root of the variance of random variable and is denoted as,

$$\sqrt{\sigma_X^2} = \sigma_X$$

The variance of a constant time random variable is the constant squared times the variance of random variable. This can be symbolically written as,

$$\text{Var } (cX) = c^2 \text{ Var } (X)$$

The variance of a sum of independent random variables equals the sum of the variances.

Thus,

$$\text{Var}(X + Y + Z) = \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z)$$

If X, Y and Z are independent of each other.

The following examples will illustrate the method of calculation of these measures of a random variable.

Example 1: Calculate the mean, the variance and the standard deviation for random variable sales from the following information provided by a sales manager of a certain business unit for a new product:

Monthly Sales (in units)	*Probability*
50	0.10
100	0.30
150	0.30
200	0.15
250	0.10
300	0.05

Solution:

The given information may be developed as shown in the following table for calculating mean, variance and the standard deviation for random variable sales:

Monthly Sales (in units)[1] X_i		*Probability* $pr(X_i)$	$(X_i)\,pr(X_i)$	$(X_i - E(X))^2$	$X_i - E(X)^2$ $pr(X_i)$
X_1	50	0.10	5.00	$(50 - 150)^2$ = 10000	1000.00
X_2	100	0.30	30.00	$(100 - 150)^2$ = 2500	750.00
X_3	150	0.30	45.00	$(150 - 150)^2$ = 0	0.00
X_4	200	0.15	30.00	$(200 - 150)^2$ = 2500	375.00
X_5	250	0.10	25.00	$(250 - 150)^2$ = 10000	1000.00
X_6	300	0.5	15.00	$(300 - 150)^2$ = 22500	1125.00
			$\Sigma(X_i)\,pr(X_i)$ = 150.00		$\Sigma[X_i - E(X)^2]$ $pr(X_i)$ = 4250.00

Mean of random variable sales = $\overline{X}$

or, $\quad E(X) = \Sigma(X_i).pr(X_i) = 150$

Variance of random variable sales,

NOTES

Or, $$\sigma_X^2 = \sum_{i=1}^{n}\left(X_i - E(X)\right)^2 .pr(X_i) = 4250$$

Standard deviation of random variable sales,

Or, $$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{4250} = 65.2 \text{ approx.}$$

The mean value calculated above indicates that in the long run the average sales will be 150 units per month. The variance and the standard deviations measure the variation or dispersion of random variable values about the mean or the expected value of random variable.

Example 2: Given are the mean values of four different random variables viz., A, B, C, and D.

$$\overline{A} = 20, \quad \overline{B} = 40, \quad \overline{C} = 10, \quad \overline{D} = 5$$

Find the mean value of the random variable $(A + B + C + D)$

Solution:

$$\because \quad E(A + B + C + D) = E(A) + E(B) + E(C) + E(D)$$
$$= \overline{A} + \overline{B} + \overline{C} + \overline{D}$$
$$= 20 + 40 + 10 + 5$$
$$= 75$$

Hence, the mean value of random variable $(A + B + C + D)$ is 75.

Example 3: If X represents the number of heads that appear when one coin is tossed and Y the number of heads that appear when two coins are tossed, compare the variances of the random variables X and Y. The probability distributions of X and Y are as follows:

X_i	$pr(X_i)$	Y_i	$pr(Y_i)$
0	1/2	0	1/4
		1	1/2
1	1/2	2	1/4
	Total = 1		Total = 1

Solution:

Variance of $X = \sigma_X^2 = \left(0-\frac{1}{2}\right)^2 .\left(\frac{1}{2}\right) + \left(1-\frac{1}{2}\right)^2 .\left(\frac{1}{2}\right)$

$$= \frac{1}{8} + \frac{1}{8} = \frac{1}{4} = 0.25$$

Variance of $Y = \sigma_Y^2 = (0-1)^2 .\left(\frac{1}{4}\right) + (1-1)^2 .\left(\frac{1}{2}\right)$

$$+ (2-1)^2 .\left(\frac{1}{4}\right) = \frac{1}{4} + 0 + \frac{1}{4} = 0.50$$

The variance of the number of heads for two coins is double the variance of the number of heads for one coin.

3.2.2 Moment Generating Functions

According to probability theory, moment generating function generates the moments for the probability distribution of a random variable X, and can be defined as:

$$M_X(t) = E\,(e^{tX}), \qquad t \in \overset{+}{\mathrm{R}}$$

When the moment generating function exists with an interval $t = 0$, the nth moment becomes,

$$E(X^n) = M_X^{(n)}\,(0) = [d^n\, M_X\,(t)\,/\,dt^n]_{\,t=0}$$

The moment generating function, for probability distribution condition being continuous or not, can also be given by Riemann-Stieltjes integral,

$$M_X(t) = \int_{-\infty}^{\infty} e^{tX}\, dF(x)$$

Where F is the cumulative distribution function.

The probability density function $f(x)$, for X having continuous moment generating function becomes,

$$M_X(t) = \int_{-\infty}^{\infty} e^{tX} f(x)\; dx$$
$$= \int_{-\infty}^{\infty} (1 + tx + t^2x^2/2!\; + \ldots.)\, f(x)\; dx$$

***Note*:** The moment generating function of X always exists, when the exponential function is positive and is either a real number or a positive infinity.

1. Prove that when X shows a discrete distribution having density function f, then,

$$M_X(t) = \sum_{x \in S} e^{tx} f(x)$$

2. When X is continuous with density function f, then,

$$M_X(t) = \int_S e^{tx} f(x)dx$$

3. Consider that X and Y are independent. Show that,

$$M_{X+Y}(t) = M_X(t)\, M_Y(t)$$

3.2.3 Probability Density Functions—Discrete and Continuous

Probability density may be defined as a measure of existence. Logic can be analysed with the help of a graph. In graphical representation, a probability density functions gives information about the existence of samples at various locations and the entire graph area can be considered as a sample space.

A mathematical description of samples in regions can be given under a coordinate system. The regions and the existence of samples at various locations can be described in this coordinate system by mathematical functions. For example, to locate a point in a sample space it is expressed as a pair of (x,y) coordinates and a mathematical function describes a location for each sample.

The function that shows the location of each sample is called a 'density function'. By showing locations of each sample, density of the samples can be understood. If we assume that the odds of selecting any one sample is the same as for any other, then the PDF (probability density function), $p(x,y)$, gives the probability of a point in sample space to lie between two given limits of the variables.

Thus, in case of a discrete variable x, the probability function, PDF(x), is the probability of occurrence of x, variable of continuous nature, PDF(x) shows probability density of a variable x. Thus, probability of a value between x and $x+dx$ is PDF(x) × dx.

A 'cumulative density function', CDF(x), shows the probability that the said variable assumes a value $<= x$.

The *mean* or the average value is available in majority of cases. This MEAN is taken as sum of products x × PDF(x). In case of continuous variables, MEAN is given by an integral of X × PDF(x), integrated in the range.
Let f be a non-negative fucntion mapped $\mathbf{R} \to \mathbf{R}$ then the probability distribution has density f probability in the interval $[a, b]$ is given as,

$\int_a^b f(x)\,dx$ in case of any two numbers a and b

Total integral of f must be 1. It converse is also true which tells that a function f with total integral is equal to 1 then, f represents the probability density for some probability distribution.

A probability density function is in fact, a refined version of a histogram with very small or infinitesimal interval. Due to this, the curve is smooth. If sampling is done, taking many values of a random variable which is of continuous nature, a histogram is produced that depicts a histogram showing probability density, in a very narrow output range. This can be termed a probability density function if and only if it is non-negative and the area under the graph is 1. Putting mathematically with logical connective showing 'conjunction', it is given as:

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

All distributions are not showing density function. It is said to have a density function *f(x)* only if its CDF, denoted as $F(x)$, is continuous.

This is expressed mathematically as:

$$\frac{d}{dx}F(x) = f(x)$$

Link between Discrete and Continuous Distributions

A PDF describes association of variable in a distribution that is continuous: taking a set of two state variable in the interval [*a, b*].

Discrete random variables may be represented probability distribution using a delta function. If we consider a binary discrete random variable by taking two distinct value, say 1 which are equally likely, probability density of such a variable is given by:

$$f(t) = \frac{1}{2}(\delta(t+1) + \delta(t-1))$$

We may generalize this as follows: If a discrete variable assumes '*n*' different values in the set of real numbers, then the associated probability density function is given by:

$$f(t) = \sum_{i=1}^{n} P_i \delta(t - x_i)$$

Where i = 1, 2,3....., n, and x_1 ......,. x_n stand for discrete values for the variables and SP_i (where i = 1, 2, …., n) are probabilities associated with these values.

The method is used for knowing the characteristics the mean, its variance and kurtosis.

The method is used to show mathematically the characteristic of Brownian movement and deciding on its initial configuration.

Probability Functions Associated with Multiple Variables

In case of random variables, of continuous nature, SX_i, where i = 1, 2,3, 4, ….., n, one can define a PDF for the whole set. This is a term coined as joint PDF which is defined as a function with n variables. This is known by a different name MDF which means marginal density function.

Independence

Random variables X_1, …, …, X_n of continuous nature are independent of each other if and only if $fX_1, ..., X_n (x_1, ..., x_n) = fX_1(x_1) \ldots fX_n(x_n)$.

Corollary

If a JPDF (joint probability distribution function) of a vector of n random variables shown as a product of n functions of one variable $fX_1, ..., X_n (x_1, ..., x_n) = f_1(x_1)$

NOTES

$\cdot f_n(x_n)$, then all these variables are independent of each other. MPDF (marginal probability density function) for each is expressed as:

$$fX_i(x_i) = \frac{f_i(x_i)}{\int f_i(x)\,dx}$$

For example, this illustrates the definition of a MDPF (multidimensional probability density functions) as stated above. This condition is considered for function with a set of two variables. This condition is considered in case of a function of a set of two variables. Let us call $\vec{R}$ a 2-dimensional random vector of coordinates (X,Y): the probability to obtain $\vec{R}$ in the quarter plane of positive x and y is

$$\Pr(X > 0, Y > 0) = \int_0^\infty \int_0^\infty fX, Y(x, y)\,dx\,dy$$

Sums of Independent Random Variables

Let there be two random variables, u and v. Here each of which has a PDF; then the sum of these two PDFs, is taken as convolution of these two and is the convolution of their separate density functions. This is given mathematically, as below:

$$fu + v(x) = \int_{-\infty}^{\infty} fu(y)\,fv\ (x - y)\,dy$$

Check Your Progress

1. Define variance of random variable.
2. Explain discrete and continuous random variables.
3. Define variance of random variable.
4. Explain the moment generating function with reference to probability theory.
5. What is a density function?
6. What is meant by independence of a random variable?

3.3 EXPECTATION OF RANDOM VARIABLE

The **expected value** (or mean) of X is the weighted average of the possible values that X can take. Here X is a discrete random variable and each value is being weighted according to the probability of the possibility of the occurrence of the event. The expected value of X is usually written as $E(X)$ or μ.

$$E(X) = \Sigma \times P(X = x)$$

Hence, the expected value is the sum of:

Each of the possible outcomes + The probability of the outcome occurring

Therefore, the expectation is the outcome you expect of an experiment.

Let us consider the following example,

What is the expected value when we roll a fair die?

There are six possible outcomes 1, 2, 3, 4, 5, 6. Each one of these has a probability of 1/6 of occurring. Let X be the outcome of the experiment.

Then,

$P(X=1) = 1/6$ (this shows that the probability that the outcome of the experiment is 1 is 1/6)

$P(X=2) = 1/6$ (the probability that you throw a 2 is 1/6)

$P(X=3) = 1/6$ (the probability that you throw a 3 is 1/6)

$P(X=4) = 1/6$ (the probability that you throw a 4 is 1/6)

$P(X=5) = 1/6$ (the probability that you throw a 5 is 1/6)

$P(X=6) = 1/6$ (the probability that you throw a 6 is 1/6)

$E(X) = 1\times P(X = 1) + 2\times P(X = 2) + 3\times P(X = 3) + 4\times P(X=4) + 5\times P(X=5) + 6 \times P(X = 6)$

Therefore,

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

Hence, the expectation is 3.5, which is also the halfway between the possible values the die can take, and so this is what you should have expected.

Expected Value of a Function of X

To find $E[f(X)]$, where $f(X)$ is a function of X, we use the following formula:

$$E[f(X)] = \Sigma f(x)P(X = x)$$

Let us consider the above example of die, and calculate $E(X^2)$

Using the notation above, $f(x) = x^2$

$f(1) = 1, f(2) = 4, f(3) = 9, f(4) = 16, f(5) = 25, f(6) = 36$

$P(X = 1) = 1/6, P(X = 2) = 1/6$, etc.

Hence, $E(X^2) = 1/6 + 4/6 + 9/6 + 16/6 + 25/6 + 36/6 = 91/6 = 15.167$

The expected value of a constant is just the constant, as for example $E(1) = 1$. Multiplying a random variable by a constant multiplies the expected value by that constant.

Therefore, $E[2X] = 2E[X]$

An important formula, where a and b are constants, is:

$$E[aX + b] = aE[X] + b$$

Hence, we can say that the expectation is a linear operator.

Variance

The variance of a random variable tells us something about the spread of the possible values of the variable. For a discrete random variable X, the variance of X is written as Var(X).

$$\text{Var}(X) = E[(X - \mu)^2]$$

Where, μ is the expected value $E(X)$

This can also be written as:

$$\text{Var}(X) = E(X^2) - \mu^2$$

The standard deviation of X is the square root of Var(X).

Note: The variance does not behave in the same way as expectation, when we multiply and add constants to random variables.

$$\text{Var}[aX + b] = a^2\text{Var}(X)$$

Because, $\text{Var}[aX + b] = E[\,(aX + b)^2\,] - (E\,[aX + b])^2$

$$= E[\,a^2X^2 + 2abX + b^2] - (aE(X) + b)^2$$

$$= a^2E(X^2) + 2abE(X) + b^2 - a^2E^2(X) - 2abE(X) - b^2$$

$$= a^2E(X^2) - a^2E^2(X) = a^2\text{Var}(X)$$

3.4 CHEBYSHEV'S INEQUALITY

The Chebyshev polynomials $\{T_n(x)\}$ are orthogonal on (–1, 1) with respect to the weight function $w(x) = (1 - x^2)^{-1/2}$. The Chebyshev polynomial is defined by the following relation:

For $x \in [-1, 1]$, define

$$T_n(x) = \cos\,(n \cos^{-1}(x)) \text{ for each } n \geq 0.$$

It is not obvious from this definition that $T_n(x)$ is an n^{th} degree polynomial in x, but now it will be proved that it is. First note that

$$T_0(x) = \cos 0 = 1 \quad \text{and} \quad T_1(x) = \cos\,(\cos^{-1}(x)) = x.$$

For $n \geq 1$, introduce the substitution $\theta = \cos^{-1}x$ to change this equation to

$$T_n(x) = T_n(\theta(x)) \equiv T_n(\theta) = \cos(n\theta), \quad \text{where } \theta \in |\,0, \pi\,|.$$

A recurrence relation is derived by noting that:

$$T_{n+1}(\theta) = \cos(n\theta + \theta) = \cos(n\theta)\cos\theta - \sin(n\theta)\sin\theta$$

And $$T_{n-1}(\theta) = \cos(n\theta - \theta) = \cos(n\theta)\cos\theta + \sin(n\theta)\sin\theta.$$

Adding these equations gives the following:

$$T_{n+1}(\theta) + T_{n-1}(\theta) = 2\cos(n\theta)\cos\theta.$$

Returning to the variable x and solving for $T_{n+1}(x)$ you have, for each $n \geq 1$,

$$T_{n+1}(\theta) = 2\cos(n \arccos x).x - T_{n-1}(x) = 2T_n(x)\,x - T_{n-1}(x).$$

Since $T_0(x)$ and $T_1(x)$ are both polynomials in x, $T_{n+1}(x)$ will be a polynomial in x for each n.

[Chebyshev Polynomials] $T_0(x) = 1, \quad T_1(x) = x,$

and, for $n \geq 1$, $T_{n+1}(x)$ is the polynomial of degree $n + 1$ given by

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).$$

The recurrence relation implies that $T_n(x)$ is a polynomial of degree n, and it has leading coefficient 2^{n-1}, when $n \geq 1$. The next three Chebyshev polynomials therefore are as follows:

$$T_2(x) = 2xT_1(x) - T_0(x) = 2x^2 - 1.$$

$$T_3(x) = 2xT_2(x) - T_1(x) = 4x^3 - 3x.$$

And $$T_4(x) = 2xT_3(x) - T_2(x) = 8x^4 - 8x^2 + 1.$$

The graphs of $T_1, T_2, T_3,$ and T_4 are shown in Figure 3.1. Notice that each of the graphs is symmetric to either the origin or the y-axis, and that each assumes a maximum value of 1 and a minimum value of –1 on the interval [–1, 1].

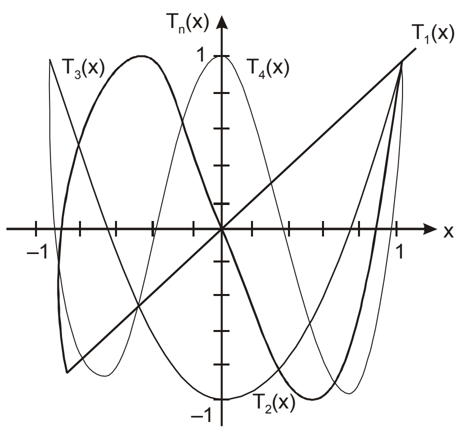


***Fig. 3.1** Graph of Chebyshev Polynomials*

NOTES

The first few Chebyshev polynomials are summarized as follows:

$$T_0(x) = 1$$
$$T_1(x) = x$$
$$T_2(x) = 2x^2 - 1$$
$$T_3(x) = 4x^3 - 3x$$
$$T_4(x) = 8x^4 - 8x^2 + 1$$
$$T_5(x) = 16x^5 - 20x^3 + 5x$$
$$T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1$$

Note: The coefficients of x^n in $T_n(x)$ is always 2^{n-1}.

By using the Chebyshev polynomials, you can express $1, x, x^2, \ldots x^6$ as follows:

$$1 = T_0(x)$$

$$x = T_1(x)$$

$$x^2 = \frac{1}{2}[T_2(x) + T_0(x)]$$

$$x^3 = \frac{1}{4}[3T_1(x) + T_3(x)]$$

$$x^4 = \frac{1}{8}[3T_0(x) + 4T_2(x) + T_4(x)]$$

$$x^5 = \frac{1}{16}[10T_1(x) + 5T_3(x) + T_5(x)]$$

and $$x^6 = \frac{1}{32}[10T_0(x) + 15T_2(x) + 6T_4(x) + T_6(x)]$$

These expressions are useful in the economization of power series.

Thus, after omitting $T_5(x)$ you have, $\sin x - 0.8802\ T_1(x) - 0.03906\ T_3(x)$

substituting $T_1(x) = x$ and $T_3(x) = 4x^3 - 3x$ you have,

$$\sin x = 0.9974x - 0.1562x^3$$

which gives $\sin x$ correctly upto three significant digits with only two terms for any value of x.

Chebychev Inequality

If X is a random variable having μ as expected value and σ^2 as finite variance, then for a positive real number k

NOTES

$$\Pr(|X - \mu| \geq k\alpha) \leq \frac{1}{k^2}.$$

Only when, $k > 1$, we can get useful information. This is equivalent to

$$\Pr(|X - \mu| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}.$$

For example, when $k = \sqrt{2}$, at least half of these values lie in the open interval $(\mu - \sqrt{\sigma}, \sigma\mu + \sqrt{2}\sigma)$.

The theorem provides loose bounds. However, bounds provided as per Chebychev's inequality cannot be improved upon. For example, when $k > 1$, following example having $\sigma = 1/k$, meets such bounds exactly.

$$\Pr(X = -1) = 1/(2k^2),$$
$$\Pr(X = 0) = 1 - 2k^2,$$
$$\Pr(X = 1) = 1/(2k^2),$$
$$\Pr(|X - \mu| \geq k\sigma) = 1/k^2,$$

Equality holds exactly in case of a linear transformation whereas inequality holds for distribution which is non-linear transformation.

Theorem is useful as it applies to random variables for any distribution and these bounds can be computed by knowing only mean and variance.

Any observation, howsoever accurate it may be, is never more than few standard deviations away from the mean. Chebyshev's inequality gives following bounds that apply to all distributions in which it is possible to define standard deviation.

At least:

- 50% of these values lie within standard deviations = $\sqrt{2}$
- 75% of the values lie within standard deviations = 2
- 89% lie within standard deviations = 3
- 94% lie within standard deviations = 4
- 96% lie within standard deviations = 5
- 97% lie within standard deviations = 6

Standard deviations are always taken from the mean.

Generally:

Minimum $(1 - 1/k^2) \times 100\%$ lie with standard deviations = k.

Example 4: Represent $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} \ldots$ by using Chebyshev polynomial to obtain three significant digits accuracy in the computation of $\sin x$.

NOTES

Solution: Using Chebyshev polynomial, sin x is approximated as:

$$\sin x = T_1(x) - \frac{1}{24}[3T_1(x) + T_3(x)] + \frac{1}{1920}[10T_1(x) + 5T_3(x) + T_5(x)]$$

$$\simeq 0.8802\, T_1(x) - 0.03906\, T_3(x) + 0.00052\, T_5(x)$$

As the coefficient of $T_5(x)$ is 0.00052 and as $|\,T_5(x)\,| \le 1$ for all x, therefore the truncation error if we omit the last term above still gives three significant digit accuracy.

Example 5: Express the Taylor series expansion of

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} + ...$$

in terms of Chebyshev polynomials.

Solution: The Chebyshev polynomial representation of

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + ...$$

as $e^{-x} = T_0(x) - T_1(x) + \frac{1}{4}[T_2(x) + T_0(x)] - \frac{1}{24}[3T_1(x) + T_3(x)]$

$$+\frac{1}{195}[3T_0(x) + 4T_2(x) + T_4(x)] - \frac{1}{1920}[10T_1(x) + 5T_3(x) + T_5(x)] + ...$$

Thus,

$$e^{-x} = 1.26606\, T_0(x) - 1.13021\, T_1(x) + 0.27148\, T_2(x)$$
$$-0.04427\, T_3(x) + 0.0054687\, T_4(x) \; -0.0005208\, T_5(x) + ...$$

Now, if you expand $T_0(x), T_1(x), T_2(x), T_3(x), T_4(x)$ and $T_5(x)$ using their polynomial equivalents and truncate after six terms, than you have,

$$e^{-x} = 1.00045 - 1.000022x + 0.4991992x^2 - 0.166488x^3$$
$$+ 0.043794x^4 - 0.008687x^5$$

Comparing this representation with the Taylor series representation, you observe that there is a slight difference in the coefficients of different powers of x. The main advantage of this representation as a sum of Chebyshev polynomials is that, for a given error bound, you can truncate the series with a smaller number of terms compared to the Taylor series. Also, the error is more uniformly distributed for various arguments. The possibility of a series with a lower number of terms is called economization of power series. The maximum error in the six terms of Chebyshev representation of e^{-x} is 0.00045 whereas the error in the six terms of Taylor series representation of e^{-x} is 0.0014. Thus, you have to add one more term in Taylor series to ensure that the error is less than that in the Chebyshev approximation.

NOTES

Check Your Progress

7. What is an expected value?
8. Define variance with respect to discrete random variable.
9. Summarize the first few Chebyshev polynomials.

3.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A random variable is a variable that takes on different values as a result of the outcomes of a random experiment.
2. A random variable can be either discrete or continuous. If a random variable is allowed to take on only a limited number of values, it is a discrete random variable but if it is allowed to assume any value within a given range, it is a continuous random variable.
   (*i*) A probability cannot be less than zero or greater than one, i.e., $0 \leq pr \leq 1$, where, *pr* represents probability.
   (*ii*) The sum of all the probabilities assigned to each value of the random variable must be exactly one.
3. The variance of random variable is defined as the sum of the squared deviations of the values of random variable from the expected value weighted by their probability.
4. According to probability theory, moment generating function generates the moments for the probability distribution of a random variable *X,* and can be defined as:

   $M_X(t) = E\,(e^{tX}), \qquad t \in \overset{+}{R}$
5. The function that shows the location of each sample is called a 'density function'. By showing locations of each sample, density of the samples can be understood.
6. Random variables $X_1, \ldots\ldots, X_n$ of continuous nature are independent of each other if and only if $fX_1, \ldots., X_n\,(x_1, \ldots, x_n) = fX_1(x_1) \cdots fX_n(x_n)$.
7. The expected value (or mean) of *X* is the weighted average of the possible values that *X* can take. Here *X* is a discrete random variable and each value is being weighted according to the probability of the possibility of the

NOTES

occurrence of the event. The expected value of X is usually written as $E(X)$ or μ.

8. The variance of a random variable tells us something about the spread of the possible values of the variable. For a discrete random variable X, the variance of X is written as Var(X).

   $\text{Var}(X) = E[(X - \mu)^2]$

   Where, μ is the expected value $E(X)$.

9. The first few Chebyshev polynomials can be summarized as follows:

$$T_0(x) = 1$$

$$T_1(x) = x$$

3.6 SUMMARY

- A function which assigns numerical values to each element of the set of events that may occur (i.e., every element in the sample space) is termed a random variable.
- If a random variable is allowed to take on only a limited number of values, it is a discrete random variable but if it is allowed to assume any value within a given range, it is a continuous random variable.
- The sum of all the probabilities assigned to each value of the random variable must be exactly one.
- we want to measure the dispersion of random variable (X) about its expected value, i.e., E(X). The variance and the standard deviation provide measures of this dispersion.
- The variance of random variable is defined as the sum of the squared deviations of the values of random variable from the expected value weighted by their probability.
- According to probability theory, moment generating function generates the moments for the probability distribution of a random variable X, and can be defined as:

$$M_X(t) = E\,(e^{tX}), \qquad t \in \overset{+}{R}$$

- In graphical representation, a probability density functions gives information about the existence of samples at various locations and the entire graph area can be considered as a sample space.

NOTES

- The mean or the average value is available in majority of cases. This MEAN is taken as sum of products $x \times$ PDF(x). In case of continuous variables, MEAN is given by an integral of X × PDF(x), integrated in the range.
- If a JPDF (joint probability distribution function) of a vector of n random variables shown as a product of n functions of one variable $fX_1, ..., X_n (x_1, ..., x_n) = f_1(x_1) \cdot f_n(x_n)$, then all these variables are independent of each other.
- The maximum error in the six terms of Chebyshev representation of e^{-x} is 0.00045 whereas the error in the six terms of Taylor series representation of e^{-x} is 0.0014.

3.7 KEY WORDS

- **Expected value:** The **expected value** (or mean) of X is the weighted average of the possible values that X can take.
- **Variance:** The variance of a random variable tells us something about the spread of the possible values of the variable.
- **Chebyshev polynomials:** The Chebyshev polynomials $\{T_n(x)\}$ are orthogonal on (–1, 1) with respect to the weight function $w(x) = (1 - x^2)^{-1/2}$.

3.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Explain the techniques of assigning probability.
2. What is moment generating function?
3. Explain the concept of expectation of a random variable.
4. Explain briefly the terms 'expectation' and 'expected value'.
5. Define variance.

Long-Answer Questions

1. Discuss the techniques of assigning probabilities.
2. Give the mathematically representation of the variance of random variable.
3. Define the various theories of moment generating function with syntax and example.
4. Explain the probability density functions of discrete and continuous type.
5. Drive the relation of expected value of a function of X.
6. Briefly describe about the Chebyshev inequality.

3.9 FURTHER READINGS

NOTES

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

NOTES

UNIT 4 DISTRIBUTION OF RANDOM VARIABLES

Structure

4.0 INTRODUCTION

A random variable is a phenomenon of interest in which the observed outcomes of an activity are entirely by chance, are absolutely unpredictable and may differ from response to response. By definition of randomness, each possible entity has the same chance of being considered. For instance, lottery drawings are considered to be random drawings so that each number has exactly the same chance of being picked up. Similarly, the value of the outcome of a toss of a fair coin is random, since a head or a tail has the same chance of occurring. A random variable may be qualitative or quantitative in nature. The qualitative random variables yield categorical responses so that the responses fit into one category or another. For example, a response to a question such as 'Are you currently unemployed?' would fit in the category of either 'Yes' or 'No'. On the other hand, quantitative random variables yield numerical responses.

In this unit, you will study about the multivariate distribution, distribution of two random variable, conditional distribution and expectation.

4.1 OBJECTIVES

After going through this unit, you will be able to:

- Analyse the multivariate distributions
- Explain the distribution of two random variable
- Discuss about the conditional distribution and expectation

4.2 MULTIVARIATE DISTRIBUTION

NOTES

Uniform or Rectangular Distribution

Each possible value of the random variable x has the same probability in the uniform distribution. If x takes vaues $x_1, x_2,....,x_k$, then,

$$p(x_i,\ k) = \frac{1}{k}$$

The numbers on a die follow the uniform distribution,

$$p(x_i,\ 6) = \frac{1}{6} \text{ (Here } x = 1, 2, 3, 4, 5, 6)$$

Bernoulli Trials

In a Bernoulli experiment, an Event E either happens or does not happen (E'). Examples are, getting a head on tossing a coin, getting a six on rolling a die, and so on.

The Bernoulli random variable is written,

$$X = 1 \text{ if } E \text{ occurs}$$
$$= 0 \text{ if } E' \text{ occurs}$$

Since there are two possible value it is a case of a discrete variable where,

Probability of success = $p = p(E)$

Profitability of failure = $1 - p = q = \text{p}(E')$

We can write,

For $k = 1, f(k) = p$

For $k = 0, f(k) = q$

For $k = 0$ or $1, f(k) = p^k q^{1-k}$

Geometric Distribution

Suppose the probability of success p in a series of independent trials remains constant.

Suppose, the first success occurs after x failures, i.e., there are x failures preceding the first success. The probability of this event will be given by $p(x) = q^x p$ $(x = 0, 1, 2, ...)$

This is the geometric distribution and can be derived from the negative binomial. If we put $r = 1$ in the Negative Binomial distribution:

$$p(x) = {}^{x+r-1}C_{r-1} p^{r-1} q^x$$

We get the geometric distribution,

$$p(x) = {}^{x}C_0 p^1 q^x = pq^x$$

$$\Sigma p(x) = \sum_{n=0}^{p} q^x p = \frac{p}{1-q} = 1$$

$$E(x) = \text{Mean} = \frac{p}{q}$$

$$\text{Variance} = \frac{p}{q^2}$$

$$\text{Mode} = \left(\frac{1}{2}\right)^x$$

Example 1: Find the expectation of the number of failures preceding the first success in an infinite series of independent trials with constant probability p of success.

Solution: The probability of success in,

1st trial $= p$ (Success at once)

2nd trial $= qp$ (One failure then success, and so on)

3rd trial $= q^2p$ (Two failures then success, and so on)

The expected number of failures preceding the success,

$$E(x) = 0 \,.\, p + 1.\, pq + 2p^2p + \ldots\ldots\ldots$$

$$= pq(1 + 2q + 3q^2 + \ldots\ldots)$$

$$= pq\frac{1}{(1-q)^2} = qp\frac{1}{p^2} = \frac{q}{p}$$

Since $p = 1 - q$.

Hypergeometric Distribution

From a finite population of size N, a sample of size n is drawn without replacement.

Let there be N_1 successes out of N.

The number of failures is $N_2 = N - N_1$

The disribution of the random variable X, which is the number of successes obtained in the above case, is called the Hypergeometric distribution.

$$p(x) = \frac{{}^{N_1}C_x {}^{N}C_{n-x}}{{}^{N}C_n} \quad (X = 0, 1, 2, \ldots., n)$$

Here x is the number of successes in the sample and $n - x$ is the number of failures in the sample.

NOTES

It can be shown that,

$$\text{Mean : } E(X) = n\frac{N_1}{N}$$

$$\text{Variance : Var (X)} = \frac{N-n}{N-1}\left(\frac{nN_1}{N} - \frac{nN_1^2}{N}\right)$$

Example 2: There are 20 lottery tickets with three prizes. Find the probability that out of 5 tickets purchased exactly two prizes are won.

Solution: We have $N_1 = 3$, $N_2 = N - N_1 = 17$, $x = 2$, $n = 5$.

$$p(2) = \frac{{}^3C_2\ {}^{17}C_3}{{}^{20}C_5}$$

The probability of getting no prize $p(0) = \dfrac{{}^3C_0\ {}^{17}C_5}{{}^{20}C_5}$

The probability of getting exactly 1 prize $p(1) = \dfrac{{}^3C_1\ {}^{17}C_4}{{}^{20}C_5}$

Example 3: Examine the nature of the distibution of r balls are drawn, one at a time without replacement, from a bag containing m white and n black balls.

Solution: It is the hypergeometric distribution. It corresponds to the probability that x balls will be white out of r balls so drawn and is given by,

$$p(x) = \frac{{}^xC_x\ {}^nC_{r-x}}{{}^{m+n}C_r}$$

Multinomial

There are k possible outcomes of trials, viz., $x_1, x_2, ..., x_k$ with probabilities $p_1, p_2, ..., p_k$, n independent trials are performed. The multinomial distibution gives the probability that out of these n trials, x_1 occurs n_1 times, x_2 occurs n_2 times, and so on. This is given by the following equation: $\dfrac{n!}{n_1!n_2!....n_k!}\ p_1^{n_1} p_2^{n_2} ....p_k^{n}$

Where, $\displaystyle\sum_{i=1}^{k} n_i = n$

Characteristic Features of the Binomial Distribution

The following are the characteristics of Binomial distribution:

1. It is a discrete distribution.
2. It gives the probability of x successes and $n-x$ failures in a specific order.
3. The experiment consists of n repeated trials.

4. Each trial results in a success or a failure.
5. The probability of success remains constant from trial to trial.
6. The trials are independent.
7. The success probability p of any outcome remains constant over time. This condition is usually not fully satisfied in situations involving management and economics, for example, the probability of response from successive informants is not the same. However, it may be assumed that the condition is reasonably well satisfied in many cases and that the outcome of one trial does not depend on the outcome of another. This condition too, may not be fully satisfied in many cases. An investigator may not approach a second informant with the same set-up of mind as used for the first informant.
8. The Binomial distribution depends on two parameters n and p. Each set of different values of n, p has a different Binomial distribution.
9. If $p = 0.5$, the distribution is *symmetrical.* For a symmetrical distribution, in n

   $$\text{Prob } (X = 0) = \text{Prob } (X = n)$$

   i.e., the probabilities of 0 or n successes in n trials will be the same. Similarly,

   $$\text{Prob } (X = 1) = \text{Prob}(X = n - 1), \text{ and so on.}$$

   If $p > 0.5$, the distribution is not symmetrical. The probabilities on the right are larger than those on the left. The reverse case is when $p < 0.5$.

   When n becomes large, the distribution becomes bell shaped. Even when n is not very large but $p \cong 0.5$, it is fairly bell shaped.
10. The Binomial distribution can be approximated by the normal. As n becomes large and p is close to 0.5, the approximation becomes better.

Example 4: If the ratio n/N, i.e., sample size to population size is small, the result given by the Binomial may not be reliable. Comment.

Solution: When the distribution is Binomial, each successive trial, being independent of other trials, has constant probability of success. If he sampling of n items is without replacement from a population of size N, the probability of success of any event depends upon what happened in the previous events. In this case the Bionomial cannot be used unless the ratio n/N is small. Even then there is no guarantee of getting accurate results.

The Binomial should be used only if the ratio $\frac{n}{N}$ is very small, say less that 0.05.

Example 5: Explain the concept of a discrete probability distribution.

Solution: If a random variable x assumes n discrete values $x_1, x_2, ...,x_n$, with respective probabilities $p_1, p_2,..........,p_n (p_1 + p_2 + ...+ p_n = 1)$ then, the

NOTES

distribution of values x_i with probabilities p_i (= 1, 2,....,n), is called the discrete probability distribution of x.

The frequency function or frequency distribution of x is defined by $p(x)$ which for different values $x_1, x_2, ...x_n$ of x, gives the corresponding probabilities:

$$p(x_i) = p_i \text{ where, } p(x) \geq 0 \ \Sigma p(x) = 1$$

Example 6: For the following probability distribution, find $p(x > 4)$ and $p(x \geq 4)$:

x	0	1	2	3	4	5
$p(x)$	0	a	$a/2$	$a/2$	$a/4$	$a/4$

Solution: The solution is obtained as follows:

Since, $$\Sigma p(x) = 1, 0 + a + \frac{a}{2} + \frac{a}{2} + \frac{a}{4} + \frac{a}{4} = 1$$

$$\therefore \quad \frac{5}{2}a = 1 \quad \text{or} \quad a = \frac{2}{5}$$

$$p(x > 4) = p(x = 5) = \frac{9}{4} = \frac{1}{10}$$

$$p(x \leq 4) = 0 + a + \frac{a}{2} + \frac{a}{2} + \frac{a}{4} + \frac{9a}{4} = \frac{9}{10}$$

Example 7: A fair coin is tossed 400 times. Find the mean number of heads and the corresponding standard deviation.

Solution: This is a case of Binomial distribution with $p = q = \frac{1}{2}, n = 400$

The mean number of heads is given by $\mu = np = 400 \times \frac{1}{2} = 200$

And S. D. $\sigma = \sqrt{npq} = \sqrt{400 \times \frac{1}{2} \times \frac{1}{2}} = 10$

Example 8: A manager has thought of 4 planning strategies each of which has an equal chance of being successful. What is the probability that at least one of his strategies will work if he tries them in 4 situations? Here $p = \frac{1}{4}, q = \frac{3}{4}$.

Solution: The probability that none of the strategies will work is given by,

$$p(0) = {}^4C_0\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^4 = \left(\frac{3}{4}\right)^4$$

The probability that at least one will work is given by $1 - \left(\frac{3}{4}\right)^4 = \frac{175}{256}$

Example 9: Suppose the proportion of people preferring a car C is 0.5. Let X denote the number of people out of a set of 3 who prefer C. The probabilities of 0, 1, 2, 3 of them preferring C are,

Solution: The solution is obtained as follows:

$$p(X=0) = {}^3C_0(0.5)^0(0.5)^3 = \frac{1}{8}$$

$$p(X=1) = {}^3C_1(0.5)^1(0.5)^2 = \frac{3}{8}$$

$$p(X=2) = {}^3C_2(0.5)^2(0.5)^1 = \frac{3}{8}$$

$$p(X=3) = {}^3C_3(0.5)^3(0.5)^0 = \frac{1}{8}$$

$$\mu = E(X) = \Sigma x_i p_i = 0\times\frac{1}{8}+1\times\frac{3}{8}+2^2\frac{3}{8}+2\times\frac{3}{8}+3\times\frac{1}{8}=1.5$$

$$\sigma^2 = E(X-\mu)^2 = E(X^2)-\mu^2 = \Sigma x_i^2 p_i - \mu^2$$

$$= 0^2\times\frac{1}{8}+1^2\times\frac{3}{8}+2^2\times\frac{3}{8}+3^2\times\frac{1}{8}-1.5^2$$

$$= 0.75$$

Example 10: For the Poisson distribution, write the probabilities of 0, 1, 2, ... successes.

Solution: The solution is obtained as follows:

x	$p(x) = e^{-m}\dfrac{m^x}{x!}$
0	$p(0) = e^{-m}\, m^0/0!$
1	$p(1) = e^{-m}\dfrac{m}{1!} = p(0).m$
2	$e^{-m}\dfrac{m^2}{2!} = p(2) = p(1).\dfrac{m}{2}$
3	$e^{-m}\dfrac{m^3}{3!} = p(3) = p(2).\dfrac{m}{3}$
$\vdots$	

, and so on.

Total of all probabilities $\Sigma p(x) = 1$

NOTES

Example 11 What are the raw moments of Poisson distribution?

Solution: First raw moment $\mu'_1 = m$

Second raw moment $\mu'_2 = m^2 + m$

Third raw moment $\mu'_3 = m^3 + 3m^2 + m$

Example 12: For a Poisson distribution $p(0) = p(1)$, find $p(x > 0)$.

Solution: We have $e^{-m}\, m^0/0! = e^{-m}\, m^0/1!$ so that $m = 1$

$\therefore\ p(0) \quad = e^{-1} = 1/2.718 = 0.37$

$p(x > 0) \quad = 1 - p(0) = 0.63$

Example 13: It is claimed that the TV branded M is demanded by 70% of customers. If X is the number of TVs demanded, find the probability distribution of X when there are four customers.

Solution: If all the four demand M then $p(4) = 0.7^4 = 0.2401$. The probability distribution is:

X	4	3	2	1	0
p	0.2401	0.4116	0.2646	0.0756	0.0081

These values may be plotted.

Note: Poisson Approximation to the Binomial is given by when the number of trials n is large and the probability p of a success is small, the Binomial distribution can be approximated by the Poisson. This approximation is useful is practice.

If $p = 0.002$, $n = 1000$, $m = np = 2$

The probability of 2 successes in 1000 trials is,

$$p(X = 2) = {}^{n}C_x p^x q^{n-x} = {}^{100}C_2 (0.002)^2\ (0.998)^{998}$$

Similarly, $$p(X = 3) = {}^{100}C_3 (0.002)^3\ (0.998)^{997}, \text{ etc.}$$

These terms are difficult to calculate. If we employ the Poisson, we have a much easier task before us.

$$m = np = 100 \times 0.002 = 2$$

$$p(X = 2) = \frac{e^{-m} m^x}{x!} = \frac{e^{-2} 2^2}{2!} = 0.1353 \times 20.2706$$

$$p(X = 3) = \frac{e^{-m} m^x}{x!} = \frac{e^{-2} 2^3}{3!} = 0.1804$$

Example 14: One in every 100 items made on a machine is defective. Out of 25 items picked, find the probability of 1 item being defective.

Solution: $p = 0.01$, $q = 0.99$, $n = 25$, $np = 0.25$

$$\text{Binomial} : p(1) = {}^{25}C_1 (0.1)^1 (0.99)^{24} = 0.1964$$

$$\text{Poisson} : p(1) = \frac{e^{-25} (0.25)^1}{1!} = 0.1947$$

Exponential Distribution

In probability theory and statistics, the exponential distributions, also known as negative exponential distributions, are a set of continuous probability distributions. They describe the times between events in a Poisson process, i.e., a process in which events occur continuously and independently at a constant average rate.

Probability Density Function

The Probability Density Function (PDF) of an exponential distribution is defined as,

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

Here $\lambda > 0$ is the parameter of the distribution and is frequently termed as the *rate parameter*, λ. The distribution is based on the interval (0, •). When a random variable X has exponential distribution, then it is written as $X \sim \text{Exp}(\lambda)$.

Cumulative Distribution Function

The cumulative distribution function is defined as,

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

Alternative Parameterization

The Probability Density Function (PDF) of an exponential distribution can also be defined using alternative parameterization as,

$$f(x;\beta) = \begin{cases} \dfrac{1}{\beta} e^{-x/\beta}, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

Here $\beta > 0$ is a scale parameter of the distribution. It is the reciprocal of the rate parameter, λ. In this specific notation, β is considered as a *survival parameter* if a random variable X which is the duration of time that manages system to survive and $X \sim \text{Exponential}(\beta)$ then $E[X] = \beta$. Thus, the expected duration of survival of the system is β units of time. The parameterization involves the 'rate parameter' that arises in the context of events arriving at a rate λ, when the time between events has a mean of $\beta = \lambda^{-1}$.

Occurrence and Uses

The exponential distribution occurs automatically while the lengths of the inter-arrival times are described in a homogeneous Poisson process. It can be analysed

NOTES

as a continuous counterpart of the geometric distribution for describing the number of Bernoulli trials necessary for a discrete process to change state. Thus, the exponential distribution describes the time for a continuous process to change state.

Exponential variables can be used to model situations where specific events occur with a constant probability per unit length. In queuing theory, the service times of agents in a system are frequently modeled as exponentially distributed variables, for example the time taken by a bank teller to serve a customer. The length of a process can be considered as a sequence of several independent events and is modeled using a variable following the Erlang distribution, which is the distribution of the sum of several independent exponentially distributed variables.

The exponential distribution is also used in reliability theory and reliability engineering. Because this distribution has the *memoryless* property, hence it is quite compatible to model the constant hazard rate portion of the bathtub curve in reliability theory. In physics, when a gas is observed at a fixed temperature and pressure in a uniform gravitational field, then the altitudes of the various molecules also adhere to an approximate exponential distribution.

Properties of Exponential Distribution

Mean: The mean or expected value of an exponentially distributed random variable X with rate parameter λ is given as,

$$E[X] = \frac{1}{\lambda}.$$

Variance: The variance of X is given as,

$$\mathrm{Var}[X] = \frac{1}{\lambda^2}$$

Median: The median of X is given as,

$$m[X] = \frac{\ln 2}{\lambda} < E[X],$$

Here, ***ln*** refers to the natural logarithm. Hence, the absolute difference between the mean and median is given as shown below in accordance with the median-mean inequality.

$$|E[X] - m[X]| = \frac{1 - \ln 2}{\lambda} < \frac{1}{\lambda} = \text{Standard deviation}$$

Memorylessness: It is an important property of the exponential distribution. This explains that when a random variable T is exponentially distributed, then its conditional probability follows the following notation:

$$\Pr[T > s + t \mid T > s] = \Pr[T > t] \quad \text{for all } s, t \geq 0.$$

Let us use the above equation to explain memorylessness. For example, the conditional probability that we need to wait more than another 10 seconds before the first arrival specified that the first arrival has not still happened even after 30 seconds, then it is equal to the initial probability that we need to wait more than 10 seconds for the first arrival. Thus, if we have waited for 30 seconds and the first arrival did not happen ($T > 30$), then there is a probability that we need to wait for another 10 seconds for the first arrival ($T > 30 + 10$). This is similar to the initial probability that we need to wait more than 10 seconds for the first arrival ($T > 10$). This does not mean that the events $T > 40$ and $T > 30$ are independent events. The exponential distributions and the geometric distributions are the only memoryless probability distributions. The exponential distribution also has a constant hazard function.

Quartiles: The quartile function or inverse cumulative distribution function for Exponential (λ) is given as,

$$F^{-1}[p;\lambda] = \frac{-ln(1-p)}{\lambda},$$

For $0 \leq p \leq 1$. Therefore the quartiles are as follows:

First Quartile - ln (4/3)/λ

Median - ln (2)/λ

Third Quartile - ln (4)/λ

Kullback–Leibler Divergence: The directed Kullback–Leibler divergence between Exp(λ_0) for 'True' distribution and Exp(λ) for 'Approximating' distribution is given as,

$$\Delta[\lambda_0 \,||\, \lambda] = \log(\lambda_0) - \log(\lambda) + \frac{\lambda}{\lambda_0} - 1.$$

Maximum Entropy Distribution: The Exponential distribution with $\lambda = 1/\mu$ has the largest entropy with all continuous probability distributions having support (0, ∞) and mean μ.

Distribution of the Minimum of Exponential Random Variables

Let $X_1, ..., X_n$ be the independent exponentially distributed random variables with rate parameters $\lambda_1, ..., \lambda_n$. Then, min $\{X_1, ..., X_n\}$ is also exponentially distributed with parameter $l = \lambda_1 + ... + \lambda_n$

This can be defined using the complementary cumulative distribution function as shown below:

$$\Pr(\min\{X_1,\ldots,X_n\} > x) = \Pr[X_1 > x \text{ and} \ldots \text{and } X_n > x)$$
$$= \prod_{i=1}^{n} \Pr(X_i > x) = \prod_{i=1}^{n} \exp(-x\lambda_i) = \exp\left(-x\sum_{i=1}^{n}\lambda_i\right).$$

NOTES

The index of the variable which achieves the minimum is distributed according to the law,

$$\Pr(X_k = \min\{X_1,\ldots,X_n\}) = \frac{\lambda_k}{\lambda_i + \cdots + \lambda_n}.$$

Remember that min $\{X_1, ..., X_n\}$ is not exponentially distributed.

Check Your Progress

1. What are the Bernoulli trials?
2. Define the term hypergeometric distribution.
3. Give some characteristics features of the binomial distribution.
4. What is exponential distribution?

4.3 DISTRIBUTION OF TWO RANDOM VARIABLE

Mean of Random Variable or The Expected Value of Random Variable

Mean of *random variable* is the sum of the values of the random variable weighted by the probability that the random variable will take on the value. In other words, it is the sum of the product of the different values of the random variable and their respective probabilities. Symbolically, we write the mean of a random variable, say X, as $\overline{X}$. *The Expected value of the random variable* is the average value that would occur if we have to average an infinite number of outcomes of the random variable. In other words, it is the average value of the random variable in the long run. The expected value of a random variable is calculated by weighting each value of a random variable by its probability and summing over all values. The symbol for the expected value of a random variable is $E(X)$. Mathematically, we can write the mean and the expected value of a random variable, X, as follows:

$$\overline{X} = \sum_{i=1}^{n} (X_i).pr.(X_i)$$

And,

$$E(X) = \sum_{i=1}^{n} (X_i).pr.(X_i)$$

Thus, the mean and expected value of a random variable are conceptually and numerically the same but usually denoted by different symbols and as such the two symbols, viz., $\overline{X}$ and $E(X)$ are completely interchangeable. We can, therefore, express the two as follows:

$$E(X) = \sum_{i=1}^{n} (X_i).pr.(X_i) = \overline{X}$$

Where X_i is the ith value X can take.

NOTES

Sum of Random Variables

If we are given the means or the expected values of different random variables, say X, Y, and Z to obtain the mean of the random variable ($X + Y + Z$), then it can be obtained as under:

$$E(X + Y + Z) = E(X) + E(Y) + E(Z) = \overline{X} + \overline{Y} + \overline{Z}$$

Similarly, the expectation of a constant time a random variable is the constant time the expectation of the random variable. Symbolically, we can write this as under:

$$E(cX) = cE(X) = c\overline{X}$$

Where cX is the constant time random variable.

4.4 CONDITIONAL DISTRIBUTION AND EXPECTATION

Expectation (Conditional)

The expectation of a random variable X with probability density function (PDF) $p(x)$ is theoretically defined as:

$$E[X] = \int xp(x)dx$$

If we consider two random variables X and Y (not necessarily independent), then their combined behaviour is described by their joint probability density function $p(x, y)$ and is defined as:

$$p\{x \le X < x + dx,\ y \le Y < y + dy\} = p(x,\ y).dx.dy$$

The marginal probability density of X is defined as,

$$p_X(x) = \int p(x,\ y)dy$$

For any fixed value y of Y, the distribution of X is the conditional distribution of X, where $Y = y$, and it is denoted by $p(x\ ,\ y)$.

Expectation (Iterated)

The expectation of the random variable is expressed as:

$$E[X] = E[E[X \mid Y]]$$

NOTES

This expression is known as the '**Theorem of Iterated Expectation**' or '**Theorem of Double Expectation**'. Symbolically, it can be expressed as:

(*i*) For the discrete case:

$$E[X] = \Sigma_y E[X \mid Y = y].P\{Y = y\}$$

(*ii*) For the continuous case:

$$E[X] = \int_{-\infty}^{+\infty} E[X|Y = y].f(y).dy$$

Expectation: Continuous Variables

If x is a continuous random variable we define that,

$$E(x) = \int_{-\infty}^{\infty} x P(x) dx = u$$

The expectation of a function $h(x)$ is,

$$Eh(x) = \int_{-\infty}^{\infty} h(x) P(x) dx$$

The rth moment about the mean is,

$$E(x - \mu)^r = \int_{-\infty}^{\infty} (x - \mu)^r P(x) dx$$

Example 15: A newspaper earns Rs. 100 a day if there is suspense in the news. He loses Rs. 10 a day if it is an eventless newspaper. What is the expectation of his earnings if the probability of suspense news is 0.4?

Solution:

$$\begin{aligned} E(x) &= p_1x_1 + p_2x_2 \\ &= 0.4 \times 100 - 0.6 \times 10 \\ &= 40 - 6 = 34 \end{aligned}$$

Example 16: A player tossing three coins, earns Rs. 10 for 3 heads, Rs. 6 for 2 heads and Re. 1 for 1 head. He loses Rs. 25, if 3 tails appear. Find his expectation.

Solution:

$$p(HHH) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8} = p_1 \text{ say} \qquad \text{3 heads}$$

$$p(HHT) = {}^3C_2 \cdot \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{3}{8} = p_2 \text{ say} \qquad \text{2 heads, 1 tail}$$

$$p(HTT) = {}^3C_1 \cdot \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{3}{8} = p_3 \text{ say} \qquad \text{1 head, 2 tails}$$

NOTES

$$p(TTT) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{8} = p_4 \text{ say} \qquad \text{3 tails}$$

$$E(x) = p_1x_1 + p_2x_2 + p_3x_3 - p_4x_4$$

$$= \frac{1}{8} \times 10 + \frac{3}{8} \times 6 + \frac{3}{8} \times 2 - \frac{1}{8} \times 25$$

$$= \frac{9}{8} = \text{Rs } 1.125$$

Example 17: *A* and *B* roll a die. Whoever gets a 6, first wins Rs. 550. Find their individual expectations if *A* makes a start. What will be the answer if *B* makes a start?

Solution: *A* may not get 6 in 1st trial, $p_1 = \frac{1}{6}$

A may not get in 1st, *B* may not get in 2nd and *A* may get in 3rd,

$$p_3 = \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} = \left(\frac{5}{6}\right)^2 \frac{1}{6} \text{ and so on.}$$

$$A\text{'s winning chance} = \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} = \left(\frac{5}{6}\right)^2 \frac{1}{6} + \ldots\ldots\ldots\ldots$$

Where $p(x)$ is the density function x,

$$= \frac{1}{6} \frac{1}{1 - \left(\frac{5}{6}\right)^2} = \frac{6}{11}$$

$$\left(\text{Geometric progression with common ratio} \left(\frac{5}{6}\right)^2\right)$$

$$\therefore B\text{'s winning chance} = 1 - \frac{6}{11} = \frac{5}{11}$$

A wins Rs. 550 with probability $p = \frac{6}{11}$

A gets nothing if he loses with probability $q = \frac{5}{11}$

NOTES

Expectation of $A = p\,.\,x + q\,.\,0$

$$= \frac{6}{11} \times 550 + \frac{5}{11} \times 0 = 300$$

Similarly B wins with $p = \frac{5}{11}$

B gets nothing if he loses with $q = \frac{6}{11}$

Expectation of $B = \frac{5}{11} \times 550 + \frac{6}{11} \times 0 = 200$

If B starts first his expectation would be 300 and A's would be 250.

Example 18: Calculate the standard deviation (S.D.) when x takes the values 0, 1, 2 and 9 with probability, 0.4, 0 – 0.1.

Solution: x takes the values 0, 1, 2, 9 with probability 0.4, 0.2, 0.3, 0.1.

$$\mu = E(x) = \Sigma(x_i p_i) = 0 \times 0.4 + 1^2 \times 0.2 + 3 \times 0.3 + 9 \times 0.1 = 2.0$$

$$E(x^2) = \Sigma x_i^2 p_i = 0^2 \times 0.4 + 1^2 \times 0.2 + 3 \times 0.3 + 9^2 \times 0.1 = 11.0$$

$$V(x) = E(x^2) - \mu^2 = 11 - 2 = 9$$

$$\text{S.D.}(x) = \sqrt{9} = 3$$

Example 19: The purchase of some shares can give a profit of Rs. 400 with probability 1/100 and Rs. 300 with probability 1/20. Comment on a fair price of the share.

Solution: Expected value $E(x) = \Sigma x_i p_i = 400 \times \frac{1}{100} + 300 \times \frac{1}{20} = 19$

Example 20: Find the variance and standard deviation (S.D.) of the following probability distribution:

x_1	1	2	3	4
p_1	0.1	0.3	0.2	0.4

Solution: $E(x) = \Sigma p_i x_i = 0.1 \times 1 + 0.3 \times 2 + 0.2 \times 3 + 0.4 \times 4 = 2.9$

Variance $(x) = V(x) = E(x^2) - \bar{x}^2$

=

$$\Sigma p_i x_i^2 - \bar{x}^2 = 0.1 \times 1^2 + 0.3 \times 2^2 + 0.2 \times 3^2 + 0.4 \times 4^2 - 2.9^2$$

$$= 0.1 + 1.2 + 1.8 + 6.4 - 8.41 = 1.0.9$$

$$\text{S.D.}(x) = \sqrt{V(x)} = \sqrt{1.09}1.044$$

NOTES

Example 21: Prove that the variance of a constant is zero.

Solution: If k is a constant, it has no variability.

$$V(k) = 0$$

If the constant k is attached to a variable x then,

$$V(kx) = k^2 V(x)$$

$$V(2x) = 4V(x)$$

$$V(2 + 3x) = V(3x) = 9\ V(x)$$

Example 22: A box contains 2^n tickets of which ${}^nC_i$ tickets bear the number $i(i = 1, 2, 3, ..., n)$. If a set of m tickets is drawn, find the expected value of the sum of their numbers.

Solution: To find $E(A)$, where $A = x_1 + x_2 + ... + x_m$. Here x_i can take values 0, 1, 2,...., n with probabilities,

$$\frac{{}^nC_0}{2^n}, \frac{{}^nC_1}{2^n}, ... \frac{{}^nC_n}{2^n}$$

$$E(x_i) = \Sigma p_i x_i = \frac{{}^nC_0}{2^n}.0 + \frac{{}^nC_1}{2^n}.1 + \frac{{}^nC_2}{2^n}.2 + ... + \frac{{}^nC_n}{2^n}.\ n$$

$$= \frac{1}{2^n}[{}^nC_1 + 2\,{}^nC_2 + ... + n\,{}^nC_n]$$

$$= \frac{n}{2^n}[1 + {}^{n-1}C_1 + {}^{n-1}C_2 + ... + {}^{n-1}C_{n-1}] \text{ (By taking } n \text{ common)}$$

$$= \frac{n}{2^n}(1+1)^{n-1} = \frac{n}{2} \text{ [Expand } (1 + 1)^{n-1} \text{ and check]}$$

$$\therefore \quad E(A) = \sum_{i=1}^{m} E(x_i) = \frac{mn}{2}$$

Check Your Progress

5. Explain the mean of random variable.
6. What is conditional expectation?

4.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. In a Bernoulli experiment, an Event E either happens or does not happen (E'). Examples are, getting a head on tossing a coin, getting a six on rolling a die, and so on.

NOTES

2. The disribution of the random variable *X,* which is the number of successes obtained in the above case, is called the Hypergeometric distribution.

$$p(x) = \frac{{}^{N_1}C_x^{\,N}C_{n-x}}{{}^{N}C_n} \quad (X = 0, 1, 2, ..., n)$$

3. The following are the characteristics of Binomial distribution:
   (a) It is a discrete distribution.
   (b) It gives the probability of x successes and $n-x$ failures in a specific order.
   (c) The experiment consists of n repeated trials.
   (d) Each trial results in a success or a failure.
4. In probability theory and statistics, the exponential distributions, also known as negative exponential distributions, are a set of continuous probability distributions.
5. Mean of *random variable* is the sum of the values of the random variable weighted by the probability that the random variable will take on the value.
6. The expectation of a random variable *X* with probability density function (PDF) $p(x)$ is theoretically defined as:

   $E[X] = \int xp(x)dx$

   If we consider two random variables *X* and *Y* (not necessarily independent), then their combined behaviour is described by their joint probability density function $p(x, y)$ and is defined as:

   $p\{x \le X < x + dx,\ y \le Y < y + dy\} = p(x,\ y).dx.dy$

   The marginal probability density of *X* is defined as,

   $p_X(x) = \int p(x,\ y)dy$

   For any fixed value *y* of *Y*, the distribution of *X* is the conditional distribution of *X,* where $Y = y$, and it is denoted by $p(x\ , y)$.

4.6 SUMMARY

- Each possible value of the random variable *x* has the same probability in the uniform distribution. If *x* takes vaues $x_1, x_2....,x_k$, then,

  $$p(x_i,\ k) = \frac{1}{k}$$

- The multinomial distibution gives the probability that out of these *n* trials, x_1 occurs n_1 times, x_2 occurs n_2 times, and so on. This is given by the following equation: $\dfrac{n!}{n_1!n_2!....n_k!}\, p_1^{n_1} p_2^{n_2} ....p_k^{n}$

- The Binomial distribution depends on two parameters n and p. Each set of different values of n, p has a different Binomial distribution.
- The Binomial distribution can be approximated by the normal. As n becomes large and p is close to 0.5, the approximation becomes better.
- The Probability Density Function (PDF) of an exponential distribution can also be defined using alternative parameterization as,

$$f(x;\beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

- The Expected value of the random variable is the average value that would occur if we have to average an infinite number of outcomes of the random variable.

4.7 KEY WORDS

- **Cumulative distribution function:** The cumulative distribution function is defined as,

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

- **Quartiles:** The quartile function or inverse cumulative distribution function for Exponential (λ) is given as,

$$F^{-1}[p;\lambda] = \frac{-ln(1-p)}{\lambda},$$

- **Maximum entropy distribution:** The Exponential distribution with $\lambda = 1/\mu$ has the largest entropy with all continuous probability distributions having support $(0, \infty)$ and mean μ.

4.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Explain about the uniform or rectangular distribution.
2. What is geometric distribution?
3. What is meant by multinomial?
4. Explain the sum of random variables.
5. What is iterated expectation?

NOTES

Long-Answer Questions

1. Briefly describe the probability density function.
2. Discuss the properties of exponential distribution.
3. Describe the expected value of random variables.
4. Explain the continuous variables expectation.

4.9 FURTHER READINGS

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

NOTES

BLOCK - II
CORRELATION COEFFICIENT, SPECIAL DISTRIBUTIONS AND DISTRIBUTION FUNCTION OF RANDOM VARIABLES

UNIT 5 CORRELATION

Structure

5.0 INTRODUCTION

Correlation analysis looks at the indirect relationships in sample survey data and establishes the variables which are most closely associated with a given action or mind-set. It is the process of finding how accurately the line fits using the observations. Correlation analysis can be referred as the statistical tool used to describe the degree to which one variable is related to another. The relationship, if any, is usually assumed to be a linear one. In fact, the word correlation refers to the relationship or interdependence between two variables. There are various phenomena which have relation to each other. The theory by means of which quantitative connections between two sets of phenomena are determined is called the 'Theory of Correlation'. On the basis of the theory of correlation you can study the comparative changes occurring in two related phenomena and their cause-effect relation can also be examined. Thus, correlation is concerned with relationship between two related and quantifiable variables and can be positive or negative.

In this unit you will study about the correlation, types of correlation and properties of correlation coefficient, methods of studying correlation, scatter diagram, Karl Pearson's coefficient, rank coefficient and coefficient of determination.

5.1 OBJECTIVES

NOTES

After going through this unit, you will be able to:

- Define correlation
- Explain the different types of correlation
- Understand the function of correlation coefficient
- Understand the different methods of studying correlation

5.2 CORRELATION

Correlation analysis is the statistical tool generally used to describe the degree to which one variable is related to another. The relationship, if any, is usually assumed to be a linear one. This analysis is used quite frequently in conjunction with regression analysis to measure how well the regression line explains the variations of the dependent variable. In fact, the word correlation refers to the relationship or interdependence between two variables. There are various phenomena which have relation to each other. For instance, when demand of a certain commodity increases, then its price goes up and when its demand decreases then its price comes down. Similarly, with age the height of the children increases; with height the weight of the children increases, with money supply the general level of prices go up. Such sort of relationship can as well be noticed for several other phenomena. The theory by means of which quantitative connections between two sets of phenomena are determined is called the *Theory of Correlation.*

On the basis of the theory of correlation one can study the comparative changes occurring in two related phenomena and their cause-effect relation can be examined. It should, however, be borne in mind that relationship like 'black cat causes bad luck', 'filled-up pitchers result in good fortune' and similar other beliefs of the people cannot be explained by the theory of correlation since they are all imaginary and are incapable of being justified mathematically. Thus, correlation is concerned with the relationship between two related and quantifiable variables. If two quantities vary in sympathy so that a movement (an increase or decrease) in the one tends to be accompanied by a movement in the same or opposite direction in the other and the greater the change in the one, the greater is the change in the other, the quantities are said to be correlated. This type of relationship is known as correlation or what is sometimes called, in statistics, co-variation.

For correlation it is essential that the two phenomena, should have cause-effect relationship. If such relationship does not exist then one should not talk of correlation. For example, if the height of the students as well as the height of the trees increases, then one should not call it a case of correlation because the two phenomena, viz., the height of students and the height of trees are not causally related. But the relationship between the price of a commodity and its demand, the price of a commodity and its supply, the rate of interest and savings, etc., are examples of correlation since in all such cases the change in one phenomenon is explained by a change in the other phenomenon.

Check Your Progress

1. What are the different types of correlations?
2. Explain the meaning of correlation analysis.
3. What is the scatter diagram method?
4. What is the least-squares method?

5.3 TYPES OF CORRELATION

It is appropriate here to mention that correlation in case of phenomena pertaining to natural sciences can be reduced to absolute mathematical terms, for example, heat always increases with light. But in phenomena pertaining to social sciences, it is often difficult to establish any absolute relationship between two phenomena. Hence, in social sciences we must take the fact of correlation being established if in a large number of cases, two variables always tend to move in the same or the opposite direction.

Correlation can either be positive or it can be negative. Whether correlation is positive or negative would depend upon the direction in which the variables are moving. If both variables are changing in the same direction, then correlation is said to be positive but when the variations in the two variables take place in opposite direction, the correlation is termed as negative. This can be explained as follows:

Changes in Independent Variable	*Changes in Dependent Variable*	*Nature of Correlation*
Increase (+)↑	Increase (+)↑	Positive (+)
Decrease (–)↓	Decrease (–)↓	Positive (+)
Increase (+)↑	Decrease (–)↓	Negative (–)
Decrease (–)↓	Increase (+)↑	Negative (–)

Correlation can either be linear or it can be non-linear. Non-linear correlation is also known as curvilinear correlation. The distinction is based upon the constancy of the ratio of change between the variables. When the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable then the correlation is said to be linear. In such a case, if the values of the variables are plotted on a graph paper, then a straight line is obtained. This is why the correlation is known as linear correlation. But when the amount of change in one variable does not bear a constant ratio to the amount of change in the other variable, i.e., the ratio happens to be variable instead of constant, then the correlation is said to be non-linear or curvilinear. In such a situation we shall obtain a curve if the values of the variables are plotted on a graph paper.

NOTES

Correlation can either be simple correlation or it can be partial correlation or it can be multiple correlation. The study of correlation for two variables (of which one is independent and the other is dependent) involves application of simple correlation. When more than two variables are involved in a study relating to correlation then it can either be as of multiple correlation or of partial correlation. Multiple correlation studies the relationship between a dependent variable and two or more independent variables. In partial correlation, we measure the correlation between a dependent variable and one particular independent variable assuming that all other independent variables remain constant.

Statisticians have developed two measures for describing the correlation between two variables viz., the coefficient of determination and the coefficient of correlation.

5.4 PROPERTIES OF CORRELATION COEFFICIENT

The coefficient of correlation symbolically denoted by 'r' is an important measure to describe how well one variable is explained by another. It measures the degree of relationship between the two causally-related variables. The value of this coefficient can never be more than + 1 or less than –1. Thus, + 1 and –1 are the limits of this coefficient. For a unit change in independent variable, if there happens to be a constant change in the dependent variable in the same direction then the value of the coefficient will be + 1 indicative of the perfect positive correlation; but if such a change occurs in the opposite direction, the value of the coefficient will be –1, indicating perfect negative correlation. In practical life the possibility of obtaining either a perfect positive or perfect negative correlation is very remote, particularly in respect of phenomena concerning social sciences. If the coefficient of correlation has a zero value then it means that there exists no correlation between the variables under study.

There are several methods of finding the coefficient of correlation but the following ones are considered important:[1]

(*i*) Coefficient of correlation by the method of least squares

(*ii*) Coefficient of correlation through product moment method or Karl Pearson's coefficient of correlation

(*iii*) Coefficient of correlation using simple regression coefficients

Whichever of these three methods, we adopt we get the same value of r. Now, we explain in brief each one of these three methods of finding *'r'*.

5.5 METHODS OF STUDYING CORRELATION

5.5.1 Scatter Diagram

Least squares method of fitting a line (the line of best fit or the regression line) through the scatter diagram is a method which minimizes the sum of the squared vertical deviations from the fitted line. In other words, the line to be fitted will pass through the points of the scatter diagram in such a fashion that the sum of the squares of the vertical deviations of these points from the line will be a minimum.

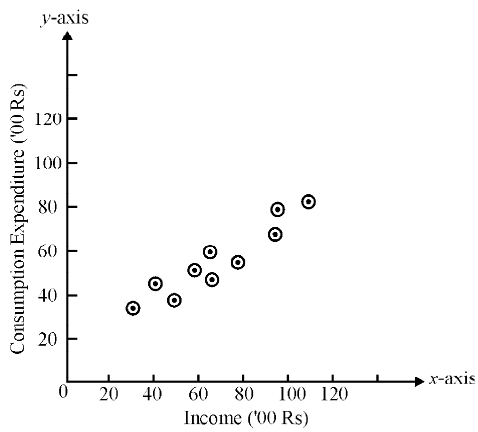


Fig 5.1 *Scatter Diagram*

The meaning of the least squares criterion can be better understood more easily through reference to the following Figure 5.2 where the Scatter diagram has been reproduced along with a line which represents the least squares fit to the data.

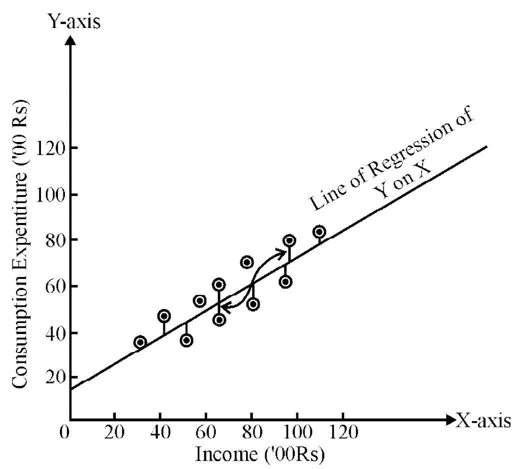


Fig 5.2 *Scatter Diagram, Regression Line and Short Vertical Lines Representing 'e'*

In Figure 5.2, the vertical deviations of the individual points from the line are shown as the short vertical lines joining the points to the least squares line. These deviations are denoted by the symbol 'e'. The value 'e' varies from one point to another. In some cases it is positive, in others it is negative. If the line drawn happens to be the least squares line then the values of $\sum e_i^o$ is the least possible. It is because of this feature the method is known as Least Squares Method.

Why we insist on minimizing the sum of squared deviations is a question that needs explanation. If we denote the deviations from the actual value Y to the estimated value as $(Y - \hat{Y})$ or e_i, it is logical that we want the $\Sigma(Y - \hat{Y})$ or $\sum_{i=1}^{n} e,$ to

NOTES

be as small as possible. However, mere examining $\Sigma(Y - \hat{Y})$ or $\sum_{i=1}^{n} e,$ is inappropriate since any e_i can be positive or negative and large positive values and large negative values would cancel one another.

But large values of e_i regardless of their sign, indicate a poor prediction. Even if we ignore the signs while working out $\sum_{i=1}^{n} | e_i |$, the difficulties may continue to be there. Hence, the standard procedure is to eliminate the effect of signs by squaring each observation. Squaring each term accomplishes two purposes viz., (*i*) It magnifies (or penalizes) the larger errors, and (*ii*) It cancels the effect of the positive and negative values (since a negative error squared becomes positive). The choice of minimizing the squared sum of errors rather than the sum of the absolute values implies that we would make many small errors rather than a few large errors. Hence, in obtaining the regression line we follow the approach that the sum of the squared deviations be minimum and on this basis work out the values of its constants viz., '*a*' and '*b*' or what is known as the intercept and the slope of the line. This is done with the help of the following two normal equations:[3]

$$\Sigma Y = na + b\Sigma X$$

$$\Sigma XY = a\Sigma X + \mathrm{b}\Sigma X^2$$

In these two equations, '*a*' and '*b*' are unknowns and all other values viz., ΣX, ΣY, ΣX^2, ΣXY are the sum of the products and the cross products to be calculated from the sample data and '*n*' means the number of observations in the sample. Hence, one can solve these two equations for finding the unknown values. Once these values are found, the regression line is said to have been defined for the given problem. Statisticians have also derived a short cut method through which these two equations can be rewritten so that the values of '*a*' and '*b*' can be directly obtained as follows:

$$b = \frac{n\sum XY - \sum X.\sum Y}{n\sum X^2 - \left(\sum X\right)^2}$$

$$a = \frac{\sum Y}{n} - b\frac{\Sigma X}{n}$$

5.5.2 Karl Pearson's Coefficient

Karl Pearson's method is the most widely used method for measuring the relationship between two variables. This coefficient is based on the following assumptions:

(*i*) There is a linear relationship between the two variables which means that a straight line would be obtained if the observed data is plotted on a graph.

(*ii*) The two variables are causally related which means that one of the variables is independent and the other one is dependent.

(*iii*) A large number of independent causes operates in both the variables so as to produce a normal distribution.

NOTES

According to Karl Pearson, 'r' can be worked out as follows:

$$r = \frac{\sum xy}{n\sigma_x \sigma_y}$$

Here,

$$x = (x - x)$$

$$y = (y - y)$$

σ_x = Standard deviation of X series and is equal to $\sqrt{\frac{\sum x^2}{n}}$

σ_y = Standard deviation of Y series and is equal to $\sqrt{\frac{\sum y^2}{n}}$

n = Number of pairs of X and Y observed

A short cut formula known as the Product Moment Formula can be derived from the earlier formula:

$$r = \frac{\sum xy}{n\sigma_x \sigma_y}$$

$$= \frac{\sum xy}{\sqrt{\frac{\sum x^2}{n} \cdot \frac{\sum y^2}{n}}}$$

$$= \frac{n \sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

These formulae are based on obtaining the true means (viz., $\bar{x}$ and $\bar{y}$) first and then performing all other calculations.

5.5.3 Rank Coefficient

If observations on two variables are given in the form of ranks and not as numerical values, it is possible to compute what is known as rank correlation between the two series.

The rank correlation, written ρ, is a descriptive index of agreement between ranks over individuals. It is the same as the ordinary coefficient of correlation computed on ranks, but its formula is simpler.

$$\rho = 1 - \frac{6\Sigma D_i^2}{n(n^2-1)}$$

NOTES

Here, n is the number of observations and D_i the positive difference between ranks associated with the individuals i.

Like r, the rank correlation lies between –1 and +1.

Example 1: The ranks given by two judges to 10 individuals are as follows:

	Rank given by			
Individual	*Judge I*	*Judge II*	*D*	D^2
	x	y	$= x - y$	
1	1	7	6	36
2	2	5	3	9
3	7	8	1	1
4	9	10	1	1
5	8	9	1	1
6	6	4	2	4
7	4	1	3	9
8	3	6	3	9
9	10	3	7	49
10	5	2	3	9
				$\Sigma D^2 = 128$

Solution: The rank correlation is given by,

$$\rho = 1 - \frac{6\Sigma D^2}{n^3 - n} = 1 - \frac{6 \times 128}{10^3 - 10} = 1 - 0.776 = 0.224$$

The value of $\rho = 0.224$ shows that the agreement between the judges is not high.

Example 2: In the previous case, compute r and compare.

Solution: The simple coefficient of correlation r for the previous data is calculated as follows:

x	y	x^2	y^2	xy
1	7	1	49	7
2	5	4	25	10
7	8	49	64	56
9	10	81	100	90
8	9	64	81	72
6	4	36	16	24
4	1	16	1	4
3	6	9	36	18
10	3	100	9	30
5	2	25	4	10
$\Sigma x = 55$	$\Sigma y = 55$	$\Sigma x^2 = 385$	$\Sigma y^2 = 385$	$\Sigma xy = 321$

$$r = \frac{321 - 10\times\frac{55}{10}\times\frac{55}{10}}{\sqrt{385-10\times\left(\frac{55}{10}\right)^2}\sqrt{385-10\times\left(\frac{55}{10}\right)^2}} = \frac{18.5}{\sqrt{82.5\times 82.5}} = \frac{18.5}{82.5}$$

$$= 0.224$$

NOTES

This shows that the Spearman ρ for any two sets of ranks is the same as the Pearson *r* for the set of ranks. But it is much easier to compute ρ.

Often, the ranks are not given. Instead, the numerical values of observations are given. In such a case, we must attach the ranks to these values to calculate ρ.

Example 3:

Marks in Maths	*Marks in Stats*	*Rank in Maths*	*Rank in Stats*	*D*	D^2
45	60	4	2	2	4
47	61	3	1	2	4
60	58	1	3	2	4
38	48	5	4	1	1
50	46	2	5	3	9
				ΣD^2 =	22

$$\rho = 1 - \frac{6\Sigma D^2}{n^3 - n} = 1 - \frac{6\times 22}{125-5} = -0.1$$

Solution: This shows a negative, though small, correlation between the ranks.

If two or more observations have the same value, their ranks are equal and obtained by calculating the means of the various ranks.

If in this data, marks in maths, are 45 for each of the first two students, the rank of each would be $\frac{3+4}{2}$ = 3.5. Similarly, if the marks of each of the last two students in statistics are 48, their ranks would be $\frac{4+5}{2}$ = 4.5

The problem takes the following shape:

			Rank		
	Marks in	*Marks in*	*x*	*y*	*D*
D^2		*Maths*	*Stats*		
45	60	3.5	2	1.5	2.25
45	61	3.5	1	2.5	6.25
60	58	1	3	2	4.00
38	48	5	4.5	1.5	2.25
50	48	2	4.5	2.5	6.25

$$\rho = 1 - \frac{6\Sigma D^2}{n^3 - n} = 1 - \frac{6 \times 21}{120} = -0.05$$

An elaborate formula which can be used in case of equal ranks is

NOTES

$$\rho = 1 - \frac{6}{n^3 - n}\left[\Sigma D^2 + \frac{1}{12}\Sigma(m^3 - m)\right].$$

Here, $\frac{1}{12}\Sigma(m^3 - m)$ is to be added to ΣD^2 for each group of equal ranks, m being the number of equal ranks each time.

For the given data, we have

For series x, the number of equal ranks $m = 2$.

For series y, also, $m = 2$; so that,

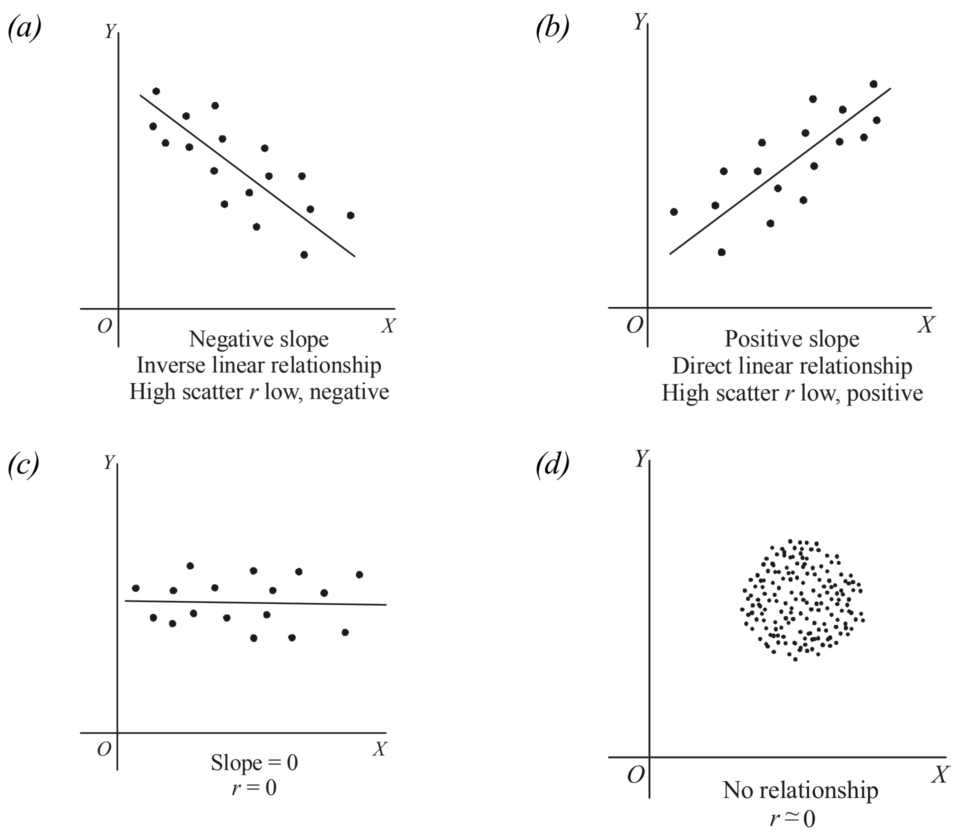
$$\rho = 1 - \frac{6}{5^3 - 5}\left[21 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]$$

$$= 1 - \frac{6}{120}\left[21 + \frac{6}{12} + \frac{6}{12}\right]$$

$$= 1 - \frac{6 \times 22}{120} = -0.1$$

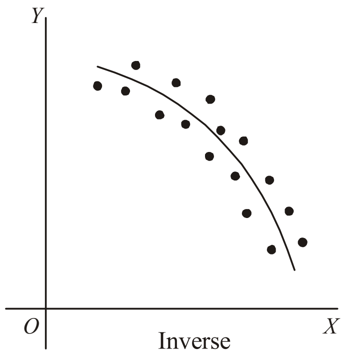
Example 4: Show by means of diagrams various cases of scatter expressing correlation between x, y.

Solution:



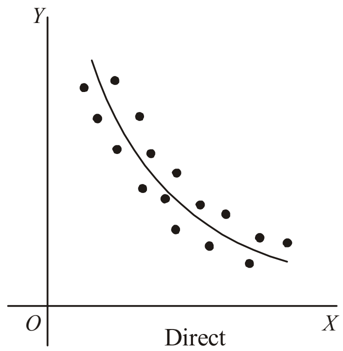
NOTES

(e)



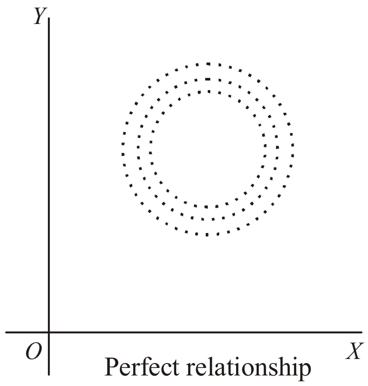
Inverse
curvilinear relationship

(f)



Direct
curvilinear relationship

(g)



Perfect relationship
But, $r = 0$ because of
non-linear relation

Correlation analysis helps us in determining the degree to which two or more variables are related to each other.

When there are only two variables we can determine the degree to which one variable is linearly related to the other. Regression analysis helps in determining the pattern of relationship between one or more independent variables and a dependent variable. This is done by an equation estimated with the help of data.

5.6 COEFFICIENT OF DETERMINATION

Coefficient of determination (r^2) which is the square of the coefficient of correlation (r) is a more precise measure of the strength of the relationship between the two variables and lends itself to more precise interpretation because it can be presented as a proportion or as a percentage.

The coefficient of determination (r^2) can be defined as the proportion of the variation in the dependent variable Y, that is explained by the variation in independent variable X, in the regression model. In other words:

$$r^2 = \frac{\text{Explained variation}}{\text{Total variation}}$$

$$= \frac{\Sigma(Y_c - \overline{Y})^2}{\Sigma(Y - \overline{Y})^2}$$

$$= \frac{b_0 \Sigma Y + b_1 \Sigma XY - \frac{(\Sigma Y)^2}{n}}{\Sigma(Y)^2 - \frac{(\Sigma Y)^2}{n}}$$

NOTES

Example 5: The heights of fathers and their sons are given. Calculate the coefficient of correlation r and the coefficient of determination (r^2). Also, given that

$b_0 = 26.25$,

$b_1 = 0.625$

Father (X)	Son (Y)
63	66
65	68
66	65
67	67
67	69
68	70

Solution: Now,

$$r^2 = \frac{b_0 \Sigma Y + b_1 \Sigma XY - \frac{(\Sigma Y)^2}{n}}{\Sigma(Y)^2 - \frac{(\Sigma Y)^2}{n}}$$

X	Y	X^2	XY	Y^2
63	66	3969	4158	4356
65	68	4225	4420	4624
66	65	4356	4290	4225
67	67	4489	4489	4489
67	69	4489	4623	4761
68	70	4624	4760	4900
$\Sigma X = 396$	$\Sigma Y = 405$	$\Sigma X^2 = 26152$	$\Sigma XY = 26740$	$\Sigma Y^2 = 27355$

Hence,

$$r^2 = \frac{26.25(405) + 0.625(26740) - \frac{(405)^2}{6}}{27355 - \frac{(405)^2}{6}}$$

$$= \frac{10631.25 + 16712.5 - 27337.5}{27355 - 27377.5}$$

$$= \frac{6.25}{17.5} = 0.357$$

and $$r = \sqrt{r^2} = \sqrt{0.357} = 0.597$$

While the value of $r = 0.597$ is more of an abstract figure, the value of $r^2 = 0.357$ tells us that 35.7% of the variation in Y is explained by the variation in X. This indicates a weak relationship since the value of $r^2 = 0$, means no relationship at all and the value of $r = 1$ or 100% means perfect relationship. In general, for a high degree of correlation which leads to better estimates and prediction, the coefficient of determination r^2 must have a high value.

Check Your Progress

5. What do you mean by coefficient of correlation?
6. State the assumptions for finding the coefficient of correlation by Karl Pearson's method.
7. Define coefficient of determination, r^2.

5.7 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. There are several types of correlations. They are:
   (a) Positive or negative correlations
   (b) Linear or non-linear correlations
   (c) Simple, partial or multiple correlations
2. Correlation analysis is the statistical tool that is generally used to describe the degree to which one variable is related to another. The relationship, if any, is usually assumed to be a linear one. This analysis is used quite frequently in conjunction with regression analysis to measure how well the regression line explains the variations of the dependent variable. In fact, the word correlation refers to the relationship or interdependence between two variables. There are various phenomena, which are related to each other. For instance, when demand of a certain commodity increases, then its price goes up and when its demand decreases then its price comes down.
3. Scatter diagram is the method to calculate the constants in regression models that makes use of scatter diagram or dot diagram. A scatter diagram is a diagram that represents two series with the known variables, i.e., independent variable plotted on the X-axis and the variable to be estimated, i.e., dependent variable to be plotted on the Y-axis.
4. The least squares method is a method to calculate the constants in regression models for fitting a line through the scatter diagram that minimizes the sum of the squared vertical deviations from the fitted line. In other words, the line to be fitted will pass through the points of the scatter diagram in such a fashion that the sum of the squares of the vertical deviations of these points from the line will be a minimum.

NOTES

5. The coefficient of correlation, which is symbolically denoted by r, is another important measure to describe how well one variable explains another. It measures the degree of relationship between two casually related variables. The value of this coefficient can never be more than +1 or –1. Thus, +1 and –1 are the limits of this coefficient.
6. Karl Pearson's method is most widely used method of measuring the relationship between two variables. This coefficient is based on the following assumptions:
   (*i*) There is a linear relationship between the two variables which means that a straight line would be obtained if the observed data is plotted on a graph.
   (*ii*) The two variables are casually related which means that one of the variables is independent and the other is dependent.
   (*iii*) A large number of independent causes are operating in both the variables so as to produce a normal distribution.
7. The coefficient of determination (r^2), the square of the coefficient of correlation (r), is a more precise measure of the strength of the relationship between the two variables and lends itself to more precise interpretation because it can be presented as a proportion or as a percentage.

5.8 SUMMARY

- Correlation analysis is the statistical tool generally used to describe the degree to which one variable is related to another. The relationship, if any, is usually assumed to be a linear one. This analysis is used quite frequently in conjunction with regression analysis to measure how well the regression line explains the variations of the dependent variable.
- Correlation can either be positive or it can be negative. Whether correlation is positive or negative would depend upon the direction in which the variables are moving.
- Non-linear correlation is also known as curvilinear correlation. The distinction is based upon the constancy of the ratio of change between the variables.
- Least squares method of fitting a line (the line of best fit or the regression line) through the scatter diagram is a method which minimizes the sum of the squared vertical deviations from the fitted line.
- There is a linear relationship between the two variables which means that a straight line would be obtained if the observed data is plotted on a graph.
- The two variables are causally related which means that one of the variables is independent and the other one is dependent.
- If observations on two variables are given in the form of ranks and not as numerical values, it is possible to compute what is known as rank correlation between the two series.

- Coefficient of determination (r^2) which is the square of the coefficient of correlation (r) is a more precise measure of the strength of the relationship between the two variables and lends itself to more precise interpretation because it can be presented as a proportion or as a percentage.

NOTES

5.9 KEY WORDS

- **Correlation analysis:** It is the statistical tool to describe the degree to which one variable is related to another.
- **Scatter diagram:** It is a graph of observed plotted points where each point represents the values of X and Y as a coordinate.
- **Coefficient of determination:** It can be defined as the proportion of the variation in the dependent variable Y that is explained by the variation in independent variable X.
- **Rank correlation:** In this type of correlation, observations on two variables are given in the form of ranks instead of numerical values. Rank correlation is a descriptive index of agreement between ranks over individuals.

5.10 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is a scatter diagram?
2. Under what conditions will you say that a correlation is linear?
3. How does a scatter diagram help in studying the correlation between two variables?
4. List the different types of correlation.
5. Define correlation analysis.

Long-Answer Questions

1. Obtain the estimating equation by the method of least squares from the following information:

X (*Independent variable*)	Y (*Dependent variable*)
2	18
4	12
5	10
6	8
8	7
11	5

NOTES

2. Find out the coefficient of correlation between the two kinds of assessment of M.A. students' performance.
   (*i*) By adopting Karl Pearson's method
   (*ii*) By the method of least squares

S.N. of Students	*Internal assessment (Marks obtained out of 100)*	*External assessment (Marks obtained out of 100)*
1	51	49
2	63	72
3	73	74
4	46	44
5	50	58
6	60	66
7	47	50
8	36	30
9	60	35

Also, work out r_2 and interpret the same.

3. Calculate correlation coefficient from the following results:

$$n = 10; \sum X = 140; \sum Y = 150$$
$$\sum(X-10)^2 = 180; \sum(Y-15)^2 = 215$$
$$(X-10)(Y-15) = 60$$

4. Given is the following information:

Observation	*Test score* X	*Sales* (,000 Rs) Y
1	73	450
2	78	490
3	92	570
4	61	380
5	87	540
6	81	500
7	77	480
8	70	430
9	65	410
10	82	490
Total	766	4740

You are required to:
   (*i*) Graph the scatter diagram for the given data.
   (*ii*) Find the regression equation and draw the line corresponding to the equation on the scatter diagram.
   (*iii*) Make an estimate of sales if the test score happens to be 75.

5. Calculate correlation coefficient and the two regression lines for the following information:

		Ages of Wives (in years)				
		10–20	*20–30*	*30–40*	*40–50*	*Total*
Ages of	10–20	20	26	—	—	46
Husbands	20–30	8	14	37	—	59
(in	30–40	—	4	18	3	25
years)	40–50	—	—	4	6	10
Total		28	44	59	9	140

6. To know what relationship exists between unemployment and suicide attempts, a sociologist surveyed twelve cities and obtained the following data:

S.N. of the city	*Unemployment rate per cent*	*Number of suicide attempts per 1000 residents*
1	7.3	22
2	6.4	17
3	6.2	9
4	5.5	8
5	6.4	12
6	4.7	5
7	5.8	7
8	7.9	19
9	6.7	13
10	9.6	29
11	10.3	33
12	7.2	18

(*i*) Develop the estimating equation that best describes the given relationship.

(*ii*) Find a prediction interval (with 95% confidence level) for the attempted suicide rate when unemployment rate happens to be 6%.

5.11 FURTHER READINGS

NOTES

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 6 DISTRIBUTIONS

NOTES

Structure

6.0 INTRODUCTION

Binomial distribution is used in finite sampling problems where each observation is one of two possible outcomes ('success' or 'failure'). Poisson distribution is used for modelling rates of occurrence. Exponential distribution is used to describe units that have a constant failure rate. The term 'normal distribution' refers to a particular way in which observations will tend to pile up around a particular value rather than be spread evenly across a range of values, i.e., the Central Limit Theorem (CLT). It is generally most applicable to continuous data and is intrinsically associated with parametric statistics (For example, ANOVA, *t*-test, regression analysis). Graphically, the normal distribution is best described by a bell-shaped curve. This curve is described in terms of the point at which its height is maximum, i.e., its mean and its width or standard deviation.

In this unit, you will study about the binomial distribution, Poisson distribution, limiting form of binomial and Poisson fitting and discrete theoretical distributions.

6.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Understand about the binomial distribution
- Analyse the Poisson distribution
- Discuss the limiting form of binomial and Poisson fitting
- Explain the discrete theoretical distributions

6.2 BINOMIAL DISTRIBUTION

NOTES

Binomial distribution (or the Binomial probability distribution) is a widely used probability distribution concerned with a discrete random variable and as such is an example of a discrete probability distribution. The binomial distribution describes discrete data resulting from what is often called as the Bernoulli process. The tossing of a fair coin a fixed number of times is a Bernoulli process and the outcome of such tosses can be represented by the binomial distribution. The name of Swiss mathematician Jacob Bernoulli is associated with this distribution. This distribution applies in situations where there are repeated trials of any experiment for which only one of two mutually exclusive outcomes (often denoted as 'success' and 'failure') can result on each trial.

Bernoulli Process

Binomial distribution is considered appropriate in a Bernoulli process which has the following characteristics:

(a) **Dichotomy.** This means that each trial has only two mutually exclusive possible outcomes, for example, 'Success' or 'failure', 'Yes' or 'No', 'Heads' or 'Tails' and the like.

(b) **Stability.** This means that the probability of the outcome of any trial is known (or given) and remains *fixed* over time, i.e., remains the same for all the trials.

(c) **Independence.** This means that the trials are statistically independent, i.e., to say the happening of an outcome or the event in any particular trial is independent of its happening in any other trial or trials.

Probability Function of Binomial Distribution

The random variable, say X, in the Binomial distribution is the number of 'successes' in n trials. The probability function of the binomial distribution is written as under:

$$f(X = r) = {}^nC_r p^r q^{n-r}$$

$$r = 0,\ 1,\ 2 \ldots n$$

Where, n = Numbers of trials.

p = Probability of success in a single trial.

$q = (1 - p)$ = Probability of 'failure' in a single trial.

r = Number of successes in 'n' trials.

Parameters of Binomial Distribution

This distribution depends upon the values of p and n which in fact are its parameters. Knowledge of p truly defines the probability of X since n is known by definition of the problem. The probability of the happening of exactly r events in n trials can be found out using the above stated binomial function.

NOTES

The value of p also determines the general appearance of the binomial distribution, if shown graphically. In this context the usual generalizations are:

(a) When p is small (say 0.1), the binomial distribution is skewed to the right, i.e., the graph takes the form shown in Figure 6.1.

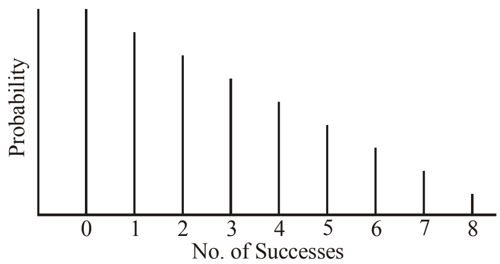


Fig. 6.1

(b) When p is equal to 0.5, the binomial distribution is symmetrical and the graph takes the form as shown in Figure 6.2.

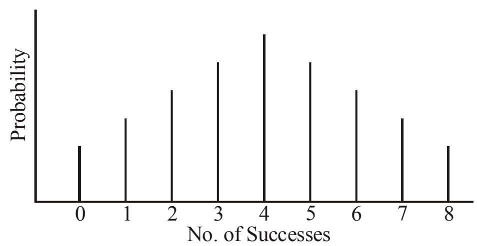


Fig. 6.2

(c) When p is larger than 0.5, the binomial distribution is skewed to the left and the graph takes the form as shown in Figure 6.3.

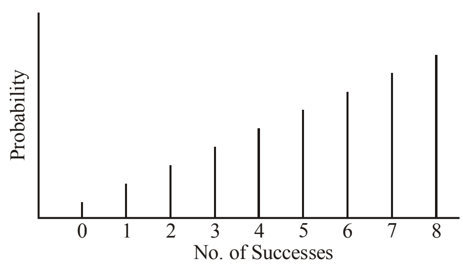


Fig. 6.3

But if 'p' stays constant and 'n' increases, then as 'n' increases the vertical lines become not only numerous but also tend to bunch up together to form a bell shape, i.e., the binomial distribution tends to become symmetrical and the graph takes the shape as shown in Figure 6.4.

NOTES

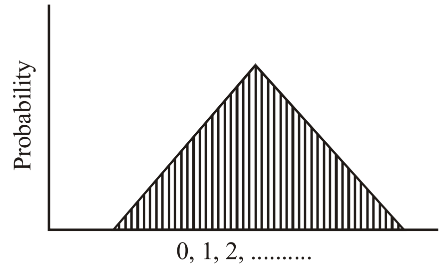


Fig. 6.4

Important Measures of Binomial Distribution

The expected value of random variable [i.e., $E(X)$] or mean of random variable (i.e., $\bar{X}$) of the binomial distribution is equal to $n.p$ and the variance of random variable is equal to $n.\ p.\ q$ or $n.\ p.\ (1 - p)$. Accordingly the standard deviation of binomial distribution is equal to $\sqrt{n.p.q}$. The other important measures relating to binomial distribution are as under:

$$\text{Skewness} = \frac{1-2p}{\sqrt{n.p.q}}$$

$$\text{Kurtosis} = 3 + \frac{1-6p+6q^2}{n.p.q}$$

When to Use Binomial Distribution

The use of binomial distribution is most appropriate in situations fulfilling the conditions outlined above. Two such situations, for example, can be described as follows.

(a) When we have to find the probability of 6 heads in 10 throws of a fair coin.

(b) When we have to find the probability that 3 out of 10 items produced by a machine, which produces 8 per cent defective items on an average, will be defective.

Example 1: A fair coin is thrown 10 times. The random variable X is the number of head(s) coming upwards. Using the binomial probability function, find the probabilities of all possible values which X can take and then verify that binomial distribution has a mean: $\overline{X} = n.p.$ and variance: $\sigma^2 = n.p.q$

Solution: Since the coin is fair and so, when thrown, can come either with head upward or tail upward. Hence, p (head) $= \frac{1}{2}$ and q (no head) $= \frac{1}{2}$. The required probability function is,

$$f(X=r) = {}^nC_r\, p^r q^{n-r}$$
$$r = 0,\ 1,\ 2 \ldots 10$$

The following table of binomial probability distribution is constructed using this function.

X_i (Number of Heads)	Probability pr_i			$X_i pr_i$	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^2.p_i$
0	${}^{10}C_0 p^0 q^{10}$	=	1/1024	0/1024	–5	25	25/1024
1	${}^{10}C_1 p^1 q^9$	=	10/1024	10/1024	–4	16	160/1024
2	${}^{10}C_2 p^2 q^8$	=	45/1024	90/1024	–3	9	405/1024
3	${}^{10}C_3 p^3 q^7$	=	120/1024	360/1024	–2	4	480/1024
4	${}^{10}C_4 p^4 q^6$	=	210/1024	840/1024	–1	1	210/1024
5	${}^{10}C_5 p^5 q^5$	=	252/1024	1260/1024	0	0	0/1024
6	${}^{10}C_6 p^6 q^4$	=	210/1024	1260/1024	1	1	210/1024
7	${}^{10}C_7 p^7 q^3$	=	120/1024	840/1024	2	4	480/1024
8	${}^{10}C_8 p^8 q^2$	=	45/1024	360/1024	3	9	405/1024
9	${}^{10}C_9 p^9 q^1$	=	10/1024	90/1024	4	16	160/1024
10	${}^{10}C_{10} p^{10} q^0$	=	1/1024	10/1024	5	25	25/1024
				$\Sigma\bar{X}$ = 5120/1024 $\bar{X}$ = 5			Variance = σ^2 = $\Sigma\ (X_i - \bar{X})^2.pr_i$ = 2560/1024 = 2.5

NOTES

The mean of the binomial distribution[1] is given by $n.\ p. = 10 \times \frac{1}{2} = 5$ and the variance of this distribution is equal to $n.\ p.\ q. = 10 \times \frac{1}{2} \times \frac{1}{2} = 2.5$

These values are exactly the same as we have found them in the above table.

Hence, these values stand verified with the calculated values of the two measures as shown in the table.

Check Your Progress

1. Explain the Bernoulli process.
2. Write the probability function of binomial distribution.
3. Explain the different parameters of binomial distributions.
4. Explain the important measures of binomial distribution.
5. Under what circumstances will you use binomial distribution?

6.3 THE POISSON DISTRIBUTION

Poisson distribution is also a discrete probability distribution with which is associated the name of a Frenchman, Simeon Denis Poisson who developed this distribution. This distribution is frequently used in context of Operations Research and for this reason has a great significance for management people. This distribution plays an important role in *Queuing* theory, Inventory control problems and also in Risk models.

NOTES

Unlike binomial distribution, Poisson distribution cannot be deducted on purely theoretical grounds based on the conditions of the experiment. In fact, it must be based on experience, i.e., on the empirical results of past experiments relating to the problem under study. Poisson distribution is appropriate specially when probability of happening of an event is very small (so that q or $(1-p)$ is almost equal to unity) and n is very large such that the average of series (viz., $n. p.$) is a finite number. Experience has shown that this distribution is good for calculating the probabilities associated with X occurrences in a given time period or specified area.

The random variable of interest in Poisson distribution is number of occurrences of a given event during a given interval (interval may be time, distance, area, etc.). We use capital X to represent the discrete random variable and lower case x to represent a specific value that capital X can take. The probability function of this distribution is generally written as under:

$$f(X_i = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$x = 0, 1, 2\ldots$$

Where, λ = Average number of occurrences per specified interval.[2] In other words, it is the mean of the distribution.

e = 2.7183 being the basis of natural logarithms.

x = Number of occurrences of a given event.

Poisson Process

The distribution applies in case of Poisson process which has following characteristics.

- Concerning a given random variable, the mean relating to a given interval can be estimated on the basis of past data concerning the variable under study.
- If we divide the given interval into very very small intervals we will find:
  (a) The probability that exactly one event will happen during the very very small interval is a very small number and is constant for every other very small interval.
  (b) The probability that two or more events will happen within a very small interval is so small that we can assign it a zero value.
  (c) The event that happens in a given very small interval is independent, when the very small interval falls during a given interval.
  (d) The number of events in any small interval is not dependent on the number of events in any other small interval.

Parameter and Important Measures of Poisson Distribution

NOTES

Poisson distribution depends upon the value of λ, the average number of occurrences per specified interval which is its only parameter. The probability of exactly x occurrences can be found out using Poisson probability function stated above. The expected value or the mean of Poisson random variable is λ and its variance is also λ. The standard deviation of Poisson distribution is, $\sqrt{\lambda}$.

Underlying the Poisson model is the assumption that if there are on the average λ occurrences per interval t, then there are on the average k λ occurrences per interval kt. For example, if the number of arrivals at a service counted in a given hour, has a Poisson distribution with λ = 4, then y, the number of arrivals at a service counter in a given 6 hour day, has the Poisson distribution λ = 24, i.e., 6×4.

When to Use Poisson Distribution

The use of Poisson distribution is resorted to those cases when we do not know the value of 'n' or when 'n' can not be estimated with any degree of accuracy. In fact, in certain cases it does not make any sense in asking the value of 'n'. For example, the goals scored by one team in a football match are given, it cannot be stated how many goals could not be scored. Similarly, if one watches carefully one may find out how many times the lightning flashed but it is not possible to state how many times it did not flash. It is in such cases we use Poisson distribution. The number of death per day in a district in one year due to a disease, the number of scooters passing through a road per minute during a certain part of the day for a few months, the number of printing mistakes per page in a book containing many pages, are a few other examples where Poisson probability distribution is generally used.

Example 2: Suppose that a manufactured product has 2 defects per unit of product inspected. Use Poisson distribution and calculate the probabilities of finding a product without any defect, with 3 defects and with four defects.

Solution: If the product has 2 defects per unit of product inspected. Hence, λ = 2.

Poisson probability function is as follows:

$$f(X_i = x) = -\frac{\lambda^x . e^{-\lambda}}{x!}$$

$$x = 0, 1, 2, \ldots$$

Using the above probability function, we find the required probabilities as under:

$$P(\text{without any defects, i.e., } x = 0) = \frac{2^0 . e^{-2}}{0}$$

$$= \frac{1.(0.13534)}{1} = 0.13534$$

NOTES

$$P(\text{with 3 defects, i.e., } x = 3) = \frac{2^3.e^{-2}}{3} = \frac{2\times2\times2(0.13534)}{3\times2\times1}$$

$$= \frac{0.54136}{3} = 0.18045$$

$$P(\text{with 4 defects, i.e., } x = 4) = \frac{2^4.e^{-2}}{4} = \frac{2\times2\times2\times2(0.13534)}{4\times3\times2\times1}$$

$$= \frac{0.27068}{3} = 0.09023$$

Example 3: How would you use a Poisson distribution to find approximately the probability of exactly 5 successes in 100 trials the probability of success in each trial being $p = 0.1$?

Solution: In the question we have been given,

$n = 100$ and $p = 0.1$

$\therefore \quad \lambda = n.p = 100 \times 0.1 = 10$

To find the required probability, we can use Poisson probability function as an approximation to Binomial probability function as shown below:

$$f(X_i = x) = \frac{\lambda^x.e^{-\lambda}}{x!} = \frac{(n.p)^x.e^{-(n.p)}}{x!}$$

or $$P(5)^7 = \frac{10^5.e^{-10}}{5} = \frac{(100000)(0.00005)}{5\times4\times3\times2\times1} = \frac{5.00000}{5\times4\times3\times2\times1}$$

$$= \frac{1}{24} = 0.042$$

Check Your Progress

6. What is Poisson distribution?
7. Where and when will you use Poisson distribution?

6.4 LIMITING FORM OF BINOMIAL AND POISSON FITTING

Fitting a Binomial Distribution

When a binomial distribution is to be fitted to the given data, then the following procedure is adopted:

(a) Determine the values of 'p' and 'q' keeping in view that $\bar{X} = n.\ p.$ and $q = (1 - p)$.

(b) Find the probabilities for all possible values of the given random variable applying the binomial probability function, viz.,

$$f(X_i = r) = {}^nC_r p^r q^{n-r}$$
$$r = 0,\ 1,\ 2,\ldots n$$

(c) Work out the expected frequencies for all values of random variable by multiplying N (the total frequency) with the corresponding probability.

(d) The expected frequencies, so calculated, constitute the fitted binomial distribution to the given data.

Fitting a Poisson Distribution

When a Poisson distribution is to be fitted to the given data, then the following procedure is adopted:

(a) Determine the value of λ, the mean of the distribution.

(b) Find the probabilities for all possible values of the given random variable using the Poisson probability function, viz.,

$$f\left(X_i = x\right) = \frac{\lambda^x . e^{-\lambda}}{x}$$

$$x = 0,\ 1,\ 2,\ldots$$

(c) Work out the expected frequencies as follows:

$$n.p.(X_i = x)$$

(d) The result of case (c) above is the fitted Poisson distribution to the given data.

Poisson Distribution as an Approximation of Binomial Distribution

Under certain circumstances Poisson distribution can be considered as a reasonable approximation of Binomial distribution and can be used accordingly. The circumstances which permit all this, are when 'n' is large approaching to infinity and p is small approaching to zero (n = Number of Trials, p = Probability of 'Success'). Statisticians usually take the meaning of large n, for this purpose, when $n \geq 20$ and by small 'p' they mean when $p \leq 0.05$. In cases where these two conditions are fulfilled, we can use mean of the binomial distribution (viz., $n.p.$) in place of the mean of Poisson distribution (viz., λ) so that the probability function of Poisson distribution becomes as stated below:

$$f\left(X_i = x\right) = \frac{\left(n.p\right)^x . e^{-(np)}}{x}$$

We can explain Poisson distribution as an approximation of the Binomial distribution with the help of following example.

Example 4: Given is the following information:

(a) There are 20 machines in a certain factory, i.e., $n = 20$.

(b) The probability of machine going out of order during any day is 0.02.

NOTES

What is the probability that exactly 3 machines will be out of order on the same day? Calculate the required probability using both Binomial and Poissons Distributions and state whether Poisson distribution is a good approximation of the Binomial distribution in this case.

Solution: Probability, as per *Poisson probability function* (using *n.p* in place of λ)

(since $n \geq 20$ and $p \leq 0.05$)

$$f(X_i = x) = \frac{(n.p)^x .e^{-np}}{x!}$$

Where, x means number of machines becoming out of order on the same day.

$$P(X_i = 3) = \frac{(20 \times 0.02)^3 .e^{-(20\times0.02)}}{3}$$

$$= \frac{(0.4)^3 .(0.67032)}{3 \times 2 \times 1} = \frac{(0.064)(0.67032)}{6}$$

$$= 0.00715$$

Probability, as per *Binomial probability function,*

$$f(X_i = r) = {}^nC_r p^r q^{n-r}$$

Where, $n = 20$, $r = 3$, $p = 0.02$ and hence $q = 0.98$

$$\therefore \quad f(X_i = 3) = {}^{20}C_3 (0.02)^3 \ (0.98)^{17}$$

$$= 0.00650$$

The difference between the probability of 3 machines becoming out of order on the same day calculated using probability function and binomial probability function is just 0.00065. The difference being very very small, we can state that in the given case Poisson distribution appears to be a good approximation of Binomial distribution.

Example 5: How would you use a Poisson distribution to find approximately the probability of exactly 5 successes in 100 trials the probability of success in each trial being $p = 0.1$?

Solution: In the question we have been given,

$$n = 100 \text{ and } p = 0.1$$

$$\therefore \quad \lambda = n.p = 100 \times 0.1 = 10$$

To find the required probability, we can use Poisson probability function as an approximation to Binomial probability function, as shown below:

NOTES

$$f\left(X_i = x\right) = \frac{\lambda^x . e^{-\lambda}}{x!} = \frac{\left(n.p\right)^x . e^{-(n.p)}}{x!}$$

Or, $$P(5)^7 = \frac{10^5 . e^{-10}}{5} = \frac{(100000)(0.00005)}{5 \times 4 \times 3 \times 2 \times 1} = \frac{5.00000}{5 \times 4 \times 3 \times 2 \times 1}$$

$$= \frac{1}{24} \quad = 0.042$$

Check Your Progress

8. How can we measure the area under the curve?
9. Under what circumstances, is a Poisson distribution considered as an approximation of binomial distribution?

6.5 DISCRETE THEORETICAL DISTRIBUTIONS

Theoretical Distributions: If a certain hypothesis is assumed, it is sometimes possible to derive mathematically, what the frequency distributions of certain universes should be. Such distributions are called **theoretical distributions.**

Binomial Distribution

Binomial distribution was discovered by James Bernoulli in the year 1700.

Let there be an event the probability of its success* is P and the probability of its failure is Q is one trial, so $P + Q = 1$

Consider a set of n independent trials and the probability P of success is the same in every trial, then $Q = 1 - P$ is the probability of failure in any trial.

Let the set of n trials be repeated N times, where N is a very large number. Out of these N, there will be sets with few success and also with number of successes and so on.

Now the probability that the first k trials are successes and the remaining $(n - k)$ trial are failures is $P^k Q^{(n-k)}$.

Since k can be chosen out of n-trials in ${}^n c_k$ ways, the probability of k-successes, $P(k)$ in a series of n-independent trials is given by,

$$P(k) \quad = {}^n c_k . P^k . Q^{(n-k)}$$

The probability distribution of the number of successes, so obtained is called the 'Binomial probability distribution'.

The probabilities of 0, 1, 2, ... n successes are ${}^n C_0 P^0 Q^n$, ${}^n C_1 P^1 Q^{n-1}$, ...${}^n C_n P^n Q^0$, are the successive terms of binomial expansion $(Q + P)^n$.

NOTES

Definition: A random variable X is said to follow binomial distribution if it assumes only non-negative values and its probability mass function is given by,

$$P(X=k)=P(k)=\begin{cases} {}^nc_k.P^k.Q^{(n-k)} & ; k=0,1,2,\ldots Q=1-P \\ 0 & ; \text{Otherwise} \end{cases}$$

Usually we use the notation $X \sim B(n, P)$ to denote that the random variable X follows binomial distribution with parameters n and P.

Notes**:** 1. Since n and P are independent constants in the binomial distribution , they are called the parameters of distribution.

2. $$\sum_{k=0}^{n} P(k) = \sum_{k=0}^{n} {}^nC_k.P^k.Q^{n-k} = (Q+P)^n = 1$$

3. If the experiment consisting of n-trials is repeated n times, then the frequency function of the binomial distribution is given by,

$$f(k) \quad = N.P(k) = N.\ [nC_k.P^k.Q^{(n-k)}];\ k = 0,\ 1,\ 2,\ \ldots$$

and the expected frequencies of 0, 1, 2, ...n successes are the successive terms of the binomial expansion, $N(Q+P)^n$; $(Q+P=1)$

Moments

The first four moments about origin of binomial distribution are obtained as follows:

(i) Mean or First Moment about Origin

$$\mu'_1 = E(X) = \sum_{x=0}^{n} x.{}^nC_x.P^x.Q^{n-x}$$

$$= nP\sum_{x=1}^{n} {}^{n-1}C_{x-1}.P^{x-1}.Q^{n-x}$$

$$= nP[Q^{n-1} + {}^{n-1}C_1P.Q^{n-2} + \ldots + P^{n-1}]$$

$$= nP(Q+P)^{n-1} = nP \quad \text{as } P+Q=1$$

So, mean or $\mu'_1 = nP$.

(ii) Second Moment about Origin

$$\mu'_2 = E(X^2) = \sum_{k=0}^{n} x^2.{}^nC_x.P^x.Q^{n-x}$$

$$= \sum_{x=0}^{n}\left[(x + x(x-1))^n C_x.P^x.Q^{n-x}\right]$$

$$= \sum_{x=0}^{n} x.^nC_x.P^x.Q^{n-x} + \sum_{x=0}^{n} x(x-1).^nC_x.P^x.Q^{n-x}$$

$$= nP + n(n-1)P^2. \sum_{x=2}^{n} {}^{n-2}C_{x-2}.P^{x-2}.Q^{n-x}$$

$$= nP + n(n-1)P^2.(Q - P)^{n-2}$$

$$\mu'_2 = nPQ + n^2P^2.$$

NOTES

(iii) Third Moment about Origin

$$\mu'_3 = E(X^3) = \sum_{k=0}^{n} x^3.^nC_x.P^x.Q^{n-x}$$

On simplifying, we get,

$$\mu'_3 = n(n-1)(n-2)P^3 + 3n(n-1)P^2 + nP.$$

(iv) Fourth Moment about Origin

$$\mu'_4 = E(X^4) = \sum_{k=0}^{n} x^4.^nC_x.P^x.Q^{n-x}$$

And on simplifying, we get,

$$\mu'_4 = n(n-1)(n-2)(n-3)P^4 + 6n(n-1)(n-2)P^3 + 7n(n-1)P^2 + nP.$$

Now, the moment about 'Mean of Binomial Distribution will be Discussed'.

(i) **First Moment about Mean**

$$\mu_0 = 0 \quad \text{(always)}$$

(ii) **Second Moment about Mean**

$$\mu_2 = \mu'_2 - (\mu'_1)^2 = E(X^2) - [E(x)]^2$$

$$= nPQ + n^2P^2 - n^2P^2 = nPQ$$

Thus, Var $(X) = \mu_2 = nPQ$ and standard deviation $\sigma(X) = \sqrt{nPQ}$

(iii) **Third and Fourth Moment about Mean**

(*a*) $\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3$

$= nPQ\,(Q - P)$ (On simplification)

(*b*) $\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2(\mu'_1)^2 - 2(\mu'_1)^4$

$= 3n^2P^2Q^2 - 6nP^2Q^2 + nPQ$

$= nPQ[1 + 3(n-2)PQ]$

Hence, $\beta_1 = \dfrac{\mu^2{}_3}{\mu^3{}_2} = \dfrac{n^2P^2Q^2(Q-P)^2}{n^3P^3Q^3} = \dfrac{(1-2P)^2}{nPQ}$

$$\beta_2 = \frac{\mu_4}{\mu^2{}_2} = \frac{nPQ[1+3(n-2)PQ]}{n^2P^2Q^2} = 3 + \frac{1-6PQ}{nPQ}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{Q-P}{\sqrt{nPQ}} \text{ and } \gamma_2 = \beta_2 - 3 = \frac{1-6PQ}{nPQ}$$

NOTES

Moment Generating Function of Binomial Distribution: Let x have a binomial distribution with probability function,

$$P(x) = {}^nC_x P^x Q^{n-x}\,; \quad x = 0, 1, 2, \ldots n$$

The moment generating function about the origin is given as

$$\mathrm{M}_0(t) = E(e^{tx}) = \sum_{x=0}^{n} e^{tx}.{}^nC_x\, P^x.Q^{n-x}$$

$$= \sum_{x=0}^{n} {}^nC_x\,(Pe^t)^x.Q^{n-x}$$

$$M_0(t) = (Q + Pe^t)^n$$

Moment generating function about mean nP is given by,

$$M_{nP}(t) = E[e^{t(x-nP)}] = E(e^{tx}.e^{-nPt})$$

$$= e^{-nPt}.E(e^{tx}) = e^{-nPt}.M_0(t)$$

$$= e^{-nPt}(Q + e^t.P)^n$$

$$= (Q.e^{-Pt} + P.e^{(1-P)t})^n$$

$$M_{nP}(t) = (Q.e^{-Pt} + P.e^{Qt})^n$$

Poisson Distribution with Mean and Variance

The Poisson distribution is a limiting case of binomial distribution when the probability of success or failure (i.e., P or Q) is vary small and the number of trial n is very large (i.e. $n \to \infty$) enough so that nP is a finite constant say λ i.e. $nP = \lambda$. Under these conditions, $P(x)$ the probability of x success in the binomial distribution,

$$P(x) = P(X-x) = {}^nC_x P^x Q^{n-x} \text{ can be written as,}$$

$$P(x) = \frac{n!}{x!.(n-x)!}.\left(\frac{\lambda}{n}\right)^x.\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^n \frac{n!}{n^x(n-x)!\left(1-\frac{\lambda}{n}\right)^x}$$

Using $\mathrm{Lim}\; n \to \infty$, we have,

$$\mathrm{P}(x) = \frac{\lambda^x}{x!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n \lim_{n\to\infty}\frac{n!}{(n-x)!\left(1-\frac{\lambda}{n}\right)^x.n^x} \qquad \ldots(6.1)$$

We know that $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}$ and $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{x} = 1$

And using Stirling's formula for $n!$, we have,

$$n! = \sqrt{2\pi}.\, n^{n+1/2}e^{-n}$$

$$\frac{n!}{(n-x)!} = \frac{\sqrt{2\pi}.n^{n+1/2}.e^{-n}}{\sqrt{2\pi}.(n-x)^{n-x+1/2}.e^{-n+x}}$$

So, $$\frac{n!}{(n-x)!} = \frac{n^{n+1/2}.e^{-n}}{(n-x)^{n-x+1/2}.e^{-n+x}.n^{x}}$$

Thus, Equation (6.1) becomes,

$$P(x) = \frac{\lambda^{x}}{x!}.e^{-\lambda}.\lim_{n\to\infty}\frac{n^{n+1/2}.e^{-n}}{(n-x)^{n-x+1/2}.e^{-n+x}.n^{x}}$$

$$= \frac{\lambda^{x}.e^{-\lambda}}{x!.e^{x}}.\lim_{n\to\infty}\frac{n^{n+1/2}.e^{-n}}{(n-x)^{n-x+1/2}.e^{-n}.n^{x}}$$

$$= \frac{\lambda^{x}.e^{-\lambda}}{x!.e^{x}}.\lim_{n\to\infty}\frac{n^{n-x+1/2}}{n^{n-x+1/2}\left(1-\frac{x}{n}\right)^{n-x+\frac{1}{2}}}$$

$$= \frac{\lambda^{x}.e^{-\lambda}}{x!.e^{x}}.\lim_{n\to\infty}\left[\frac{1}{\left(1-\frac{x}{n}\right)^{n}.\left(1-\frac{x}{n}\right)^{-x+\frac{1}{2}}}\right]$$

$$= \frac{\lambda^{x}.e^{-\lambda}}{x!.e^{x}}.\frac{1}{e^{-x}.1} = \frac{\lambda^{x}.e^{-\lambda}}{x!}$$

Thus,

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

When $n \to \infty$, $nP = \lambda$ and $P \to 0$

Here, λ is known as the parameter of Poisson distribution.

Definition: A random variable X is said to follow a Poisson distribution if it assumes only non-negative values and its probability mass function is given by,

$$P(x, \lambda) = P(X = x) = \frac{\lambda^x . e^{-\lambda}}{x!}; x = 0, 1, 2, ...; \lambda > 0.$$

Constants of the Poisson Distribution:

NOTES

(i) Mean, $\mu_1' = \lambda = E(X)$.

(ii) $\mu_2' = E(X^2) = \lambda + \lambda^2$.

(iii) $\mu_3' = E(X^3) = \lambda^3 + 3\lambda^2 + \lambda$.

(iv) $\mu_4' = E(X^4) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$.

(v) First moment about mean $\mu_1 = 0$.

(vi) $V(X) = \mu_2 = E(X^2) - [E(X)]^2 = \mu_2' - \mu_1'^2$

$= \lambda + \lambda^2 - \lambda^2 = \lambda$

Note that, Mean = Variance = λ.

(vii) Standard deviation $\sigma(X) = \sqrt{\lambda}$.

(viii) μ_3, i.e. the third moment about mean

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

$$= (\lambda^3 + 3\lambda^2 + \lambda) - 3\lambda(\lambda^2 + \lambda) + 2\lambda^3 = \lambda.$$

(ix) $\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4 = 3\lambda^2 + \lambda.$

Thus, coefficients of skewness and kurtosis are given by,

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\lambda^2}{\lambda^3} = \frac{1}{\lambda} \text{ and } \gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{\lambda}}$$

$$\text{Also, } \beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{\lambda} \text{ and } \gamma_2 = \beta_2 - 3 = \frac{1}{\lambda}$$

Hence, the Poisson distribution is always a skewed distribution if $\lim\limits_{\lambda \to \infty}$, we get,

$$\beta_1 = 0, \beta_2 = 3.$$

Negative Binomial Distribution: Suppose there are n-trials of an event. We assume that

(i) The n-trials are independent.

(ii) P is the probability of success which remains constant from trial to trial.

Let $f(x; r, P)$ denote the probability that there are x failures preceeding the rth success in $(x + r)$ trials.

Now, in $(x + r)$ trials the last trial must be a success with probability P. Then in the remaining $(x + r - 1)$ trials, we must have $(r - 1)$ successes whose probability is given by

$$^{x+r-1}C_{r-1} . P^{r-1} . Q^x.$$

Therefore, the compound probability theorem $f(x; r, P)$ is given by the product of these two probabilities.

So $\quad f(x; r, P) = {}^{x+r-1}C_{r-i}.\ P^{r-1}.\ Q^x.P$

$$= {}^{x+r-1}C_{r-i}.\ P^r.\ Q^x.$$

Moment Generating Function of Poisson Distribution

NOTES

Let $P(X = x) = e^{-\lambda}.\dfrac{\lambda^x}{x!}$; $x = 0, 1, 2, \ldots \infty$; $\lambda > 0$ be a Poisson distribution.

Then the moment generating function is given by,

$$M(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx}.\frac{e^{-\lambda}.\lambda^x}{x!}$$

$$= e^{-\lambda}.\sum_{x=0}^{\infty}\frac{(\lambda.e^t)^x}{x!}$$

$$= e^{-\lambda}.e^{\lambda e^t} = e^{\lambda(e^t-1)}$$

and moment generating function about mean is,

$$M_\lambda(t) = E(e^{t(x-\lambda)}) = e^{-\lambda t}.\ E(e^{tx})$$

$$= e^{-\lambda t}.e^{\lambda(e^t-1)}$$

So, $\quad M_\lambda(t) = e^{(-\lambda t + \lambda e^t - \lambda)}$

Definition: A random variable X is said to follow a negative binomial distribution, if its probability mass function is given by,

$$P(X = x) = P(x) = {}^{x+r-1}C_{r-1}.\ P^r.\ Q^x;\ x = 0, 1, 2, \ldots$$

0; otherwise.

Also, we know that ${}^nC_r = {}^nC_{n-r}$

So $\quad {}^{x+r-1}C_{r-1} = {}^{x+r-1}C_x$

$$= \frac{(x+r-1)(x+r-2)\ldots(r+1).r}{x!}$$

$$= \frac{(-1)^x\ (-r)(-r-1)\ldots(-r+x+2)(-r+x+1)}{x!} = (-1)^x.\ {}^{-r}C_x$$

So $\quad p(x) = \begin{cases} {}^{-r}C_r.\ p^r(-q)^x; & x = 0, 1, 2, \ldots \\ 0 & ;\text{ otherwise} \end{cases}$

which is the $(x + 1)$th term in the expansion of $P^r(1 - Q)^{-r}$, a binomial expansion with a negative index. Hence, the distribution is called a negative binomial distribution. Also,

$$\sum_{x=0}^{\infty} P(x) = P^r\sum_{x=0}^{\infty}{}^{-r}C_x(-Q)^x = P^r \times (1 - Q)^{-r} = 1$$

Therefore, $P(x)$ represents the probability function and the discrete variable which follows this probability function is called the negative binomial variable.

Example 6: A continuous random variable X has a probability distribution function $f(x) = 3x^2$, $0 \le x \le 1$.

Find a and b such that

NOTES

(*i*) $p\{X \le a\} = p\{X > a\}$, and

(*ii*) $p\{X > b\} = 0.05$

Solution: (*i*) Since $P\{X \le a\} = P\{X > a\}$

each must be equal to $\frac{1}{2}$, because total probability is always 1.

$$\therefore \quad P\{X \le a\} = \frac{1}{2} \Rightarrow \int_0^a f(x)dx = \frac{1}{2}$$

Or, $$3\int_0^a x^2 dx = \frac{1}{2} \Rightarrow 3\left[\frac{x^3}{3}\right]_0^a = \frac{1}{2}$$

Or, $$a^3 = \frac{1}{2} \Rightarrow a = \left(\frac{1}{2}\right)^{1/3}$$

(*ii*) $p\{X < b\} = 0.05.$

$$\Rightarrow \int_b^1 f(x)dx = 0.05 \Rightarrow 3\int_b^1 x^2 dx = 0.05$$

$$\Rightarrow 3\left[\frac{x^3}{3}\right]_b^1 = 0.05 \Rightarrow 1 - b^3 = \frac{1}{20}$$

$$b = \left(\frac{19}{20}\right)^{1/3}$$

Example 7: A probability curve $y = f(x)$ has a range from 0 to ∞. If $f(x) = e^{-x}$ find the mean and variance and the third moment about mean.

Solution: We know that, the rth moment about origin,

$$\mu'_r = \int_0^\infty x^r f(x)dx = \int_0^\infty x^r . e^{-x} dx$$

$$= \Gamma(r+1) = r! \text{ (Using Gamma Integral)}$$

Substituting, $r = 1, 2$, and 3, we have,

Mean, $\mu'_1 = 1! = 1$

$\mu'_2 = 2! = 2$ and $\mu'_3 = 3! = 6$

Thus, variance $= \mu_2 = \mu'_2 - (\mu'_1)^2 = 2 - 1 = 1$

And $\mu_3 = \mu'_3 - 3\mu'_2 . \mu'_1 + 2(\mu'_1)^3 = 6 - 3 \times 2 + 2 = 2$ is the required third moment about mean.

NOTES

Uniform Distribution

A random variable X is said to have a continuous uniform distribution over an interval (a, b) if its probability density function is constant say k, over the entire range of X.

That is, $$f(x) = \begin{cases} k; & a < X < b \\ 0; & \text{otherwise} \end{cases}$$

Since total probability is always unity, we have,

$$\int_a^b f(x)\,dx = 1 \quad \Rightarrow \quad k = \int_a^b dx = 1$$

Or, $$k = \frac{1}{b-a}$$

Thus, $$f(x) = \begin{cases} \frac{1}{b-a}; & a < X < b \\ 0\;; & \text{otherwise} \end{cases}$$

This is also known as rectangular distribution as the curve $y = f(x)$ describes a rectangle over the x-axis and between the ordinates at $x = a$ and $x = b$.

The distribution function $F(x)$ is given by,

$$F(x) = \begin{cases} 0 & ;\text{if } -\infty < x < a \\ \frac{x-a}{b-a}; & a \le x \le b \\ 1\;; & b < x < \infty \end{cases}$$

Since $F(x)$ is not continuous at $x = a$ and $x = b$, it is not differentiable at these points. Thus, $\frac{d}{dx}F(x) = f(x) = \frac{1}{b-a} \neq 0,$ exists everywhere except at the points $x = a$ and $x = b$ and consequently probability distribution function $f(x)$ is given by,

$$f(x) = \begin{cases} \frac{1}{b-a}; & a < x < b \\ 0\;; & \text{otherwise} \end{cases}$$

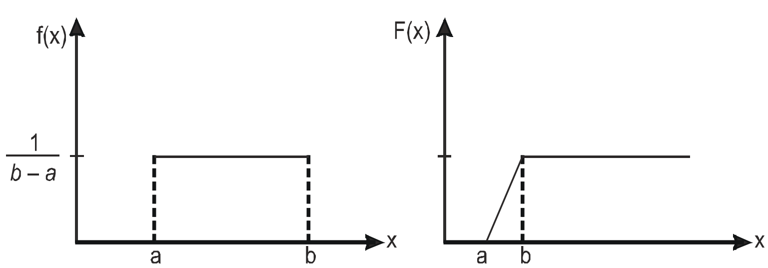


Fig. 6.5 *Uniform Distribution*

Moments of Uniform Distribution:

$$\mu'_r = \int_b^a x^r f(x)\,dx = \frac{1}{(b-a)}\left(\frac{b^{r+1}-a^{r+1}}{r+1}\right).$$

NOTES

In particular,

$$\text{mean, } \mu'_1 = \frac{1}{b-a}\left(\frac{b^2-a^2}{2}\right) = \left(\frac{b+a}{2}\right)$$

$$\text{And } \quad \mu'_2 = \frac{1}{b-a}\left(\frac{b^3-a^3}{3}\right) = \frac{1}{3}(b^2+ab+a^2)$$

$$\therefore \quad \text{Variance} = \mu_2 = \mu'_2 - (\mu'_1)^2$$

$$= \frac{1}{3}(b^2+ab+a^2) - \left(\frac{b+a}{2}\right)^2$$

$$= \frac{1}{12}(b-a)^2$$

Moment generating function is given by,

$$M_X(t) = \int_a^b e^t f(x)\,dx = \frac{e^{bt}-e^{at}}{t\,(b-a)}$$

And the characteristic function is given by,

$$\phi_x(t) = \int_a^b e^{itx} f(x)\,dx = \frac{e^{ibt}-e^{iat}}{t\,(b-a)}$$

Check Your Progress

10. Who discovered binomial distribution?
11. When is a random variable X said to follow binomial distribution?
12. What is Poisson distribution?

6.6 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Bernoulli process or Binomial distribution is considered appropriate and has the following characteristics;
   (a) Dichotomy: This means that each trial has only two mutually exclusive possible outcomes. For example, success or failure, yes or no, heads or tail, etc.
   (b) Stability: This means that the probability of the outcome of any trial is known and remains fixed over time, i.e., remains the same for all the trials.

(c) Independence: This means that the trials are statistically independent, i.e., to say the happening of an outcome or the event in any particular trial is independent of its happening in any other trial or trials.

2. The probability function of binomial distribution is written as under:

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

Where, n = Numbers of trials.

p = Probability of success in a single trial.

q = $(1-p)$ = Probability of failure in a single trial.

r = Number of successes in n trials.

3. The parameters of binomial distribution are p and n, where p specifies the probability of success in a single trial and n specifies the number of trials.
4. The important measures of binomial distribution are:

$$\text{Skewness} = \frac{1-2p}{\sqrt{n.p.q}}$$

$$\text{Kurtosis} = 3 + \frac{1-6p+6q^2}{n.p.q}$$

5. We need to use binomial distribution under the following circumstances:
   (a) When we have to find the probability of heads in 10 throws of a fair coin.
   (b) When we have to find the probability that 3 out of 10 items produced by a machine, which produces 8% defective items on an average, will be defective.
6. Poisson distribution is a discrete probability distribution that is frequently used in the context of Operations Research. Unlike binomial distribution, Poisson distribution cannot be deduced on purely theoretical grounds based on the conditions of the experiment. In fact, it must be based on the experience, i.e., on the empirical results of past experiments relating to the problem under study.
7. Poisson distribution is used when probability of happening of an event is very small and n is very large such that the average of series is a finite number. This distribution is good for calculating the probabilities associated with X occurrences in a given time period or specified area.
8. For measuring the area under a curve, we make use of the statistical tables constructed by mathematicians. Using these tables, we can find the area that the normally distributed random variable will lie within certain distances from the mean. These distances are defined in terms of standard deviations. While using the tables showing the area under the normal curve, it is considered in

NOTES

NOTES

terms of standard variate, which means standard deviations without units of measurement and it is calculated as:

$$Z = \frac{X - \mu}{\sigma}$$

Where, Z = The standard variate or number of standard deviations from X to the mean of the distribution.

X = Value of the random variable under consideration.

μ = Mean of the distribution of the random variable.

σ = Standard deviation of the distribution.

9. When n is large approaching to infinity and p is small approaching to zero, Poisson distribution is considered as an approximation of binomial distribution.
10. Binomial distribution was discovered by James Bernoulli.
11. A random variable X is said to follow binomial distribution if it assumes only non-negative values and its probability mass function is given by,

$$P(X = k) = P(k) = \begin{cases} {}^{n}c_k.P^k.Q^{(n-k)} & ; k = 0, 1, 2, ...\ Q = 1 - P \\ 0 & ; \text{otherwise} \end{cases}$$

12. Poisson distribution is a limiting case of binomial distribution when the probability of success or failure is very small and the number of trial n is very large.

6.7 SUMMARY

- The binomial distribution describes discrete data resulting from what is often called as the Bernoulli process. The tossing of a fair coin a fixed number of times is a Bernoulli process and the outcome of such tosses can be represented by the binomial distribution. The name of Swiss mathematician Jacob Bernoulli is associated with this distribution.
- The expected value of random variable [i.e., $E(X)$] or mean of random variable (i.e., $\overline{X}$) of the binomial distribution is equal to $n.p$ and the variance of random variable is equal to $n.\ p.\ q$ or $n.\ p.\ (1 - p)$.
- Unlike binomial distribution, Poisson distribution cannot be deducted on purely theoretical grounds based on the conditions of the experiment. In fact, it must be based on experience, i.e., on the empirical results of past experiments relating to the problem under study.
- Poisson distribution is appropriate specially when probability of happening of an event is very small (so that q or $(1-p)$ is almost equal to unity) and n is very large such that the average of series (viz., $n.\ p.$) is a finite number.

NOTES

- Poisson distribution depends upon the value of λ, the average number of occurrences per specified interval which is its only parameter.
- If a certain hypothesis is assumed, it is sometimes possible to derive mathematically, what the frequency distributions of certain universes should be. Such distributions are called theoretical distributions.
- A random variable X is said to follow binomial distribution if it assumes only non-negative values and its probability mass function is given by,

$$P(X = k) = P(k) = \begin{cases} {}^{n}c_k.P^k.Q^{(n-k)} & ; k = 0, 1, 2, \ldots Q = 1 - P \\ 0 & ; \text{Otherwise} \end{cases}$$

- The Poisson distribution is a limiting case of binomial distribution when the probability of success or failure (i.e., P or Q) is vary small and the number of trial n is very large (i.e. $n \to \infty$) enough so that nP is a finite constant say λ, i.e., $nP = \lambda$.
- A random variable X is said to follow a Poisson distribution if it assumes only non-negative values and its probability mass function is given by

$$P(x, \lambda) \;=\; P(X = x) = \frac{\lambda^x.e^{-\lambda}}{x!}\,; x = 0, 1, 2, \ldots; \lambda > 0.$$

- A random variable X is said to have a continuous uniform distribution over an interval (a, b) if its probability density function is constant say k, over the entire range of X.

6.8 KEY WORDS

- **Binomial distribution:** It is also called as Bernoulli process and is used to describe discrete random variable.
- **Poisson distribution:** It is used to describe the empirical results of past experiments relating to the problem and plays important role in queuing theory, inventory control problems and risk models.

6.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define probability distribution and probability functions.
2. Describe binomial distribution and its measures.
3. How a binomial distribution can be fitted to a given data?

4. Describe Poisson distribution and its important measures.
5. Poisson distribution can be an approximation of binomial distribution. Explain.

NOTES

6. When is the Poisson distribution used?
7. Write the formula for measuring the area under the curve.
8. Explain the circumstances when the normal probability distribution can be used.

Long-Answer Questions

1. Given is the following probability distribution:

X_i	$pr(X_i)$
0	1/8
1	2/8
2	3/8
3	2/8

Calculate the expected value of X_i, its variance, and standard deviation.

2. A coin is tossed 3 times. Let X be the number of runs in the sequence of outcomes: first toss, second toss, third toss. Find the probability distribution of X. What values of X are most probable?
3. (a) Explain the meaning of Bernoulli process pointing out its main characteristics.

   (b) Give a few examples narrating some situations wherein binomial *pr.* distribution can be used.
4. State the distinctive features of the Binomial, Poisson and Normal probability distributions. When does a Binomial distribution tend to become a Normal and a Poisson distribution? Explain.
5. Explain the circumstances when the following probability distributions are used:
   (a) Binomial distribution
   (b) Poisson distribution
   (c) Normal distribution
6. Certain articles were produced of which 0.5 per cent are defective, are packed in cartons, each containing 130 articles. When proportion of cartons are free from defective articles? What proportion of cartons contain 2 or more defective?
   (Given $e^{-0.5}$=0.6065).

7. The following mistakes per page were observed in a book:

No. of Mistakes Per Page	*No. of Times the Mistake Occurred*
0	211
1	90
2	19
3	5
4	0
Total	345

Fit a Poisson distribution to the data given above and test the goodness of fit.

8. In a distribution exactly normal, 7 per cent of the items are under 35 and 89 per cent are under 63. What are the mean and standard deviation of the distribution?
9. Assume the mean height of soldiers to be 68.22 inches with a variance of 10.8 inches. How many soldiers in a regiment of 1000 would you expect to be over six feet tall?.

NOTES

6.10 FURTHER READINGS

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 7 SPECIAL DISTRIBUTIONS

NOTES

Structure

7.0 INTRODUCTION

Any statistical hypothesis test, in which the test statistic has a Chi-square distribution, when the null hypothesis is true, is termed as Chi-square test. Chi-square test is a non-parametric test of statistical significance for bivariate tabular analysis, also known as cross-breaks. Amongst the several tests used in statistics for judging the significance of the sampling data, Chi-square test, developed by Prof. Fisher, is considered an important test. Chi-square, symbolically written as χ^2 (pronounced as Ki-square), is a statistical measure with the help of which it is possible to assess the significance of the difference between the observed frequencies and the expected frequencies obtained from some hypothetical universe. Chi-square tests enable us to test and compare whether more than two population proportions can be considered equal. Hence, it is a statistical test commonly used to compare observed data with expected data and testing the null hypothesis, which states that there is no significant difference between the expected and the observed result.

In this unit, you will study about the Gamma distribution, Chi-square distribution, the normal distribution and the bivariate normal distribution.

7.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand about the Gamma distribution
- Explain various Chi-square distribution
- Describe the normal distribution
- Analyse the bivariate normal distribution

7.2 THE GAMMA DISTRIBUTION

NOTES

The **Erlang distribution** is a continuous probability distribution with wide applicability primarily due to its relation to the Exponential and Gamma distributions. The Erlang distribution was developed by A. K. Erlang. He developed the Erlang distribution to examine the number of telephone calls which might be made at the same time to the operators of the switching stations. This work on telephone traffic engineering has been expanded to consider waiting times in queuing systems in general. Erlang distribution is now used in the fields of stochastic processes and biomathematics.

Given a Poisson distribution with a rate of change λ, the Distribution Function $D(x)$ giving the waiting times until the hth Poisson event is:

$$D(x) = 1 - \sum_{k=0}^{h-1} e^{-\lambda x} \frac{(\lambda x)^k}{k!}$$

$$= 1 - \frac{\Gamma(h, x\lambda)}{\Gamma(h)}$$

For $x \in (0, \infty)$, where $\Gamma(x)$ is a complete gamma function, and $\Gamma(a, x)$ an incomplete gamma function. With h explicitly an integer, this distribution is known as the Erlang distribution and has the following probability function:

$$P(x) = \frac{\lambda(\lambda x)^{h-1}}{(h-1)!} e^{-\lambda x}$$

It is closely related to the gamma distribution, which is obtained by letting a $\equiv h$ (not necessarily an integer) and defining $\theta \equiv 1/\lambda$. When $h = 1$, it simplifies to the exponential distribution.

The probability density function of the Erlang distribution is given below:

$$f(x; k, \lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)} \text{ for } x, \lambda \geq 0,$$

Where, $\Gamma(k)$ is the gamma function evaluated at k, the parameter k is called the shape parameter and the parameter λ is called the rate parameter.

An alternative but equivalent parameterization (Gamma distribution) uses the scale parameter μ, which is the reciprocal of the rate parameter (i.e., $\mu = 1/\lambda$):

$$f(x; k, \mu) = \frac{x^{k-1} e^{-\frac{x}{\mu}}}{\mu^k \Gamma(k)} \text{ for } x, \mu \geq 0.$$

When the scale parameter μ equals to 2, the distribution simplifies the Chi-square distribution with *2k* degrees of freedom. It can, therefore, be regarded as a generalized Chi-squared distribution for even numbers of degrees of freedom.

NOTES

Because of the factorial function in the denominator, the Erlang distribution is only defined when the parameter k is a positive integer. In fact, this distribution is sometimes called the **Erlang-*k* distribution** (for example, an Erlang-2 distribution is an Erlang distribution with $k = 2$). The Gamma distribution generalizes the Erlang distribution by allowing k to be any real number, using the Gamma function instead of the factorial function.

The Cumulative Distribution Function (CDF) of the Erlang distribution is given below:

$$F(x;k,\lambda) = \frac{\gamma(k,\lambda x)}{(k-1)!},$$

Where, $\Upsilon\,()$ is the lower incomplete gamma function. The CDF may also be expressed as follows:

$$F(x;k,\lambda) = 1 - \sum_{n=0}^{k-1} \frac{1}{n!} e^{-\lambda x} (\lambda x)^n.$$

An asymptotic expansion is known for the median of an Erlang distribution, for which coefficients can be computed and bounds are known.

Generating Erlang Distributed Random Numbers

Erlang distributed random numbers can be generated from uniform distribution random numbers ($U \in (0, 1)$) using the following formula:

$$E(k,\lambda) \approx -\frac{1}{\lambda} \ln \prod_{i=1}^{k} U_i$$

Waiting Times

Events that occur independently with some average rate are modeled with a Poisson process. The waiting times between k occurrences of the event are Erlang distributed.

The Erlang distribution, which measures the time between incoming calls, can be used in conjunction with the expected duration of incoming calls to produce information about the traffic load measured in Erlang units. This can be used to determine the probability of packet loss or delay, according to various assumptions made about whether blocked calls are aborted (Erlang B formula) or queued until served (Erlang C formula). The Erlang B and C formulae are still in everyday use for traffic modelling for applications, such as the design of call centers.

A.K. Erlang worked a lot in traffic modelling. Thus, there are two other Erlang distributions which used in modelling traffic. They are given below:

- **Erlang B Distribution:** This is the easier of the two distributions and can be used in a call centre to calculate the number of trunks one need to carry a certain amount of phone traffic with a certain 'target service'.

- **Erlang C Distribution:** This formula is much more difficult and is often used to calculate how long callers will have to wait before being connected to a human in a call centre or similar situation.

Stochastic Processes

NOTES

The Erlang distribution is the distribution of the sum of k independent and identically distributed random variables each having an exponential distribution. The long-run rate at which events occur is the reciprocal of the expectation of X, that is, l/k. The (age specific event) rate of the Erlang distribution is, for $k > 1$, monotonic in x, increasing from zero at x = 0, to l as x tends to infinity.

Check Your Progress

1. Define Erlang-k distribution.
2. What is Erlang C distribution?

7.3 CHI-SQUARE DISTRIBUTION

Chi-square test is a non-parametric test of statistical significance for bivariate tabular analysis (also known as cross-breaks). Any appropriate test of statistical significance lets you know the degree of confidence you can have in accepting or rejecting a hypothesis. Typically, the Chi-square test is any statistical hypothesis test, in which the test statistics has a chi-square distribution when the null hypothesis is true. It is performed on different samples (of people) who are different enough in some characteristic or aspect of their behaviour that we can generalize from the samples selected. The population from which our samples are drawn should also be different in the behaviour or characteristic. Amongst the several tests used in statistics for judging the significance of the sampling data, Chi-square test, developed by Prof. Fisher, is considered as an important test. Chi-square, symbolically written as χ^2 (pronounced as Ki-square), is a statistical measure with the help of which, it is possible to assess the significance of the difference between the observed frequencies and the expected frequencies obtained from some hypothetical universe. Chi-square tests enable us to test whether more than two population proportions can be considered equal. In order that Chi-square test may be applicable, both the frequencies must be grouped in the same way and the theoretical distribution must be adjusted to give the same total frequency which is equal to that of observed frequencies. c^2 is calculated with the help of the following formula:

$$\chi^2 = \sum \left\{ \frac{(f_0 - f_e)^2}{f_e} \right\}$$

Where, f_0 means the observed frequency; and

f_e means the expected frequency.

NOTES

Whether or not a calculated value of χ^2 is significant, it can be ascertained by looking at the tabulated values of χ^2 (given at the end of this book in appendix part) for given degrees of freedom at a certain level of confidence (generally a 5% level is taken). If the calculated value of χ^2 exceeds the table value, the difference between the observed and expected frequencies is taken as significant but if the table value is more than the calculated value of χ^2, then the difference between the observed and expected frequencies is considered as insignificant, i.e., considered to have arisen as a result of chance and as such can be ignored.

Degrees of Freedom

As already stated in the earlier unit, the number of independent constraints determines the number of degrees of freedom[2] (or *df*). If there are 10 frequency classes and there is one independent constraint, then there are (10 – 1) = 9 degrees of freedom. Thus, if n is the number of groups and one constraint is placed by making the totals of observed and expected frequencies equal, $df = (n - 1)$; when two constraints are placed by making the totals as well as the arithmetic means equal then $df = (n - 2)$ and so on. In the case of a contingency table (i.e., a table with two columns and more than two rows or table with two rows but more than two columns or a table with more than two rows and more than two columns) or in the case of a 2 × 2 table the degrees of freedom is worked out as follows:

$$df = (c - 1)(r - 1)$$

Where, c = Number of columns

r = Number of rows

Conditions for the Application of Test

The following conditions should be satisfied before the test can be applied:

(*i*) Observations recorded and used are collected on a random basis.

(*ii*) All the members (or items) in the sample must be independent.

(*iii*) No group should contain very few items say less than 10. In cases where the frequencies are less than 10, regrouping is done by combining the frequencies of adjoining groups so that the new frequencies become greater than 10. Some statisticians take this number as 5, but 10 is regarded as better by most of the statisticians.

(*iv*) The overall number of items (i.e., N) must be reasonably large. It should at least be 50, howsoever small the number of groups may be.

(*v*) The constraints must be linear. Constraints which involve linear equations in the cell frequencies of a contingency table (i.e., equations containing no squares or higher powers of the frequencies) are known as linear constraints.

NOTES

Areas of Application of Chi-Square Test

Chi-square test is applicable in large number of problems. The test is, in fact, a technique through the use of which it is possible for us to (a) Test the goodness of fit; (b) Test the homogeneity of a number of frequency distributions; and (c) Test the significance of association between two attributes. In other words, Chi-square test is a test of independence, goodness of fit and homogeneity. At times Chi-square test is used as a test of population variance also.

As a Test of Goodness of Fit, χ^2 test enables us to see how well the distribution of observe data fits the assumed theoretical distribution such as Binomial distribution, Poisson distribution or the Normal distribution.

As a Test of Independence, χ^2 test helps explain whether or not two attributes are associated. For instance, we may be interested in knowing whether a new medicine is effective in controlling fever or not and χ^2 test will help us in deciding this issue. In such a situation, we proceed on the null hypothesis that the two attributes (viz., new medicine and control of fever) are independent. Which means that new medicine is not effective in controlling fever. It may, however, be stated here that χ^2 is not a measure of the degree of relationship or the form of relationship between two attributes but it simply is a technique of judging the significance of such association or relationship between two attributes.

As a Test of Homogeneity, χ^2 test helps us in stating whether different samples come from the same universe. Through this test, we can also explain whether the results worked out on the basis of sample/samples are in conformity with well defined hypothesis or the results fail to support the given hypothesis. As such the test can be taken as an important decision-making technique.

As a Test of Population Variance. Chi-square is also used to test the significance of population variance through confidence intervals, specially in case of small samples.

Steps Involved in Finding the Value of Chi-Square

The various steps involved are as follows:

(*i*) First of all calculate the expected frequencies.

(*ii*) Obtain the difference between observed and expected frequencies and find out the squares of these differences, i.e., calculate $(f_0 - f_e)^2$.

(*iii*) Divide the quantity $(f_0 - f_e)^2$ obtained, as stated above by the corresponding expected frequency to get $\frac{(f_0 - f_e)^2}{f_e}$.

NOTES

(*iv*) Then find summation of $\frac{(f_0 - f_e)^2}{f_e}$ values or what we call

$\sum \left\{ \frac{(f_0 - f_e)^2}{f_e} \right\}$ This is the required χ^2 value.

The χ^2 value obtained as such should be compared with relevant table value of χ^2 and inference may be drawn as stated above.

The following examples illustrate the use of Chi-square test.

Example 1: A dice is thrown 132 times with the following results:

Number Turned Up	1	2	3	4	5	6
Frequency	16	20	25	14	29	28

Test the hypothesis that the dice is unbiased.

Solution: Let us take the hypothesis that the dice is unbiased. If that is so, the probability of obtaining any one of the six numbers is 1/6 and as such the expected frequency of any one number coming upward is $132 \times \frac{1}{6} = 22.$ Now, we can write the observed frequencies along with expected frequencies and work out the value of χ^2 as follows:

No. Turned Up	*Observed Frequency (or f_0)*	*Expected Frequency (or f_e)*	$(f_0 - f_e)$	$(f_0 - f_e)^2$	$\frac{(f_0 - f_e)^2}{f_e}$
1	16	22	–6	36	36/22
2	20	22	–2	4	4/22
3	25	22	3	9	9/22
4	14	22	–8	64	64/22
5	29	22	7	49	49/22
6	28	22	6	36	36/22

$$\therefore \sum \left\{ \frac{(f_0 - f_e)^2}{f_e} \right\} = 9$$

Hence, the calculated value of $\chi^2 = 9$

$\because$ Degrees of freedom in the given problem is $(n - 1) = (6 - 1) = 5$

The table value[3] of χ^2 for 5 degrees of freedom at 5% level of significance is 11.071. If we compare the calculated and table values of χ^2 we find that calculated value is less than the table value and as such could have arisen due to fluctuations of sampling. The result thus supports the hypothesis and it can be concluded that the dice is unbiased.

Example 2:

Find the value of χ^2 for the following information:

Class Observed	A	B	C	D	E
Frequency	8	29	44	15	4
Theoretical (or Expected) Frequency	7	24	38	24	7

Solution:

Since some of the frequencies are less than 10, we shall first regroup the given data as follows and then work out the value of χ^2:

Class	*Observed Frequency* (f_0)	*Expected Frequency* (f_e)	$(f_0 - f_e)$	$\frac{(f_0 - f_e)^2}{f_e}$
A and B	(8+29) = 37	(7+24) = 31	6	36/31
C	44	38	6	36/38
D and E	(15+4) = 19	(24+7) = 31	–12	144/31

$$\therefore\ \chi^2 = \sum\left\{\frac{(f_0 - f_e)^2}{f_e}\right\} = 6.76 \text{ approx.}$$

Example 3:

Two research workers classified some people in income groups on the basis of sampling studies. Their results are as follows:

	Income Groups			
Investigators	*Poor*	*Middle*	*Rich*	*Total*
A	160	30	10	200
B	140	120	40	300
Total	300	150	50	500

Show that the sampling technique of at least one research worker is defective.

Solution:

Let us take the hypothesis that the sampling techniques adopted by the research workers are similar (i.e., there is no difference between the techniques adopted by the research workers). This being so, the expectation of *A* investigator classifying the people in,

(*i*) Poor income group $= \dfrac{200 \times 300}{500} = 120$

(*ii*) Middle income group $= \dfrac{200 \times 150}{500} = 60$

(*iii*) Rich income group $= \dfrac{200 \times 50}{500} = 20$

Similarly, the expectation of *B* investigator classifying the people in

NOTES

(*i*) Poor income group $= \dfrac{300 \times 300}{500} = 180$

(*ii*) Middle income group $= \dfrac{300 \times 150}{500} = 90$

(*iii*) Rich income group $= \dfrac{300 \times 50}{500} = 30$

We can now calculate value of χ^2 as follows:

Groups	*Observed Frequency* (f_0)	*Expected Frequency* (f_e)	($f_0 - f_e$)	$(f_0 - f_e)^2/f_e$
Investigator *A*				
Classifies people as poor	160	120	40	1600/120 = 13.33
Classifies people as middle class	30	60	–30	900/60 = 15.00
Classifies people as rich	10	20	–10	100/20 = 5.00
Investigator *B*				
Classifies people as poor	140	180	–40	1600/180 = 8.88
Classifies people as middle class	120	90	30	900/90 = 10.00
Classifies people as rich	40	30	10	100/30 = 3.33

$$\because \ \chi^2 = \sum\left\{\frac{(f_0 - f_e)^2}{f_e}\right\} = 55.54$$

$$\because \text{ Degrees of freedom} = (c - 1)(r - 1)$$
$$= (3 - 1)(2 - 1) = 2$$

The table value of χ^2 for two degrees of freedom at 5% level of significance is 5.991. The calculated value of χ^2 is much higher than this table value which means that the calculated value cannot be said to have arisen just because of chance. It is significant. Hence, the hypothesis does not hold good. This means that the sampling techniques adopted by the two investigators differ and are not similar. Naturally, then the technique of one must be superior than that of the other.

Alternative Formula for Finding the Value of Chi-Square in a (2 × 2) Table

There is an alternative method of calculating the value of χ^2 in the case of a (2 × 2) table. Let us write the cell frequencies and marginal totals in case of a (2 × 2) table as follows:

NOTES

a	b	$(a + b)$
c	d	$(c + d)$
$(a + c)$	$(b + d)$	N

Then the formula for calculating the value of χ^2 will be stated as follows:

$$\chi^2 = \frac{(ad - bc)^2 N}{(a+c)(b+d)(a+b)(c+d)}$$

Where, N means the total frequency, ad means the larger cross product, bc means the smaller cross product and $(a + c)$, $(b + d)$, $(a + b)$ and $(c + d)$ are the marginal totals. *The alternative* formula is rarely used in finding out the value of Chi-square as it is not applicable uniformly in all cases but can be used only in a (2 × 2) contingency table.

Yates' Correction

F. Yates has suggested a correction in χ^2 value calculated in connection with a (2 × 2) table particularly when cell frequencies are small (since no cell frequency should be less than 5 in any case, though 10 is better as stated earlier) and χ^2 is just on the significance level. The correction suggested by Yates is popularly known as Yates' correction. It involves the reduction of the deviation of observed, from expected frequencies which of course reduces the value of χ^2. The rule for correction is to adjust the observed frequency in each cell of a (2 × 2) table in such a way as to reduce the deviation of the observed from the expected frequency for that cell by 0.5, and this adjustment is made in all the cells without disturbing the marginal totals. The formula for finding the value of χ^2 after applying Yates' correction is written as under:

$$\chi^2 \text{ (corrected)} = \frac{N.(ad - bc - 0.5N)^2}{(a+b)(c+d)(a+c)(b+d)}$$

In case we use the usual formula for calculating the value of Chi-square viz., $\left(\chi^2 = \sum \left\{\frac{(f_0 - f_e)^2}{f_e}\right\}\right)$ then Yates' correction can be applied as under:

$$\chi^2\text{(corrected)} = \frac{\left[|f_{01} - f_{e1}| - 0.5\right]^2}{f_{e1}} + \frac{\left[|f_{02} - f_{e2}| - 0.5\right]^2}{f_{e2}} + \cdots$$

It may again be emphasized that Yates' correction is made only in case of (2 × 2) table and that too when cell frequencies are small.

Chi-Square as a Test of Population Variance

χ^2 is used, at times, to test the significance of population variance $(\sigma_p)^2$ through confidence intervals. This, in other words, means that we can use χ^2 test to judge if a random sample has been drawn from a normal population with mean (μ) and with specified variance $(\sigma_p)^2$. In such a situation, the test statistic for a null hypothesis will be as under:

NOTES

$$\chi^2 = \Sigma \frac{(X_i - \bar{X}_s)^2}{(\sigma_p)^2} = \frac{n(\sigma_s)^2}{(\sigma_p)^2} \text{ with } (n\text{–}1) \text{ degrees of freedom.}$$

By comparing the calculated value (with the help of the above formula) with the table value of χ^2 for (n–1) *df* at a certain level of significance, we may accept or reject the null hypothesis. If the calculated value is equal or less than the table value, the null hypothesis is to be accepted but if the calculated value is greater than the table value, the hypothesis is rejected. All this can be made clear by an example.

Example 4:

Weight of 10 students is as follows:

Sl. No.	1	2	3	4	5	6	7	8	9	10
Weight in kg.	38	40	45	53	47	43	55	48	52	49

Can we say that the variance of the distribution of weights of all students from which the above sample of 10 students was drawn is equal to 20 square kg? Test this at 5% and 1% level of significance.

Solution:

First of all, we should work out the standard deviation of the sample (σ_s)

Calculation of the sample standard deviation:

Sl. No.	X_i *Weight in kg*	$X_i - \bar{X}_s$	$(X_i - \bar{X}_s)^2$
1	38	– 9	81
2	40	– 7	49
3	45	– 2	04
4	53	+ 6	36
5	47	+ 0	00
6	43	– 4	16
7	55	+ 8	64
8	48	+ 1	01
9	52	+ 5	25
10	49	+ 2	04
$n = 10$	$\Sigma X_i = 470$		$\Sigma (X_i - \bar{X}_s)^2 = 280$

$$\bar{X}_s = \frac{\Sigma X_i}{n} = \frac{470}{10} = 47 \text{ kg}$$

$$\therefore \quad \sigma_s = \sqrt{\frac{\Sigma(X_i - \bar{X}_s)^2}{n}} = \sqrt{\frac{280}{10}} = \sqrt{28} = 5.3 \text{ kg}$$

$\therefore\ (\sigma_s)^2 = 28$

Taking the null hypothesis as H_0: $(\sigma_p)^2 = (\sigma_s)^2$

The test statistic $\chi^2 = \dfrac{n(\sigma_s)^2}{(\sigma_p)^2} = \dfrac{10 \times 28}{20} = \dfrac{280}{20} = 14$

Degrees of freedom in this case is $(n - 1) = 10 - 1 = 9$

At 5% level of significance, the table value of $\chi^2 = 16.92$, and at 1% level of significance it is 21.67 for 9 *df*, and both these values are greater than the calculated value of χ^2 which is 14. Hence, we accept the null hypothesis and conclude that the variance of the given distribution can be taken as 20 square kg at 5% as well as at 1% level of significance.

Additive Property of Chi-Square (χ^2)

An important property of χ^2 is its additive nature. This means that several values of χ^2 can be added together and if the degrees of freedom are also added, this number gives the degrees of freedom of the total value of χ^2. Thus, if a number of χ^2 values have been obtained from a number of samples of similar data, then, because of the additive nature of χ^2, we can combine the various values of χ^2 by just simply adding them. Such addition of various values of χ^2 gives one value of χ^2 which helps in forming a better idea about the significance of the problem under consideration. The following example illustrates the additive property of the χ^2.

Example 5: The following values of χ^2 are obtained from different investigations carried to examine the effectiveness of a recently invented medicine for checking malaria.

Investigation	χ^2	*df*
1	2.5	1
2	3.2	1
3	4.1	1
4	3.7	1
5	4.5	1

What conclusion would you draw about the effectiveness of the new medicine on the basis of the five investigations taken together?

Solution: By adding all the values of χ^2, we obtain a value equal to 18.0. Also by adding the various d.f. as given in the question, we obtain a figure 5. We can now state that the value of χ^2 for 5 degrees of freedom (when all the five investigations are taken together) is 18.0.

NOTES

Let us take the hypothesis that the new medicine is not effective. The table value of χ^2 for 5 degrees of freedom at 5% level of significance is 11.070. But our calculated value is higher than this table value which means that the difference is significant and is not due to chance. As such the hypothesis is wrong and it can be concluded that the new medicine is effective in checking malaria.

Important Characteristics of Chi-Square (χ^2) Test

(*i*) This test is based on frequencies and not on the parameters like mean and standard deviation.

(*ii*) This test is used for testing the hypothesis and is not useful for estimation.

(*iii*) This test possesses the additive property.

(*iv*) This test can also be applied to a complex contingency table with several classes and as such is a very useful test in research work.

(*v*) This test is an important non-parametric (or a distribution free) test as no rigid assumptions are necessary in regard to the type of population and no need of the parameter values. It involves less mathematical details.

A Word of Caution in Using χ^2 Test

Chi-square test is no doubt a most frequently used test but its correct application is equally an uphill task. It should be borne in mind that the test is to be applied only when the individual observations of sample are independent which means that the occurrence of one individual observation (event) has no effect upon the occurrence of any other observation (event) in the sample under consideration. The researcher, while applying this test, must remain careful about all these things and must thoroughly understand the rationale of this important test before using it and drawing inferences concerning his hypothesis.

Check Your Progress

3. What is a chi-square test?
4. What do you mean by degrees of freedom?
5. What conditions should be satisfied before the application of Chi-square test?
6. What are the areas of application in which Chi-square test is applied?
7. What do you mean by goodness of fit?
8. What are the steps involved in Chi-square test?
9. Is there any alternative formula to find the value of Chi-square?
10. What is Yates' correction method?
11. Explain the additive property of Chi-square.
12. What are the important characteristics of Chi-square test?

7.4 THE NORMAL DISTRIBUTION

Among all the probability distributions the normal probability distribution is by far the most important and frequently used continuous probability distribution. This is so because this distribution well fits in many types of problems. This distribution is of special significance in inferential statistics since it describes probabilistically the link between a statistic and a parameter (i.e., between the sample results and the population from which the sample is drawn). The name of Karl Gauss, eighteenth century mathematician-astronomer, is associated with this distribution and in honour of his contribution, this distribution is often known as the Gaussian distribution.

The normal distribution can be theoretically derived as the limiting form of many discrete distributions. For instance, if in the binomial expansion of $(p + q)^n$, the value of 'n' is infinity and $p = q = \frac{1}{2}$, then a perfectly smooth symmetrical curve would be obtained. Even if the values of p and q are not equal but if the value of the exponent 'n' happens to be very very large, we get a curve of normal probability smooth and symmetrical. Such curves are called normal probability curves (or at times known as normal curves of error) and such curves represent the normal distributions.[5]

The probability function in case of normal probability distribution[6] is given as:

$$f(x) = \frac{1}{\sigma.\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Where, μ = The mean of the distribution.

σ^2 = Variance of the distribution.

The normal distribution is thus defined by two parameters viz., μ and σ^2. This distribution can be represented graphically as under:

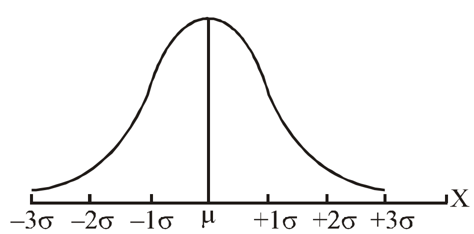


***Fig. 7.1** Curve Representing Normal Distribution*

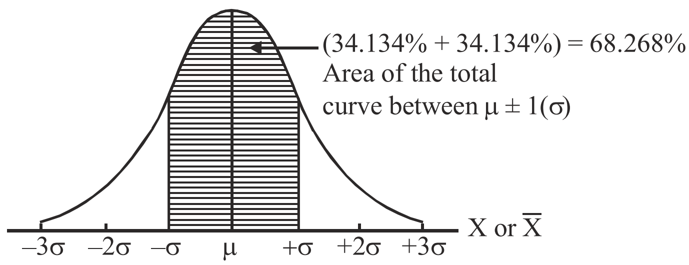
Characteristics of Normal Distribution

The characteristics of the normal distribution or that of normal curve are, as given below:

1. It is symmetric distribution.[7]

NOTES

2. The mean μ defines where the peak of the curve occurs. In other words, the ordinate at the mean is the highest ordinate. The height of the ordinate at a distance of one standard deviation from mean is 60.653% of the height of the mean ordinate and similarly the height of other ordinates at various standard deviations (σ_s) from mean happens to be a fixed relationship with the height of the mean ordinate.
3. The curve is asymptotic to the base line which means that it continues to approach but never touches the horizontal axis.
4. The variance (σ^2) defines the spread of the curve.
5. Area enclosed between mean ordinate and an ordinate at a distance of one standard deviation from the mean is always 34.134% of the total area of the curve. It means that the area enclosed between two ordinates at one sigma (S.D.) distance from the mean on either side would always be 68.268% of the total area. This can be shown as follows:



Similarly, the other area relationships are as follows:

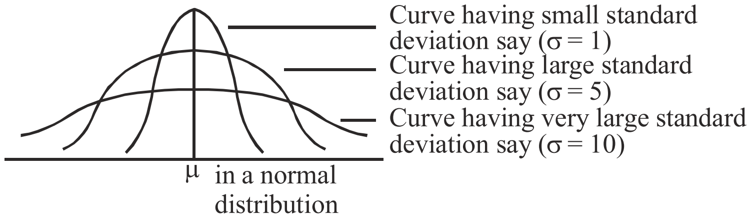
Between		*Area Covered to Total Area of the Normal Curve*[8]
$\mu \pm 1$	S.D.	68.27%
$\mu \pm 2$	S.D.	95.45%
$\mu \pm 3$	S.D.	99.73%
$\mu \pm 1.96$	S.D.	95%
$\mu \pm 2.578$	S.D.	99%
$\mu \pm 0.6745$	S.D.	50%

6. The normal distribution has only one mode since the curve has a single peak. In other words, it is always a unimodal distribution.
7. The maximum ordinate divides the graph of normal curve into two equal parts.
8. In addition to all the above stated characteristics the curve has the following properties:
   (a) $\mu = \bar{x}$
   (b) $\mu_2 = \sigma^2 = \text{Variance}$
   (c) $\mu_4 = 3\sigma^4$
   (d) Moment Coefficient of Kurtosis = 3

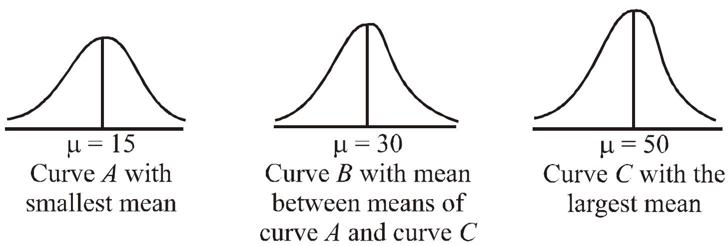
Family of Normal Distributions

We can have several normal probability distributions but each particular normal distribution is being defined by its two parameters viz., the mean (μ) and the standard deviation (σ). There is, thus, not a single normal curve but rather a family of normal curves. We can exhibit some of these as under:

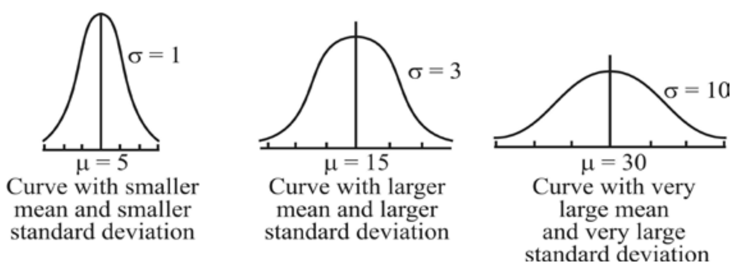
Normal curves with identical means but different standard deviations:



Normal curves with identical standard deviation but each with different means:



Normal curves each with different standard deviations and different means:



How to Measure the Area under the Normal Curve?

We have stated above some of the area relationships involving certain intervals of standard deviations (plus and minus) from the means that are true in case of a normal curve. But what should be done in all other cases? We can make use of the statistical tables constructed by mathematicians for the purpose. Using these tables we can find the area (or probability, taking the entire area of the curve as equal to 1) that the normally distributed random variable will lie within certain distances from the mean. These distances are defined in terms of standard deviations. While using the tables showing the area under the normal curve we talk in terms of standard variate (symbolically Z) which really means standard deviations without units of measurement and this 'Z' is worked out as under:

$$Z = \frac{X - \mu}{\sigma}$$

Where, Z = The standard variate (or number of standard deviations from X to the mean of the distribution).

NOTES

X = Value of the random variable under consideration.

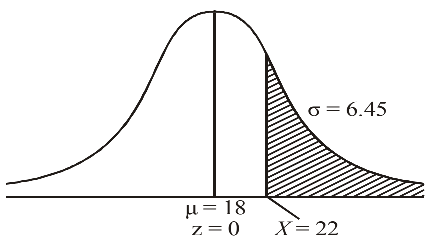
μ = Mean of the distribution of the random variable.

σ = Standard deviation of the distribution.

The table showing the area under the normal curve (often termed as the standard normal probability distribution table) is organized in terms of standard variate (or Z) values. It gives the values for only half the area under the normal curve, beginning with $Z = 0$ at the mean. Since the normal distribution is perfectly symmetrical the values true for one half of the curve are also true for the other half. We now illustrate the use of such a table for working out certain problems.

Example 6: A banker claims that the life of a regular saving account opened with his bank averages 18 months with a standard deviation of 6.45 months. Answer the following: (a) What is the probability that there will still be money in 22 months in a savings account opened with the said bank by a depositor? (b) What is the probability that the account will have been closed before two years?

Solution: (a) For finding the required probability we are interested in the area of the portion of the normal curve as shaded and shown below:



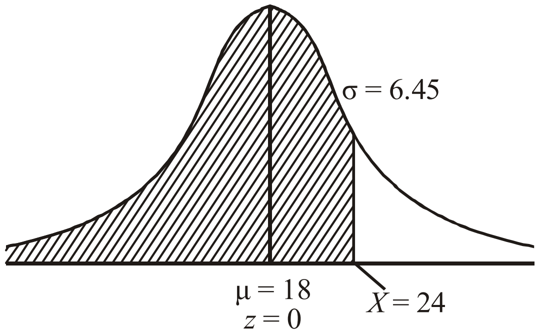
Let us calculate Z as under:

$$Z = \frac{X - \mu}{\sigma} = \frac{22 - 18}{6.45} = 0.62$$

The value from the table showing the area under the normal curve for $Z = 0.62$ is 0.2324. This means that the area of the curve between $\mu = 18$ and $X = 22$ is 0.2324. Hence, the area of the shaded portion of the curve is (0.5) – (0.2324) = 0.2676 since the area of the entire right hand portion of the curve always happens to be 0.5. Thus the probability that there will still be money in 22 months in a savings account is 0.2676.

(b) For finding the required probability we are interested in the area of the portion of the normal curve as shaded and shown in figure:

NOTES



For the purpose we calculate,

$$Z = \frac{24-18}{6.45} = 0.93$$

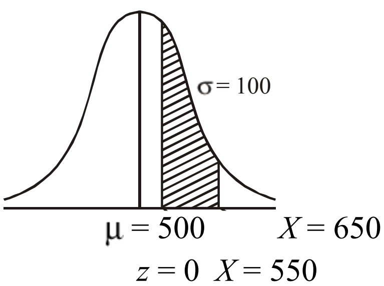
The value from the concerning table, when $Z = 0.93$, is 0.3238 which refers to the area of the curve between $\mu = 18$ and $X = 24$. The area of the entire left hand portion of the curve is 0.5 as usual.

Hence, the area of the shaded portion is (0.5) + (0.3238) = 0.8238 which is the required probability that the account will have been closed before two years, i.e., before 24 months.

Example 7: Regarding a certain normal distribution concerning the income of the individuals we are given that mean=500 rupees and standard deviation=100 rupees. Find the probability that an individual selected at random will belong to income group,

(a) Rs 550 to Rs 650 (b) Rs 420 to 570

Solution: (a) For finding the required probability we are interested in the area of the portion of the normal curve as shaded and shown below:



For finding the area of the curve between $X = 550$ to 650, let us do the following calculations:

$$Z = \frac{550-500}{100} = \frac{50}{100} = 0.50$$

Corresponding to which the area between $\mu = 500$ and $X = 550$ in the curve as per table is equal to 0.1915 and,

$$Z = \frac{650 - 500}{100} = \frac{150}{100} = 1.5$$

NOTES

Corresponding to which, the area between μ = 500 and X = 650 in the curve, as per table, is equal to 0.4332.

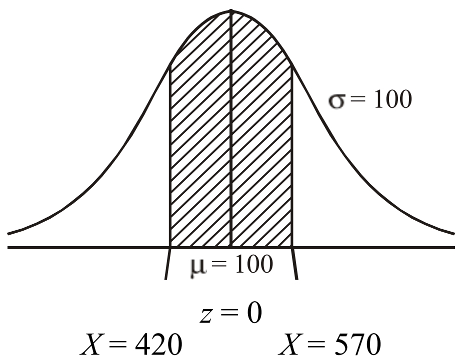
Hence, the area of the curve that lies between X = 550 and X = 650 is,

$$(0.4332) - (0.1915) = 0.2417$$

This is the required probability that an individual selected at random will belong to income group of Rs 550 to Rs 650.

(b) For finding the required probability we are interested in the area of the portion of the normal curve as shaded and shown below:

To find the area of the shaded portion we make the following calculations:



$$Z = \frac{570 - 500}{100} = 0.70$$

Corresponding to which the area between μ = 500 and X = 570 in the curve as per table is equal to 0.2580.

And $Z = \frac{420 - 500}{100} = -0.80$

Corresponding to which the area between μ = 500 and X = 420 in the curve as per table is equal to 0.2881.

Hence, the required area in the curve between X = 420 and X = 570 is,

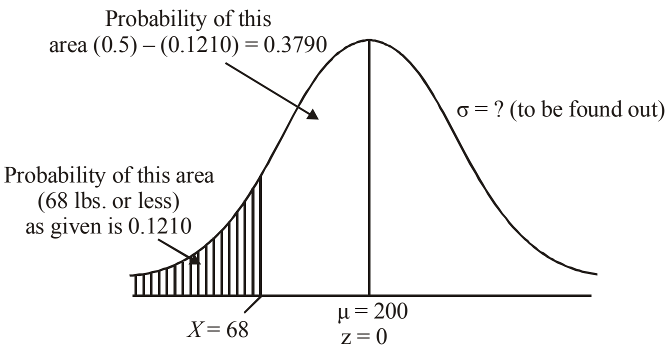
$$(0.2580) + (0.2881) = 0.5461$$

This is the required probability that an individual selected at random will belong to income group of Rs 420 to Rs 570.

Example 8: A certain company manufactures $1\frac{1''}{2}$ all-purpose rope made from imported hemp. The manager of the company knows that the average load-bearing capacity of the rope is 200 lbs. Assuming that normal distribution applies, find the standard deviation of load-bearing capacity for the $1\frac{1''}{2}$ rope if it is given that the rope has a 0.1210 probability of breaking with 68 lbs. or less pull.

NOTES

Solution: Given information can be depicted in a normal curve as shown below:



If the probability of the area falling within $\mu = 200$ and $X = 68$ is 0.3790 as stated above, the corresponding value of Z as per the table[9] showing the area of the normal curve is – 1.17 (minus sign indicates that we are in the left portion of the curve)

Now to find σ, we can write,

$$Z = \frac{X - \mu}{\sigma}$$

Or $\quad -1.17 = \dfrac{68 - 200}{\sigma}$

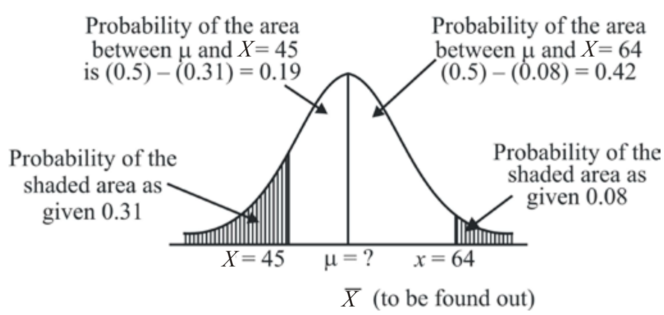
Or $\quad -1.17\sigma = -132$

Or $\quad \sigma = 112.8$ lbs. approx.

Thus, the required standard deviation is 112.8 lbs. approximately.

Example 9: In a normal distribution, 31 per cent items are below 45 and 8 per cent are above 64. Find the $\overline{X}$ and σ of this distribution.

Solution: We can depict the given information in a normal curve as shown below:



If the probability of the area falling within μ and $X = 45$ is 0.19 as stated above, the corresponding value of Z from the table showing the area of the normal curve is – 0.50. Since, we are in the left portion of the curve, we can express this as under,

$$-0.50 = \frac{45-\mu}{\sigma} \qquad (1)$$

Similarly, if the probability of the area falling within μ and $X = 64$ is 0.42, as stated above, the corresponding value of Z from the area table is, +1.41. Since, we are in the right portion of the curve we can express this as under,

NOTES

$$1.41 = \frac{64-\mu}{\sigma} \qquad (2)$$

If we solve Equations (1) and (2) above to obtain the value of μ or $\bar{X}$, we have,

$$-0.5\,\sigma = 45-\mu \qquad (3)$$

$$1.41\,\sigma = 64-\mu \qquad (4)$$

By subtracting the Equation (4) from Equation (3) we have,

$$-1.91\,\sigma = -19$$

$$\therefore \quad \sigma = 10$$

Putting $\sigma = 10$ in Equation (3) we have,

$$-5 = 45-\mu$$

$$\therefore \quad \mu = 50$$

Hence, $\bar{X}$ (or μ)=50 and σ=10 for the concerning normal distribution.

7.5 THE BIVARIATE NORMAL DISTRIBUTION

Binomial, Poisson, negative binomial and uniform distribution are some of the discrete probability distributions. The random variables in these distributions assume a finite or enumerably infinite number of values but in nature these are random variables which take infinite number of values i.e. these variables can take any value in an interval. Such variables and their probability distributions are known as continuous probability distributions.

A random variable X is the said to be normally distributed if it has the following probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ for } -\infty \le x \le \infty$$

where μ and σ > 0 are the parameters of distribution.

Normal Curve: A curve given by,

$$y_x = y_0 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Which is known as the normal curve when origin is taken at mean.

Then $$y_x = y_0 e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}.$$

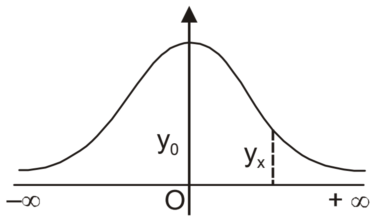


Fig. 7.2 *Normal Curve*

NOTES

Standard Normal Variate : A normal variate with mean zero and standard deviation unity, is called a standard normal variate.

That is; if X is a standard normal variate then $E(X) = 0$ and $V(X) = 1$.

Then, $X \sim N(0, 1)$

The moment generating function or MGF of a standard normal variate is given as follows:

$$M_X(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2}\Bigg]_{\substack{\mu = 0 \\ \sigma = 1}} = e^{\frac{1}{2}t^2}$$

Frequently the exchange of variable in the integral:

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}$$

is used by introducing the following new variable:

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

This new random variable Z simplifies calculations of probabilities etc. concerning normally distributed variates.

Standard Normal Distribution: The distribution of a random variable $Z = \dfrac{X-\mu}{\sigma}$ which is known as standard normal variate, is called the standard normal distribution or unit normal distribution, where X has a normal distribution with mean μ end variance σ^2.

The density function of Z is given as follows:

$$\phi(Z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}Z^2}, -\infty < Z < \infty$$

NOTES

with mean O variance one of MGF $e^{\frac{1}{2}t^2}$. Normal distribution is the most frequently used distribution in statistics. The importance of this distribution is highlighted by central limit theorem, mathematical properties, such as the calculation of height, weight, the blood pressure of normal individuals, heart diameter measurement, etc. They all follow normal distribution if the number of observations is very large. Normal distribution also has great importance in statistical inference theory.

Examples of Normal Distribution:

1. The height of men of matured age belonging to same race and living in similar environments provide a normal frequency distribution.
2. The heights of trees of the same variety and age in the same locality would confirm to the laws of normal curve.
3. The length of leaves of a tree form a normal frequency distribution. Though some of them are very short and some are long, yet they try to tend towards their mean length.

Example 10: X has normal distribution with $\mu = 50$ and $\sigma^2 = 25$. Find out

(i) The approximate value of the probability density function for $X = 50$

(ii) The value of the distribution function for $x = 50$.

Solution: (*i*) $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, -\infty \le x \le \infty.$$

for $X = 50$, $\sigma^2 = 25$, $\mu = 50$, you have

$$f(x) = \frac{1}{5\sqrt{2\pi}} \equiv 0.08.$$

Distribution function $f(x)$

$$= \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu^2)/2\sigma^2} .dx$$

$$= \int_{-\infty}^{\left(\frac{x-\mu}{\sigma}\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} dZ, \text{ where } Z = \frac{x-\mu}{\sigma}$$

$$\therefore \quad F(50) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} .dZ = 0.5.$$

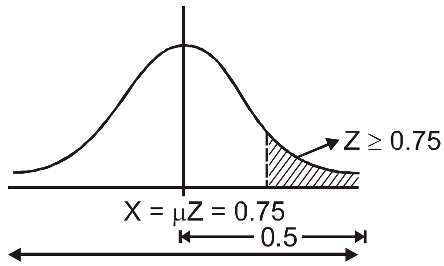
Example 11: If X is a normal variable with mean 8 and standard deviation 4, find

(i) $P[X \le 5]$

(ii) $P[5 \le X \le 10]$

NOTES

Solution: (i) $P[X \leq 5] = P\left(\frac{X-\mu}{\sigma} \leq \frac{5-8}{4}\right)$



$= P\,(Z \leq -\,0.75)$

$= P\,(Z \geq\ 0.75)$

[By Symmetry]

$= 0.5 - P(0 \leq Z \leq 0.75)$

[To use relevant table]

$= 0.5 - 0.2734$ [See Appendix for value of '2']

$= 0.2266.$

(ii) $P[5 \leq X \leq 10] = P\left(\frac{5.8}{4} \leq Z \leq \frac{10-8}{4}\right)$

$= P(-\,0.75 \leq Z \leq 0.5)$

$= P(-\,0.75 \leq Z \leq 0) + P(0 \leq Z \leq 0.5)$

$= P(-\,0 \leq Z \leq 0.75) + P(0 \leq Z \leq 0.5)$

$= 0.2734 + 0.1915$ [See Appendix]

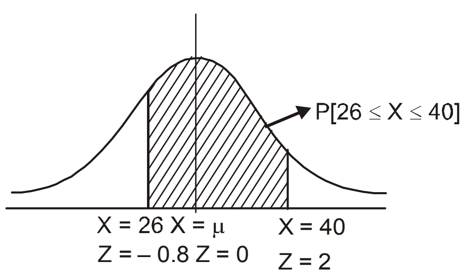
$= 0.4649.$

Example 12: X is a normal variate with mean 30 and S.D. 5. Find

(i) $P[26 \leq X \leq 40]$

(ii) $P[|X - 30| > 5]$

Solution: Here $\mu = 30$, $\sigma = 5$.



(i) When $X = 26,\ Z = \frac{X-\mu}{\sigma} = -\,0.8$

And for $X = 40,\ Z = \frac{X-\mu}{\sigma} = 2$

$\therefore\ P[26 \leq X \leq 40] = P[-\,0.8 \leq Z \leq 2]$

NOTES

$$= P[0 \le Z \le 0.8] + P[0 \le Z \le 2]$$
$$= 0.2881 + 0.4772 = 0.7653$$

(*ii*) $P[|X - 3| > 5] = 1 - P[|X - 3| \le 5]$

$$P[|X - 3| \le 5] = P[25 \le X \le 35]$$
$$= P\left(\frac{25-30}{5} \le Z \le \frac{35-30}{5}\right)$$
$$= 2.P\,(0 \le Z \le 1) = 0.$$
$$= 2 \times 0.3413 = 0.6826.$$

So $\quad P[|X - 3| > 5] = 1 - P[|X - 3| \le 5]$

$$= 1 - 0.6826 = 0.3174.$$

Central Limit Theorem

Let $X_1, X_2, ... X_n$ be n independent random variables all of which have the same distribution. Let the common expectation and variance be μ and σ_2 respectively.

Let $$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

Then, the distribution of $\bar{X}$ approches the normal distribution with mean m and variance $\frac{\sigma^2}{n}$ as $n \to \infty$

That is, the variate $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ has standard normal distribution.

Proof: Moment generating function of Z about origin is given as follows:

$$M_Z(t) = E(e^{tZ}) = E\left[e^{t\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)}\right]$$

$$= e^{-\mu t\sqrt{n}/\sigma} E(e^{t \times \sqrt{n}/\sigma})$$

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

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

NOTES

$$= e^{-\mu t \frac{\sqrt{n}}{\sigma}} . M(X_1 + X_2 + ... + X_n) . \frac{t}{\sigma \sqrt{n}}$$

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

This is because the random variables are independent and have the same MGF by using logarithms, you have:

$$\log M_z(t) = \frac{-\mu t \sqrt{n}}{\sigma} + n \log M_x \left(\frac{t}{\sigma \sqrt{n}} \right)$$

$$= \frac{-\mu t \sqrt{n}}{\sigma} + n \log \left[1 + \frac{\mu'_1 t}{\sigma \sqrt{n}} + \frac{\mu'_2 t}{2!} \left(\frac{i}{\sigma \sqrt{n}} \right)^2 + ... \right]$$

$$= \frac{-\mu t \sqrt{n}}{\sigma} + n \left[\left(\frac{\mu'_1 t}{\sigma \sqrt{n}} + \frac{\mu'_2 t}{2!} . \frac{t^2}{n \sigma^2} + ... \right) - \frac{1}{2} \left(\frac{\mu'_1 t}{\sigma \sqrt{n}} + ... \right)^2 + ... \right]$$

$$= \frac{-\mu t \sqrt{n}}{\sigma} + \frac{\mu'_1 t \sqrt{n}}{\sigma} + \frac{\mu'_2 t^2}{2\sigma^2} - \frac{\mu'^2_1 t^2}{2\sigma^2} + ...$$

$$= \frac{t^2}{2} + O(n^{-1/2}) \qquad [\because \mu_2' - \mu_1'^2 = \sigma^2 \mu_1' = \mu]$$

Hence, as $n \to \infty$

$$\log(M_z)(t) \to \frac{t^2}{2} \quad \text{i.e.} \quad M_2(t) = e^{t^2/2}$$

However, this is the M.G.F. of a standard normal random variable. Thus, the random variable Z converges to N.

This follows that the limiting distribution of $\sqrt{x}$ as normal with mean μ and variance $\frac{\sigma^2}{n}$.

Normal Approximation

Under certain circumstance one can use normal distribution to approximate a binomial distribution as well as Poisson distribution.

If $X \sim B(n, p)$ and value of n is quite large with p quite close to ½, then, X can be approximated as $N(np, npq)$, where $q = 1 - p$.

NOTES

There are cases in which use of normal distribution is found to be easier than that of a Binomial distribution.

Also, as already told, normal distribution may also be utilized for approximating Poisson distribution when value of λ is large. Here, λ is the mean of Poisson distribution.

Thus, $X \sim Po(\lambda) \rightarrow X \sim N(\lambda, \lambda)$ approximately when values of λ is large.

Continuity Correction

Normal distribution is continuous whereas both the distributions, Binomial as well as Poisson are discrete random variables. This fact has to be kept in mind while making use of normal distribution for approximating Binomial or Poisson distribution and use continuity correction.

Each probability, in case of discrete distribution, is represented using a rectangle as shown in the Figure 7.2(b).

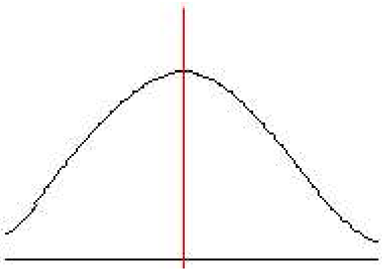


Fig. 7.2 (a) *Continuous Distribution*

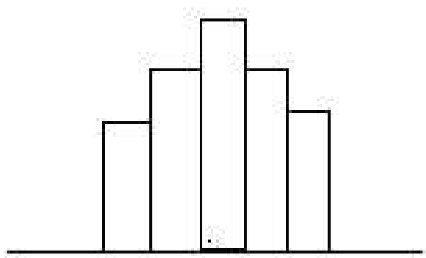


Fig. 7.3 (b) *Discrete Distribution*

While working out for probabilities, we like inclusion of the whole rectangles in applying continuity correction.

Example 13: A fair coin is tossed 20 times. Find the probability of getting between 9 and 11 heads.

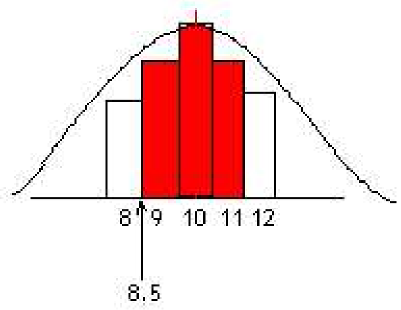
Solution: Let the random variable be represented by X that shows the number of heads thrown.

$$X \sim \text{Bin}(20, ½)$$

NOTES

As p is very near to ½, normal approximation can be used for the binomial distribution and we may write as $X \sim N(20 \times ½, 20 \times ½ \times ½) \rightarrow X \sim N(10, 5)$.

In the diagram as shown in below rectangles show binomial distribution which is discrete and curve shows normal distribution which is continuous in nature.



Using normal distribution for showing Binomial distribution

If it is desired to have P(9 ≤ X ≤ 11) as shown by shaded area one may note that first rectangle begins at 8.5 and last rectangle terminates at 11.5. By making a continuity correction, probability becomes P(8.5 < X < 11.5) in normal distribution. We may standardize this, as given below:

$$= P\left(\frac{9-10}{\sqrt{5}} < \frac{X-10}{5} < \frac{11-10}{\sqrt{5}}\right)$$

$$= \text{P}\ (-\ 0.447 < \text{Z} < 0.447)$$

$$= 2 \times 0.67 - 1 \text{ (using tables) } = 0.34$$

Check Your Progress

13. What is a normal distribution?
14. Explain any four characteristics of normal distribution.
15. Define a standard normal variate.

7.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Because of the factorial function in the denominator, the Erlang distribution is only defined when the parameter k is a positive integer. In fact, this distribution is sometimes called the Erlang-k distribution.
2. This formula is much more difficult and is often used to calculate how long callers will have to wait before being connected to a human in a call centre or similar situation.

NOTES

3. Chi-square test is a non-parametric test of statistical significance for bivariate tabular analysis. Chi-square tests enable us to test whether more than two population proportions can be considered equal.
4. The degree of freedom is determined by the number of constraints. If there are 10 frequency classes and there is one independent constraint, then there are 10 – 1 = 9 degrees of freedom.
5. Before the application of chi-square test, the following conditions need to be satisfied:
   (*i*) Observations recorded and used are collected on a random basis.
   (*ii*) All the members or items in the sample must be independent.
   (*iii*) No group should contain very few items say less than 10. In cases where the frequencies are less than 10, regrouping is done by combining the frequencies of adjoining groups so that the new frequencies become greater than 10.
   (*iv*) The overall number of items must be reasonably large; it should at least be 50 howsoever small the number of groups may be.
   (*v*) The constraints must be linear. Constraints, which involve linear equations in the cell frequencies of a contingency table, are known as linear constraints.
6. Chi-square test is applicable in large number of problems. These include:
   (*i*) Testing the goodness of fit.
   (*ii*) Testing the homogeneity of a number of frequency distributions.
   (*iii*) Testing the significance of association between two attributes.
   (*iv*) Establishing hypotheses.
   (*v*) Testing independence between two variables.
7. Goodness of fit describes that how well the theoretical distribution fits with the observed data.
8. The various steps involved in the chi-square test are:
   (*i*) Calculation of the expected frequencies.
   (*ii*) Obtaining the difference between observed and expected frequencies and finding out the squares of these differences.
   (*iii*) Dividing the quantity $(f_0 - f_e)^2$ obtained in the result by the corresponding expected frequency to get $\dfrac{(f_0 - f_e)^2}{f_e}$.
   (*iv*) Finding summation of $\dfrac{(f_0 - f_e)^2}{f_e}$ values also called as required χ^2 value, and is represented as $\sum \dfrac{(f_0 - f_e)^2}{f_e}$.

9. The alternative method of calculating the value of chi-square is used in the case of a (2×2) table. The cell frequencies and marginal totals of a (2×2) table are written as follows:

a	b	$(a + b)$
c	d	$(c + d)$
$(a + c)$	$(b + d)$	N

The alternative formula for calculating the value of χ^2 is:

$$\chi^2 = \frac{(ad - bc)^2 \cdot N}{(a + c)(b + d)(a + b)(c + d)}$$

Where, N means the total frequency, *ad* means the larger cross product, *bc* means the smaller cross product and $(a + c)$, $(b + d)$, $(a + b)$ and $(c + d)$ are the marginal totals.

10. F. Yates has suggested a correction in χ^2 value calculated in connection with a (2×2) table particularly when cell frequencies are small and χ^2 is just on the significance level. The correction suggested by Yates is popularly known as Yates' correction. It involves the reduction of the deviation of observed from expected frequencies, which of course reduces the value of χ^2. The rule for correction is to adjust the observed frequency in each cell of a (2×2) table in such a way as to reduce the deviation of the observed from the expected frequency for that cell by 0.5, and this adjustment is made in all the cells without disturbing the marginal totals.

11. It is an important property of χ^2. This means that several values of χ^2 can be added together and if the degrees of freedom are also added, this number gives the degrees of freedom of total value of χ^2. Thus, if a number of χ^2 values have been obtained from a number of samples of similar data, then because of the additive nature of χ^2 we can combine the various values of χ^2 by just simply adding them.

12. The important characteristics of chi-square test are:

    (*i*) This test is based on frequencies and not on the parameters like mean and standard deviation.

    (*ii*) This test is used for testing the hypothesis.

    (*iii*) This test possesses the additive property.

    (*iv*) This test can also be applied to a complex contingency table with several classes and as such is a very useful test in research work.

    (*v*) This test is an important non-parametric test as no rigid assumptions are necessary in regard to the type of population.. F. Yates has suggested a correction in χ^2 value calculated in connection with a (2×2) table particularly when cell frequencies are small and χ^2 is just on the significance level. The correction suggested by Yates is popularly known as Yates' correction. It involves the reduction of the deviation of observed

NOTES

from expected frequencies, which of course reduces the value of χ^2. The rule for correction is to adjust the observed frequency in each cell of a (2×2) table in such a way as to reduce the deviation of the observed from the expected frequency for that cell by 0.5, and this adjustment is made in all the cells without disturbing the marginal totals.

13. It is an important property of χ^2. This means that several values of χ^2 can be added together and if the degrees of freedom are also added, this number gives the degrees of freedom of total value of χ^2. Thus, if a number of χ^2 values have been obtained from a number of samples of similar data, then because of the additive nature of χ^2 we can combine the various values of χ^2 by just simply adding them.
14. The important characteristics of chi-square test are:
    (*i*) This test is based on frequencies and not on the parameters like mean and standard deviation.
    (*ii*) This test is used for testing the hypothesis.
    (*iii*) This test possesses the additive property.
    (*iv*) This test can also be applied to a complex contingency table with several classes and as such is a very useful test in research work.
    (*v*) This test is an important non-parametric test as no rigid assumptions are necessary in regard to the type of population.
15. A normal variate with mean zero and standard deviation unity, is called a standard normal variate.

7.7 SUMMARY

- The Erlang distribution is a continuous probability distribution with wide applicability primarily due to its relation to the Exponential and Gamma distributions.
- The Erlang distribution, which measures the time between incoming calls, can be used in conjunction with the expected duration of incoming calls to produce information about the traffic load measured in Erlang units.
- The Erlang distribution is the distribution of the sum of k independent and identically distributed random variables each having an exponential distribution.
- Chi-square test is a non-parametric test of statistical significance for bivariate tabular analysis (also known as cross-breaks). Any appropriate test of statistical significance lets you know the degree of confidence you can have in accepting or rejecting a hypothesis.
- The constraints must be linear. Constraints which involve linear equations in the cell frequencies of a contingency table (i.e., equations containing no squares or higher powers of the frequencies) are known as linear constraints.

NOTES

- χ^2 test enables us to see how well the distribution of observe data fits the assumed theoretical distribution such as Binomial distribution, Poisson distribution or the Normal distribution.
- Chi-square is also used to test the significance of population variance through confidence intervals, specially in case of small samples.
- χ^2 is used, at times, to test the significance of population variance $(\sigma_p)^2$ through confidence intervals.
- The curve is asymptotic to the base line which means that it continues to approach but never touches the horizontal axis.
- The normal distribution has only one mode since the curve has a single peak. In other words, it is always a unimodal distribution.
- The distribution of a random variable $Z = \dfrac{X - \mu}{\sigma}$ which is known as standard normal variate, is called the standard normal distribution or unit normal distribution, where X has a normal distribution with mean μ end variance σ^2.
- Under certain circumstance one can use normal distribution to approximate a binomial distribution as well as Poisson distribution.
- Normal distribution is continuous whereas both the distributions, Binomial as well as Poisson are discrete random variables. This fact has to be kept in mind while making use of normal distribution for approximating Binomial or Poisson distribution and use continuity correction.

7.8 KEY WORDS

- **Erlang B distribution:** This is the easier of the two distributions and can be used in a call centre to calculate the number of trunks one need to carry a certain amount of phone traffic with a certain 'target service'.
- **Degrees of freedom:** The number of independent constraints determines the number of degrees of freedom^2 (or *df*). If there are 10 frequency classes and there is one independent constraint, then there are $(10 - 1) = 9$ degrees of freedom.

7.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is meant by Gamma function?
2. Explain stochastic process.

NOTES

3. Explain chi-square test.
4. Why is it considered an important test in statistical analysis?
5. Describe the term 'Degrees of Freedom'.
6. Define the necessary conditions required for the application of test?
7. What are the areas of application of chi-square test?
8. How will you find the value of chi-square?
9. Define Yates' correction formula for chi-square.
10. Chi-square can be used as a test of population variance. Explain.
11. Describe the additive properties of chi-square.
12. Explain the important characteristics of chi-square test.
13. Give the characteristics of normal distribution.
14. Explain some examples of normal distribution.
15. Define central limit theorem.

Long-Answer Questions

1. Briefly discuss about the Gamma distributions.
2. What is Chi-square test? Explain its significance in statistical analysis.
3. Write short notes on the following:
   (*i*) Additive property of chi-square
   (*ii*) Chi-square as a test of 'goodness of fit'
   (*iii*) Precautions in applying chi-square test
   (*iv*) Conditions for applying chi-square test
4. On of the basis of information given below about the treatment of 200 patients suffering from a disease, state whether the new treatment is comparatively superior to the conventional treatment.

Treatment	*No. of Patients*	
	Favourable Response	*No Response*
New	60	20
Conventional	70	50

   For drawing your inference use the value of χ^2 for one degree of freedom at the 5% level of significance, viz., 3.841.

5. 200 digits were chosen at random from a set of tables. The frequencies of the digits were:

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	18	19	23	21	16	25	22	20	21	15

   Calculate χ^2.

6. The normal rate of infection for a certain disease in cattle is known to be 50%. In an experiment with seven animals injected with a new vaccine it was found that none of the animals caught infection. Can the evidence be regarded as conclusive (at 1% level of significance) to prove the value of the new vaccine?
7. Result of throwing dice were recorded as follows:

Number Falling Upwards	1	2	3	4	5	6
Frequency	27	33	31	29	30	24

   Is the dice unbiased? Answer on the basis of Chi-square test.
8. (*i*) 1000 babies were born during a certain week in a city of which 600 were boys and 400 girls. Use χ^2 test to examine the correctness of the hypothesis that the sex ratio is 1:1 in newly born babies.

   (*ii*) The percentage of smokers in a certain city was 90. A random sample of 100 persons was selected in which 85 persons were found to be smokers. Is the sample proportion significantly different from the proportion of smokers in the city? Answer on the basis of chi-square test.
9. How to measure the area under the normal curve? Give examples also.
10. Discuss about the standard normal distribution.

7.10 FURTHER READINGS

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 8 DISTRIBUTION OF FUNCTIONS OF RANDOM VARIABLE AND SAMPLING

Structure

8.0 INTRODUCTION

One of the major objectives of the field of statistical analysis is to know the true or actual values of different parameters of population. The ideal situation would be to take the entire population into consideration in determining these values. However, that is not feasible due to cost, time, labour and other constraints. Accordingly, random samples of a given size are taken from the population and these samples are properly analysed with the belief that the characteristics of these random samples represent similar characteristics of the population from which these samples have been taken. The results obtained from such analyses lead to generalizations that are considered to be valid for the entire population.

You will study two broad types of sampling: probability samples and non-probability samples. Probability samples involve simple random sampling and restricted random sampling. Non-probability samples are characterized by non-random sampling. As simple random sampling is costly and time consuming, restricted random sampling is preferred. Sampling distribution of mean is a probability distribution of all possible sample means of a given size selected from a population, while sampling distribution of proportions is a distribution of proportions of all possible random samples of a fixed size.

In this unit, you will study about the distribution of functions of random variable, sampling theory and transformation of variable of the discrete type.

8.1 OBJECTIVES

After going through this unit, you will be able to:

- Analyse the distribution of functions of random variable
- Understand the various sampling theories
- Explain the transformation of variable of the discrete type

NOTES

8.2 DISTRIBUTION OF FUNCTIONS OF RANDOM VARIABLE

A random variable takes on different values as a result of the outcomes of a random experiment. In other words, a function which assigns numerical values to each element of the set of events that may occur (i.e., every element in the sample space) is termed as random variable. The value of a random variable is the general outcome of the random experiment. One should always make a distinction between the random variable and the values that it can take on. All these can be illustrated by a few examples shown in Table 8.1.

Table 8.1 *Random Variable*

Random Variable	*Values of the Random Variable*	*Description of the Values of the Random Variable*
X	0, 1, 2, 3, 4	Possible number of heads in four tosses of a fair coin
Y	1, 2, 3, 4, 5, 6	Possible outcomes in a single throw of a die
Z	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	Possible outcomes from throwing a pair of dice
M	0, 1, 2, 3, ... S	Possible sales of newspapers by a newspaper boy, S representing his stock

All the stated random variable assignments cover every possible outcome and each numerical value represents a unique set of outcomes. A random variable can be either discrete or continuous. If a random variable is allowed to take on only a limited number of values, it is a discrete random variable, but if it is allowed to assume any value within a given range, it is a continuous random variable. Random variables presented in Table 8.1 are examples of discrete random variables. We can have continuous random variables if they can take on any value within a range of values, for example, within 2 and 5, in that case we write the values of a random variable x as,

$$2 \leq x \leq 5$$

Techniques of Assigning Probabilities

We can assign probability values to the random variables. Since the assignment of probabilities is not an easy task, we should observe following rules in this context:

(*i*) A probability cannot be less than zero or greater than one, i.e., $0 \leq pr \leq 1$, where pr represents probability.

(*ii*) The sum of all the probabilities assigned to each value of the random variable must be exactly one.

There are three techniques of assignment of probabilities to the values of the random variable that are as follows:

(*i*) Subjective Probability Assignment: It is the technique of assigning probabilities on the basis of personal judgement. Such assignment may differ from individual to individual and depends upon the expertise of the person assigning the probabilities. It cannot be termed as a rational way of assigning probabilities, but is used when the objective methods cannot be used for one reason or the other.

(*ii*) A-Priori Probability Assignment: It is the technique under which the probability is assigned by calculating the ratio of the number of ways in which a given outcome can occur to the total number of possible outcomes. The basic underlying assumption in using this procedure is that every possible outcome is likely to occur equally. However, at times the use of this technique gives ridiculous conclusions. For example, we have to assign probability to the event that a person of age 35 will live upto age 36. There are two possible outcomes, he lives or he dies. If the probability assigned in accordance with a-priori probability assignment is half then the same may not represent reality. In such a situation, probability can be assigned by some other techniques.

(*iii*) Empirical Probability Assignment: It is an objective method of assigning probabilities and is used by the decision-makers. Using this technique the probability is assigned by calculating the relative frequency of occurrence of a given event over an infinite number of occurrences. However, in practice only a finite (perhaps very large) number of cases are observed and relative frequency of the event is calculated. The probability assignment through this technique may as well be unrealistic, if future conditions do not happen to be a reflection of the past.

Thus, what constitutes the 'best' method of probability assignment can only be judged in the light of what seems best to depict reality. It depends upon the nature of the problem and also on the circumstances under which the problem is being studied.

Probability Distribution Functions: Discrete and Continuous

When a random variable x takes discrete values $x_1, x_2, ..., x_n$ with probabilities $p_1, p_2, ..., p_n$, we have a discrete probability distribution of X.

The function $p(x)$ for which $X = x_1, x_2, ..., x_n$ takes values $p_1, p_2, ..., p_n$, is the probability function of X.

The variable is discrete because it does not assume all values. Its properties are:

$$p(x_i) = \text{Probability that } X \text{ assumes the value } x$$
$$= \text{Prob } (x = x_i) = p_i$$
$$p(x) \geq 0, \ \Sigma p(x) = 1$$

For example, four coins are tossed and the number of heads X noted. X can take values 0, 1, 2, 3, 4 heads.

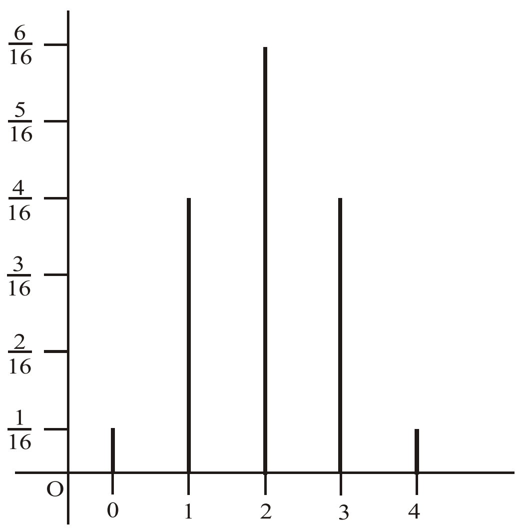
$$p(X = 0) = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

$$p(X = 1) = {}^4C_1\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3 = \frac{4}{16}$$

$$p(X = 2) = {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{6}{16}$$

$$p(X = 3) = {}^4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right) = \frac{4}{16}$$

$$p(X = 4) = {}^4C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = \frac{1}{16}$$



$$\sum_{x=2}^{4} p(x) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16} = 1$$

This is a discrete probability distribution.

NOTES

Example 1: If a discrete variable X has the following probability function, then find (*i*) a (*ii*) $p(X \leq 3)$ (*iii*) $p(X \geq 3)$.

Solution: The solution is obtained as follows:

x_1	$p(x_i)$
0	0
1	a
2	$2a$
3	$2a^2$
4	$4a^2$
5	$2a$

Since $\Sigma p(x) = 1$, $0 + a + 2a + 2a^2 + 4a^2 + 2a = 1$

$\therefore \quad 6a^2 + 5a - 1 = 0$, so that $(6a - 1)(a + 1) = 0$

$$a = \frac{1}{6} \text{ or } a = -1 \text{ (Not admissible)}$$

For $a = \frac{1}{6}$, $p(X \leq 3) = 0 + a + 2a + 2a^2 = 2a^2 + 3a = \frac{5}{9}$

$$p(X \geq 3) = 4a^2 + 2a = \frac{4}{9}$$

(*v*) Continuous Probability Distributions

When a random variate can take any value in the given interval $a \leq x \leq b$, it is a continuous variate and its distribution is a continuous probability distribution.

Theoretical distributions are often continuous. They are useful in practice because they are convenient to handle mathematically. They can serve as good approximations to discrete distributions.

The range of the variate may be finite or infinite.

A continuous random variable can take all values in a given interval. A continuous probability distribution is represented by a smooth curve.

The total area under the curve for a probability distribution is necessarily unity. The curve is always above the x axis because the area under the curve for any interval represents probability and probabilities cannot be negative.

If X is a continous variable, the probability of X falling in an interval with end points z_1, z_2 may be written $p(z_1 \leq X \leq z_2)$.

This probability corresponds to the shaded area under the curve in Figure 8.1.

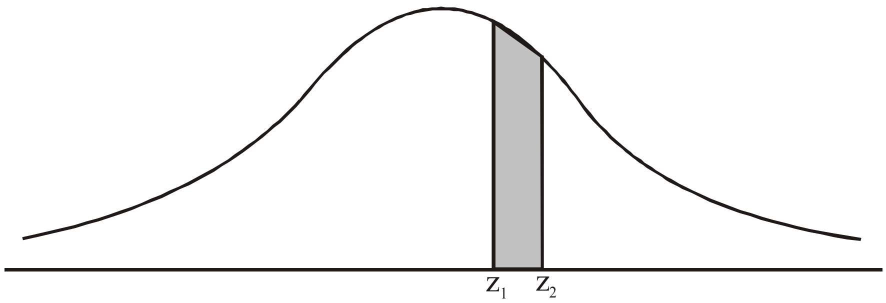


Fig. 8.1 *Continuous Probability Distribution*

A function is a probability density function if,

$\int_{-\infty}^{\infty} p(x)dx = 1, p(x) \geq 0, -\infty < x < \infty,$ i.e., the area under the curve $p(x)$ is 1 and the probability of x lying between two values a, b, i.e., $p(a < x < b)$ is positive. The most prominent example of a continuous probability function is the normal distribution.

Cumulative Probability Function (CPF)

The Cumulative Probability Function (CPF) shows the probability that x takes a value less than or equal to, say, z and corresponds to the area under the curve up to z:

$$p(x \leq z) = \int_{-\infty}^{z} p(x)dx$$

This is denoted by $F(x)$.

Extension to Bivariate Case: Elementary Concepts

If in a bivariate distribution the data is quite large, then they may be summed up in the form of a two-way table. In this for each variable, the values are grouped into different classes (not necessary same for both the variables), keeping in view the same considerations as in the case of univariate distribution. In other words, a bivariate frequency distribution presents in a table pairs of values of two variables and their frequencies.

For example, if there is m classes for the $X-$ variable series and n classes for the $Y-$ variable series then there will be $m \times n$ cells in the two-way table. By going through the different pairs of the values (x, y) and using tally marks, we can find the frequency for each cell and thus get the so called bivariate frequency table as shown in Table 8.2.

Table 8.2 Bivariate Frequency Table

x Series → / Y Series ↓	Classes / Mid Points: x_1 x_1 ... x .... x_m	Total of Frequencies of y
Classes / Mid Points: y_1, y_2, y, ..., y_n	$f(x, y)$	f_y
Total of Frequencies of x	f_x	Total $f_x = \ f_y = N$

Here, $f(x,y)$ is the frequency of the pair (x, y). The formula for computing the correlation coefficient between x and y for the bivariate frequency table is,

$$r = \frac{N\Sigma xyf(x, y) - (\Sigma xfx)(\Sigma yfy)}{\sqrt{\left[N\Sigma x^2 f_x - (\Sigma xfx)^2\right] \times \left[N\Sigma y^2 f_y - (\Sigma yfy)^2\right]}}$$

Where, N is the total frequency.

Check Your Progress

1. What is meant by subjective probability assignment?
2. What is continuous probability distribution?

8.3 SAMPLING THEORY

A sample is a portion of the total population that is considered for study and analysis. For instance, if we want to study the income pattern of professors at City University of New York and there are 10,000 professors, then we may take a random sample of only 1,000 professors out of this entire population. Then this number of 1,000 professors constitutes a sample. The summary measure that describes a characteristic, such as average income of this sample is known as a statistic.

Sampling is the process of selecting a sample from the population. It is technically and economically not feasible to take the entire population for analysis. So we must take a representative sample out of this population for the purpose of such analysis. A sample is part of the whole, selected in such a manner as to be representing the whole.

Random Sample

It is a collection of items selected from the population in such a manner that each item in the population has exactly the same chance of being selected, so that the sample taken from the population would be truly representative of the population. The degree of randomness of selection would depend upon the process of selecting the items from the sample. A true random sample would be free from all biases whatsoever. For example, if we want to take a random sample of five students from a class of twenty-five students, then each one of these twenty-five students should have the same chance of being selected in the sample. One way to do this would be writing the names of all students on separate but small pieces of paper, folding each piece of this paper in a similar manner, putting each folded piece into a container, mixing them thoroughly and drawing out five pieces of paper from this container.

Sampling without Replacement

The sample as taken in the previous example is known as sampling without replacement, because each person can only be selected once so that once a piece of paper is taken out of the container, it is kept aside so that the person whose name appears on this piece of paper has no chance of being selected again.

Sampling with Replacement

There are certain situations in which the piece of paper once selected and taken into consideration is put back into the container in such a manner that the same person has the same chance of being selected again as any other person. For example, if we are randomly selecting five persons for award of prizes so that each person is eligible for any and all prizes, then once the slip of paper is drawn out of the container and the prize is awarded to the person whose name appears on the paper, the same piece of paper is put back into the container and the same person has the same chance of winning the second prize as anybody else.

Sample Selection

The third step in the primary data collection process is selecting an adequate sample. It is necessary to take a representative sample from the population, since it is extremely costly, time-consuming and cumbersome to do a complete census. Then, depending upon the conclusions drawn from the study of the characteristics of such a sample, we can draw inferences about the similar characteristics of the population. If the sample is truly representative of the population, then the characteristics of the sample can be considered to be the same as those of the entire population. For example, the taste of soup in the entire pot of soup can be determined by tasting one spoonful from the pot if the soup is well stirred. Similarly,

NOTES

a small amount of blood sample taken from a patient can determine whether the patient's sugar level is normal or not. This is so because the small sample of blood is truly representative of the entire blood supply in the body.

Sampling is necessary because of the following reasons: First, as discussed earlier, it is not technically or economically feasible to take the entire population into consideration. Second, due to dynamic changes in business, industrial and social environment, it is necessary to make quick decisions based upon the analysis of information. Managers seldom have the time to collect and process data for the entire population. Thus, a sample is necessary to save time. The time element has further importance in that if the data collection takes a long time, then the values of some characteristics may change over the period of time so that data may no longer be up to date, thus defeating the very purpose of data analysis. Third, samples, if representative, may yield more accurate results than the total census. This is due to the fact that samples can be more accurately supervised and data can be more carefully selected. Additionally, because of the smaller size of the samples, the routine errors that are introduced in the sampling process can be kept at a minimum. Fourth, the quality of some products must be tested by destroying the products. For example, in testing cars for their ability to withstand accidents at various speeds, the environment of accidents must be simulated. Thus, a sample of cars must be selected and subjected to accidents by remote control. Naturally, the entire population of cars cannot be subjected to these accident tests and hence, a sample must be selected.

One important aspect to be considered is the size of the sample. The sampling size—which is the number of sampling units selected from the population for investigation—must be optimum. If the sample size is too small, it may not appropriately represent the population or the universe as it is known, thus leading to incorrect inferences. Too large a sample would be costly in terms of time and money. The optimum sample size should fulfil the requirements of efficiency, representativeness, reliability and flexibility. What is an optimum sample size is also open to question. Some experts have suggested that 5 per cent of the population properly selected would constitute an adequate sample, while others have suggested as high as 10 per cent depending upon the size of the population under study. However, proper selection and representation of the sample is more important than size itself. The following considerations may be taken into account in deciding about the sample size:

(a) The larger the size of the population, the larger should be the sample size.

(b) If the resources available do not put a heavy constraint on the sample size, a larger sample would be desirable.

(c) If the samples are selected by scientific methods, a larger sample size would ensure greater degree of accuracy in conclusions.

(d) A smaller sample could adequately represent the population, if the population consists of mostly homogeneous units. A heterogeneous universe would require a larger sample.

Census and Sampling

Under the census or complete enumeration survey method, data is collected for each and every unit (for example, person, consumer, employee, household, organization) of the population or universe which are the complete set of entities and which are of interest in any particular situation. In spite of the benefits of such an all-inclusive approach, it is infeasible in most of the situations. Besides, the time and resource constraints of the researcher, infinite or huge population, the incidental destruction of the population unit during the evaluation process (as in the case of bullets, explosives etc) and cases of data obsolescence (by the time census ends) do not permit this mode of data collection.

Sampling is simply a process of learning about the population on the basis of a sample drawn from it. Thus, in any sampling technique, instead of every unit of the universe, only a part of the universe is studied and the conclusions are drawn on that basis for the entire population. The process of sampling involves selection of a sample based on a set of rules, collection of information and making an inference about the population. It should be clear to the researcher that a sample is studied not for its own sake, but the basic objective of its study is to draw inference about the population. In other words, sampling is a tool which helps us know the characteristics of the universe or the population by examining only a small part of it. The values obtained from the study of a sample, such as the average and dispersion are known as 'statistics' and the corresponding such values for the population are called 'parameters'.

Although diversity is a universal quality of mass data, every population has characteristic properties with limited variation. The following two laws of statistics are very important in this regard.

1. The law of statistical regularity states that a moderately large number of items chosen at random from a large group are almost sure on the average to possess the characteristics of the large group. By random selection, we mean a selection where each and every item of the population has an equal chance of being selected.
2. The law of inertia of large numbers states that, other things being equal, larger the size of the sample, more accurate the results are likely to be.

Hence, a sound sampling procedure should result in a representative, adequate and homogeneous sample while ensuring that the selection of items should occur independently of one another.

Sampling Techniques

The various methods of sampling can be grouped under two broad categories: Probability (or random) sampling and Non-probability (or non-random) sampling.

Probability sampling methods are those in which every item in the universe has a known chance, or probability of being chosen for the sample. Thus, the sample selection process is objective (independent of the person making the study) and hence, random. It is worth noting that randomness is a property of the sampling procedure instead of an individual sample. As such, randomness can enter processed sampling in a number of ways and hence, random samples may be of many types. These methods include: (a) Simple Random Sampling, (b) Stratified Random Sampling, (c) Systematic Sampling and (d) Cluster Sampling.

Non-probability sampling methods are those which do not provide every item in the universe with a known chance of being included in the sample. The selection process is, at least, partially subjective (dependent on the person making the study). The most important difference between random and non-random sampling is that whereas the pattern of sampling variability can be ascertained in case of random sampling, there is no way of knowing the pattern of variability in the non-random sampling process. The non-probability methods include: (a) Judgement Sampling, (b) Quota Sampling and (c) Convenience Sampling.

The following Figure 8.2 depicts the broad classification and sub-classification of various methods of sampling.

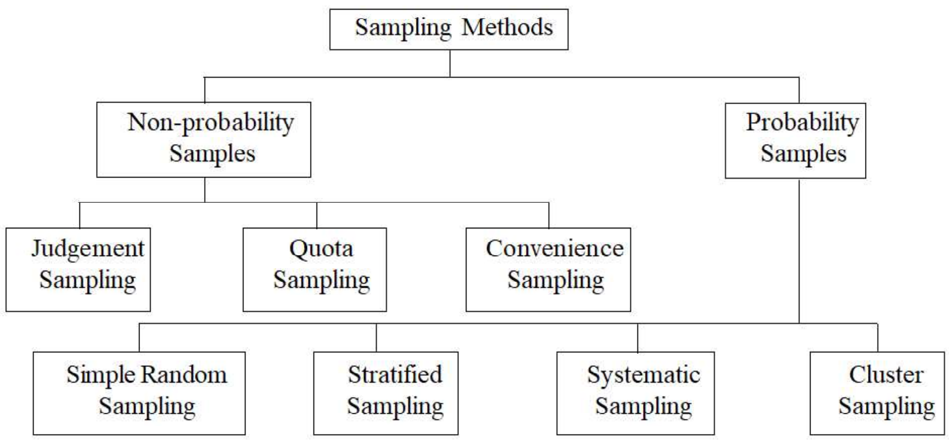


Fig. 8.2 *Methods of Sampling*

Non-Probability Sampling Methods

(a) **Judgement Sampling:** In this method of sampling, the choice of sample items depends exclusively on the judgement of the investigator. The sample here is based on the opinion of the researcher, whose discretion will clinch the sample. Though the principles of sampling theory are not applicable to judgement sampling, it is sometimes found to be useful. When we want to study some unknown traits of a population, some of whose characteristics are known, we may then stratify the population according to these known properties and select sampling units from each stratum on the basis of judgement. Naturally, the success of this method depends upon the excellence in judgement.

(b) **Convenience Sampling:** A convenience sample is obtained by selecting convenient population units. It is also called a chunk, which refers to that fraction of the population being investigated, which is selected neither by probability nor by judgement but by convenience. A sample obtained from readily available lists such as telephone directories is a convenience sample and not a random sample, even if the sample is drawn at random from such lists. In spite of the biased nature of such a procedure, convenience sampling is often used for pilot studies.

(c) **Quota Sampling:** Quota sampling is a type of judgement sampling and is perhaps the most commonly used sampling technique in non-probability category. In a quota sample, quotas (or minimum targets) are set up according to some specified characteristics, such as age, income group, religious or political affiliations, and so on. Within the quota, the selection of the sample items depends on personal judgement. Because of the risk of personal prejudice entering the sample selection process, the quota sampling is not widely used in practical works.

It is worth noting that similarity between quota sampling and stratified random sampling is confined to dividing the population into different strata. The process of selection of items from each of these strata in the case of stratified random sampling is random, while it is not so in the case of quota sampling. Quota sampling is often used in public opinion studies.

Probability Sampling Methods

The following are the probability sampling methods:

(a) Simple Random Sampling
(b) Stratified Random Sampling
(c) Systematic Sampling
(d) Multistage or Cluster Sampling

Simple Random Sampling

It refers to that sampling technique in which each and every unit of the population has an equal chance of being selected in the sample. One should not mistake the term 'arbitrary' for 'random'. To ensure randomness, one may adopt either the lottery method or consult the table of random numbers, preferably the latter. Being a random method, it is independent of personal bias creeping into the analysis besides enhancing the representativeness of the sample. Furthermore, it is easy to assess the accuracy of the sampling estimates because sampling errors follow the principles of chance. However, a completely catalogued universe is a prerequisite for this method. The sample size requirements would be usually larger under random sampling than under stratified random sampling, to ensure statistical reliability. It may escalate the cost of collecting data as the cases selected by random sampling tend to be too widely dispersed geographically.

Stratified Random Sampling

In this method, the universe to be sampled is subdivided (stratified) into groups which are mutually exclusive, but collectively exhaustive based on a variable known to be correlated with the variable of interest. Then, a simple random sample is chosen independently from each group. This method differs from simple random sampling in that, in the latter the sample items are chosen at random from the entire universe. In stratified random sampling, the sampling is designed in such a way that a designated number of items is chosen from each stratum. If the ratio of items between various strata in the population matches with the ratio of corresponding items between various strata in the sample, it is called proportionate stratified sampling; otherwise, it is known as disproportionate stratified sampling. Ideally, we should assign greater representation to a stratum with a larger dispersion and smaller representation to one with small variation. Hence, it results in a more representative sample than simple random sampling.

Systematic Sampling

It is also known as quasi-random sampling method because once the initial starting point is determined, the remainder of the items selected for the sample are predetermined by the sampling interval. A systematic sample is formed by selecting one unit at random and then selecting additional units at evenly spaced intervals until the sample has been formed. This method is popularly used in those cases where a complete list of the population from which sample is to be drawn is available. The list may be prepared in alphabetical, geographical, numerical or some other order. The items are serially numbered. The first item is selected at random generally by following the lottery method. Subsequent items are selected by taking every *K*th item from the list where '*K*' stands for the sampling interval or the sampling ratio, i.e., the ratio of the population size to the size of the sample.

Symbolically,

$K = N / n$, where K = Sampling Interval; N = Universe Size; n = Sample Size. In case K is a fractional value, it is rounded off to the nearest integer.

Cluster Sampling

Under this method, the random selection is made of primary, intermediate and final (or the ultimate) units from a given population or stratum. There are several stages in which the sampling process is carried out. At first, the stage units are sampled by some suitable method such as simple random sampling. Then, a sample of second stage units is selected from each of the selected first stage units, by applying some suitable method which may or may not be the same method employed for the first stage units. For example, in a survey of 10,000 households in AP, we may choose a few districts in the first stage, a few towns/villages/*mandals* in the second stage and select a number of households from each town/village/*mandal* selected in the previous stage. This method is quite flexible and is particularly useful in surveys of underdeveloped areas, where no frame is generally sufficiently detailed and accurate for subdivision of the material into reasonably small sampling units. However, a multistage sample is, in general, less accurate than a sample containing the same number of final stage units which have been selected by some suitable single stage process.

Sampling and Non-Sampling Errors

The basic objective of a sample is to draw inferences about the population from which such sample is drawn. This means that sampling is a technique which helps us in understanding the parameters or the characteristics of the universe or the population by examining only a small part of it. Therefore, it is necessary that the sampling technique be a reliable one. The randomness of the sample is especially important because of the principle of statistical regularity, which states that a sample taken at random from a population is likely to possess almost the same characteristics as those of the population. However, in the total process of statistical analysis, some errors are bound to be introduced. These errors may be the sampling errors or the non-sampling errors. The sampling errors arise due to drawing faulty inferences about the population based upon the results of the samples. In other words, it is the difference between the results that are obtained by the sample study and the results that would have been obtained if the entire population was taken for such a study, provided that the same methodology and manner was applied in studying both the sample as well as the population. For example, if a sample study indicates that 25 per cent of the adult population of a city does not smoke and the study of the entire adult population of the city indicates that 30 per cent are non-smokers, then this difference would be considered as the sampling error. This sampling error would be smallest if the sample size is large relative to the population, and vice versa.

Non-sampling errors, on the other hand, are introduced due to technically faulty observations or during the processing of data. These errors could also arise due to defective methods of data collection and incomplete coverage of the population, because some units of the population are not available for study, inaccurate information provided by the participants in the sample and errors occurring during editing, tabulating and mathematical manipulation of data. These are the errors which can arise even when the entire population is taken under study.

Both the sampling as well as the non-sampling errors must be reduced to a minimum in order to get as representative a sample of the population as possible.

Theory of Estimation

The theory of estimation is a very common and popular statistical method and is used to calculate the mathematical model for the data to be considered. This method was introduced by the statistician Sir R. A. Fisher, between 1912 and 1922. This method can be used in:

- Finding linear models and generalized linear models

NOTES

- Exploratory and confirmatory factor analysis
- Structural equation modelling
- Calculating time-delay of arrival (TDOA) in acoustic or electromagnetic detection
- Data modelling in nuclear and particle physics
- Finding the result for hypothesis testing

The method of estimation is used with known mean and variance. The sample mean becomes the maximum likelihood estimator of the population mean, and the sample variance becomes the close approximation to the maximum likelihood estimator of the population variance.

Interval Estimate of the Population Mean (Population Variance Known)

Since the sample means are normally distributed, with a mean of μ and a standard deviation of $\sigma_{\bar{x}}$, it follows that sample means follow normal distribution characteristics. Transforming the sampling distribution of sample means into the standard normal distribution, we get:

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

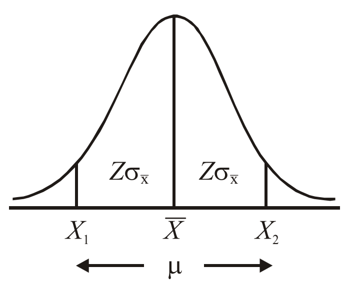
$$or \ \bar{x} - \mu = Z\sigma_{\bar{x}}$$

$$or \ \mu = \bar{x} - Z\sigma_{\bar{x}}$$

Since μ falls within a range of values equidistant from $\bar{x}$,

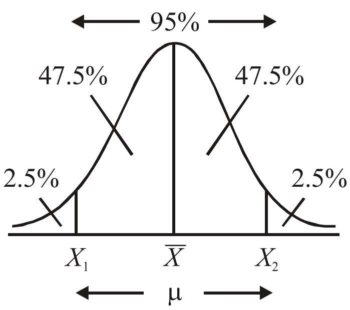
$$\mu = \bar{x} \pm Z\sigma_{\bar{x}}$$

This relationship is shown in the following illustration.



This means that the population mean is expected to lie between the values of x_1 and x_2 which are both equidistant from $\bar{x}$ and this distance depends upon the value of Z which is a function of confidence level.

Suppose that we wanted to find out a *confidence interval* around the sample mean within which the population mean is expected to lie 95 per cent of the time. (We can never be sure that the population mean will lie in any given interval 100 per cent of the time). This confidence interval is shown as follows:



The points x_1 and x_2 above define the range of the confidence interval as follows:

$$x_1 = \overline{x} - Z\sigma_{\overline{x}}$$

$$\text{and } x_2 = \overline{x} + Z\sigma_{\overline{x}}$$

Looking at the table of Z scores, (given in the Appendis) we find that the value of Z score for area 10.4750 (half of 95 per cent) is 1.96. This illustration can be interpreted as follows:

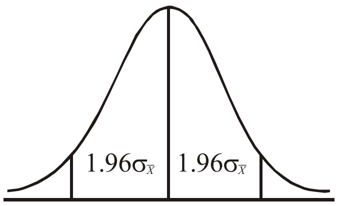
a) If all possible samples of size n were taken, then on the average 95 per cent of these samples would include the population mean within the interval around their sample means bounded by x_1 and x_2.

b) If we took a random sample of size n from a given population, the probability is 0.95 that the population mean would lie between the interval x_1 and x_2 around the sample mean, as shown.

c) If a random sample of size n was taken from a given population, we can be 95 per cent confident in our assertion that the population mean will lie around the sample mean in the interval bounded by values of x_1 and x_2 as shown. (It is also known as 95 per cent confidence interval.) At 95 per cent confidence interval, the value of Z score as taken from the Z score table is 1.96. The value of Z score can be found for any given level of confidence, but generally speaking, a confidence level of 90%, 95% or 99% is taken into consideration for which the Z score values are 1.68, 1.96 and 2.58, respectively.

Example 2: The sponsor of a television programme targeted at the children's market (age 4-10 years) wants to find out the average amount of time children spend watching television. A random sample of 100 children indicated the average time spent by these children watching television per week to be 27.2 hours. From previous experience, the population standard deviation of the weekly extent of television watched (s) is known to be 8 hours. A confidence level of 95 per cent is considered to be adequate.

Solution:



The confidence interval is given by:

$$\bar{x} \pm Z\sigma_{\bar{x}} \qquad \text{or } \bar{x} - Z\sigma_{\bar{x}} < \mu < \bar{x} + Z\sigma_{\bar{x}}$$

Where $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

Accordingly, we need only four values, namely $\bar{x}$, Z, σ and n. In our case:

$$\bar{x} = 27.2$$
$$Z = 1.96$$
$$\sigma = 8$$
$$n = 100$$

Hence $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{8}{\sqrt{100}} = \dfrac{8}{10} = .8$

Then:

$$\begin{aligned} x_1 &= \bar{x} - Z\sigma_{\bar{x}} \\ &= 27.2 - (1.96 \times .8) = 27.2 - 1.568 \\ &= 25.632 \end{aligned}$$

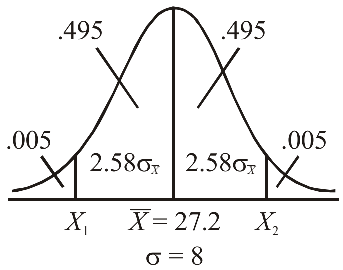
And

$$\begin{aligned} x_2 &= \bar{x} + Z\sigma_{\bar{x}} \\ &= 27.2 + (1.96 \times .8) = 27.2 + 1.568 \\ &= 28.768 \end{aligned}$$

This means that we can conclude with 95 per cent confidence that a child on an average spends between 25.632 and 28.768 hours per week watching television. (It should be understood that 5 per cent of the time our conclusion would still be wrong. This means that because of the symmetry of distribution, we will be wrong 2.5 per cent of the times because the children on an average would be watching television more than 28.768 hours and another 2.5 per cent of the time we will be wrong in our conclusion, because on an average, the children will be watching television less than 25.632 hours per week.)

Example 3: Calculate the confidence interval in the previous problem, if we want to increase our confidence level from 95% to 99%. Other values remain the same.

Solution:



If we increase our confidence level to 99 per cent, then it would be natural to assume that the range of the confidence interval would be wider, because we would want to include more values which may be greater than 28.768 or smaller than 25.632 within the confidence interval range. Accordingly, in this new situation,

$$Z = 2.58$$

$$\sigma_{\bar{x}} = .8$$

Then

$$x_1 = \bar{x} - Z\sigma_{\bar{x}}$$
$$= 27.2 - (2.58\ x\ .8) = 27.2 - 2.064$$
$$= 25.136$$

And

$$X_2 = \bar{x} + Z\sigma_{\bar{x}}$$
$$= 27.2 + 2.064$$
$$= 29.264$$

(The value of Z is established from the table of Z scores against the area of .495 or a figure closest to it. The table shows that the area close to .495 is .4949 for which the Z score is 2.57 or .4951 for which the Z score is 2.58. In practice, the Z score of 2.58 is taken into consideration when calculating 99 per cent confidence interval.)

Interval Estimate of the Population Mean (Population Variance Unknown)

As the previous example shows, in order to determine the interval estimate of μ, the variance (and hence, the standard deviation) must be known, since it figures in the formula. However, the standard deviation of the population is generally not known. In such situations and when sample size is reasonably large (30 or more), we can approximate the population standard deviation (σ) by the sample standard deviation (s), so that the confidence interval,

NOTES

$\bar{x} \pm Z\sigma_{\bar{x}}$ is approximated by the interval.

$\bar{x} \pm Zs_{\bar{x}}$, when $n \geq 30$.

where $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ and $s_{\bar{x}} = \dfrac{s}{\sqrt{n}}$.

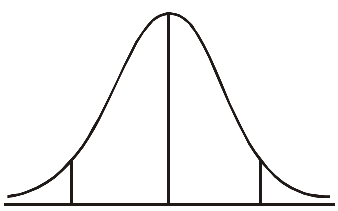
Example 4: It is desired to estimate the average age of students who graduate with an MBA degree in the university system. A random sample of 64 graduating students showed that the average age was 27 years with a standard deviation of 4 years.

a) Estimate a 95 per cent confidence interval estimate of the true average (population mean) age of all such graduating students at the university.

b) How would the confidence interval limits change if the confidence level was increased from 95 per cent to 99 per cent.

Solution: Since the sample size *n* is sufficiently large, we can approximate the population standard deviation by the sample standard deviation.

(a)



Now,

$$Z = 1.96$$

$$\bar{x} = 27$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{4}{\sqrt{64}} = \frac{4}{8} = 0.5$$

95% confidence interval of population mean μ is given by:

$$\bar{x} \pm Zs_{\bar{x}}$$

So that,

$$x_1 = \bar{x} - Zs_{\bar{x}}$$

$$= 27 - (1.96\ x\ .5) = 27 - 0.98$$

$$= 26.02$$

And

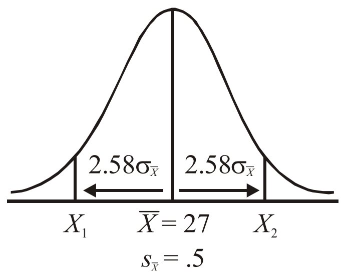
$$x_2 = \bar{x} + Zs_{\bar{x}}$$

$$=27+0.98$$

$$=27.98$$

Hence, $26.02 \geq \mu \geq 27.98$.

(b) $s_{\bar{x}} = .5$



Now, Z becomes 2.58 and the other values remain the same. Hence,

$$x_1 = \bar{x} - Zs_{\bar{x}}$$
$$= 27 - (2.58\ x\ .5) = 27 - 1.29$$
$$= 25.71$$

And

$$x_2 = \bar{x} + Zs_{\bar{x}}$$
$$= 27 + 1.29$$
$$= 28.29$$

Hence, $25.71 \leq \mu \leq 28.29$.

Sample Size Determination for Estimating the Population Mean

It is understood that the larger the sample size, the closer the sample statistic will be to the population parameter. Hence, the degree of accuracy we require in our estimate would be one factor influencing our choice of sample size. The second element that influences the choice of the sample size is the degree of confidence in ourselves that the error in the estimate remains within the degree of accuracy that is desired. Hence, the degree of accuracy has two aspects.

1. The maximum allowable error in our estimate
2. The degree of confidence that the error in our estimate will not exceed the maximum allowable error

The ideal situation would be that the sample mean $\bar{x}$ equals the population mean μ. That would be the best estimate of μ based on $\bar{x}$. If the entire population was taken as a sample then $\bar{x}$ will be equal to μ and there will be no error in our estimate. Hence, ($\bar{x}$ - μ) can be considered as *error* or *deviation* of the estimator $\bar{x}$ from the population mean μ. This maximum allowable *error* must be pre-established. Let this error be denoted by E, so that:

$$E = (\bar{x} - \mu)$$

Now, we know that

NOTES

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

$$= \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$= \frac{E}{\sigma / \sqrt{n}}$$

Or

$$Z = \frac{E\sqrt{n}}{\sigma}$$

$$Z\sigma = E\sqrt{n}$$

$$\sqrt{n} = \frac{Z\sigma}{E}$$

$$n = \left(\frac{Z\sigma}{E}\right)^2 \text{ or } \frac{Z^2\sigma^2}{E^2}$$

Based upon this formula, it can be seen that the size of the sample depends upon:

(a) Confidence interval desired. This will determine the value of Z. For example, 95 per cent confidence level yields the value of Z to be = 1.96.

(b) Maximum error allowed (E)

(c) The standard deviation of the population (σ)

It can further be seen from this formula that the sample size will increase if:

(a) The allowable error becomes smaller

(b) The degree of confidence increases

(c) The value of the variance within the population is larger

Example 5: We would like to know the average time that a child spends watching television over the weekend. We want our estimate to be within $\pm$ 1 hour of the true population average. (This means that the maximum allowable error is 1 hour.) Previous studies have shown the population standard deviation to be 3 hours. What sample size should be taken for this purpose, if we want to be 95 per cent confident that the error in our estimate will not exceed the maximum allowable error?

Solution: For 95 per cent confidence level, the values of

Z = 1.96

E = 1 hour (given)

σ = 3 hours (given)

Then,

$$n = \frac{Z^2\sigma^2}{E^2}$$

$$= \frac{(1.96)^2(3)^2}{(1)^2}$$

$$= 34.57$$

To be more accurate in our estimate, we always round off the answer to the next higher figure from the decimal. Hence, $n = 35$.

Confidence Interval Estimation of Population Proportion

So far we have discussed the estimation of population mean, which is quantitative in nature. This concept of estimation can be extended to qualitative data where the data is available in proportion or percentage form. In this situation the parameter of interest is π, which is the proportion of times a certain desirable outcome occurs. This concept lends itself to binomial distribution where we label the outcome of interest to us as success with the probability of success being π and the probability of failure being $(1-\pi)$.

When large samples of size n are selected from a population having a proportion of desirable outcomes π, then the sampling distribution of proportions is normally distributed. For large samples, when (np) as well as (nq) are both at least equal to 5, where n is the sample size, p is the probability of a desired outcome (or success) and q is the probability of failure (1-p), then the binomial distribution can also be approximated to normal distribution, with a mean of π and a standard deviation of σ_p, where σ_p is given by:

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

In such cases, we expect 95 per cent of all sample proportions to fall within the following range:

$$\pi \pm 1.96\,\sigma_p$$

Or

$$\pi \pm 1.96\sqrt{\frac{\pi(1-\pi)}{n}}$$

Similarly, 99 per cent of all such sample proportions will fall within

$$\pi \pm 2.58\sqrt{\frac{\pi(1-\pi)}{n}}$$

If all possible samples of size n are selected and the interval $p \pm 1.96\sigma_p$ is established for each sample, where p is the sample proportion, then 95 per cent of all such intervals are expected to contain π, the population proportion. Then this range of,

$$p \pm 1.96\sigma_p$$

Or

$$p \pm 1.96\sqrt{\frac{\pi(1-\pi)}{n}}$$

Is known as the 95 per cent confidence interval estimate of π.

This formula requires that we know the value of π in order to calculate σ_p. But, population proportion is generally not known. In such instances, the sample proportion is used as an approximation of π. Hence, the 95 per cent confidence interval estimate of π becomes:

$$p \pm 1.96\ \sigma_p$$

Where

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

And 99 per cent confidence interval estimate of π becomes:

$$p \pm 2.58\ \sigma_p$$

Example 6: A survey of 500 persons shopping at a mall, selected at random, showed that 350 of them used credit cards for their purchases and 150 of them used cash.

(a) Construct a 95 per cent confidence interval estimate of the proportion of all persons at the mall, who use credit card for shopping.

(b) What would our confidence level be, if we make the assertion that the proportion of shoppers at the mall who shop with a credit card is between 67 per cent and 73 per cent.

Solution: (a) There are 350 people out of a total sample of 500 who pay by credit card. Hence, the sample proportion of credit card shoppers is:

$$p = 350/500 = 0.7.$$

The 95 per cent confidence interval estimate of population proportion π is given as follows:

$$p \pm 1.96\ \sigma_p$$

NOTES

Where,

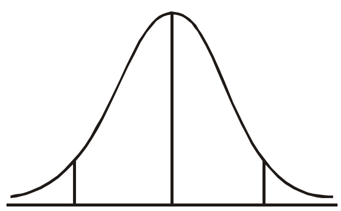
$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

(Since π is not known, we approximate sample proportion p for population proportion π).

Then, $\sigma_p = \sqrt{\frac{.7(.3)}{500}} = \sqrt{.00042} = .02$

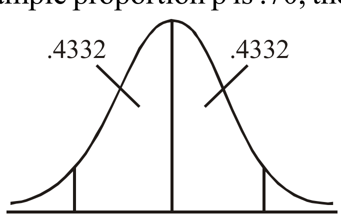
Then the confidence limits are:

$$p_1 = p - 1.96\ \sigma_p$$
$$= .7 - 1.96(.02)$$
$$= .7 - .0392 = .6608 \text{ or } 66.08\%, \text{ and}$$
$$p_2 = p + 1.96\ \sigma_p$$
$$= .7 + .0392 = .7392 \text{ or } 73.92\%.$$



This means that the population of people who pay by credit card at the mall is between 66.8 per cent and 73.92 per cent.

(b) If the population proportion of credit card shoppers is given to be between .67 and .73, when such sample proportion p is .70, then



$$p_1 = p - Z\sigma_p$$
$$.67 = .7 - Z\,(0.2)$$
$$.02Z = .7 - .67$$
$$Z_1 = \frac{.7 - .67}{.02} = \frac{.03}{.02} = 1.5$$

NOTES

Similarly,

$$p_2 = p + Z\sigma_p$$
$$.73 = .7 + Z\ (.02)$$
$$Z_2 = \frac{.73 - .7}{.02} = \frac{.03}{.02} = 1.5$$

Using the Z score table, we see that the area under the curve for Z=1.5 is .4332. This area is on each side of the mean so that the total area is .8664. In other words, our confidence level is 86.64 per cent that the proportion of shoppers using credit card is between 67 per cent and 73 per cent.

Sample Size Determination for Estimating the Population Proportion

We follow the same procedure as we did in determining the sample size for estimating the population mean. As before, there are three factors that are taken into consideration. These are:

(a) The level of confidence desired
(b) The maximum allowable error permitted in the estimate
(c) The estimated population proportion of success π

As established previously,

$$Z = \frac{p - \pi}{\sigma_p}$$

Where $\sigma_p = \sqrt{\dfrac{\pi(1-\pi)}{\sigma_p}}$

Now, $(p-\pi)$ can be considered as error (E), so that:

$$Z = \frac{E}{\sqrt{\dfrac{\pi(1-\pi)}{n}}}$$

By cross-multiplication we get,

$$E = Z\sqrt{\frac{\pi(1-\pi)}{n}}$$

Squaring both sides we get,

$$E^2 = \frac{Z^2\pi(1-\pi)}{n}$$

$$\text{or } nE^2 = Z^2\pi(1-\pi)$$

$$\text{or } n = \frac{Z^2\pi(1-\pi)}{E^2}$$

This formula assumes that we know π, the population proportion, which we are trying to estimate in the first place. Accordingly, π is unknown. However, if any previous studies have estimated this value or a sample proportion p has been calculated in previous studies, then we can approximate this p for π and hence,

$$n = \frac{Z^2 p(1-p)}{E^2}$$

However, if no previous surveys have been taken so that we do not know the value of π or p, then we assume π to be equal to 0.5, simply because, other things being given, the value of π being 0.5 will result in a larger sample size than any other value assumed by π. Hence, the sample size would be at least as large or larger than required for the given conditions. This can be established by the fact that when $\pi = 0.5$, then $\pi(1 - \pi)$ is $.5 \times .5 = 0.25$. This value is larger than any other value of $\pi (1 - \pi)$. This means that when $\pi = 0.5$ then for a given value of Z, n would be larger than any other value of π. This results in a more conservative estimate which is desirable.

Example 7: It is desired to estimate the proportion of children watching television on Saturday mornings, in order to develop a promotional strategy for electronic games. We want to be 95 per cent confident that our estimate will be within ± 2 per cent of the true population proportion.

(a) What sample size should we take if a previous survey showed that 40 per cent of children watched television on Saturday mornings?

(b) What would be the sample size, for the same degree of confidence and the same maximum allowable error, if no such previous survey had been taken?

Solution: (a) In this case, the following values are given:

$Z = 1.96$ (95% confidence interval)
$p = 0.4$
$E = 0.02$

Substituting these values in the following formula, we get:

$$\begin{aligned} n &= \frac{Z^2 p(1-p)}{E^2} \\ &= \frac{(1.96)^2(.4)(.6)}{(.02)^2} \\ &= \frac{0.922}{0.0004} \\ &= 2304.96 \\ &= 2305 \end{aligned}$$

For the sake of accuracy, we always round off to the next higher figure, in case of answer being a fraction.

(b) In this case, since no previous surveys have been taken, we assume $p = 0.5$ and follow the earlier procedure.

$$n = \frac{Z^2 p(1-p)}{E^2}$$
$$= \frac{(1.96)^2 (.5)(.5)}{(.02)^2}$$
$$= \frac{0.9604}{0.0004}$$
$$= 2401$$

8.4 TRANSFORMATION OF VARIABLE OF THE DISCRETE TYPE

A random variable is defined as a function from Ω to R, which constantly takes the numerical values. Here, Ω is referred as the set of possible outcomes of a probability experiment, consequently a random variable as a function can be written as $X : \Omega \to \mathrm{R}$, which specifies the method of assigning a numerical value to each respective outcome of the probability experiment. Generally, the outcome of a probability experiment can be identified or known when the experiment is carried out, therefore *X* or any another variable name is used for representing this outcome of the experiment before actually knowing the exact value.

The following examples explain the concept of random variables. When a coin is flipped, then the possibility is that either there is the Head (H) or the Tail (T), however you can define the random variable associated/related with the flipping of coin as *X*, by means of

$$X(\mathrm{H}) = 0$$
$$X(\mathrm{T}) = 1$$

Or,

$$Y(\mathrm{H}) = 1$$
$$Y(\mathrm{T}) = -1$$

The numerical values depend on the problem statistics, for example, when the coin is flipped in a chance game then the outcome as head gives Rupee 5 while the tail gives Rupee 8. The outcomes of the probability experiment, as heads and tails, can be represented by {–8, 5} using the profit-loss concept. This can be represented using W for the random variable signifying the winnings, by means of,

$$\mathrm{W} : \{\mathrm{H}, \mathrm{T}\} \to \{5, -8\}$$

A **discrete random variable** is the variable that takes values in a finite or countable infinite subset of R. The value of the random variable for the specific range can be denoted by *I*.

Example 8: A coin is flipped three times. Find out the appearance of heads in three flips and the range of random variable.

Solution: Let X be the random variable which represents the number of times heads can appear in the three flips.

Then,

$$X : \Omega \rightarrow \{0, 1, 2, 3\} \subset R$$

The range of X is given by,

$$I = \{0, 1, 2, 3\}$$

Example 9: Two dice are rolled simultaneously. What will be the random variable and the range for the maximum roll?

Solution: Consider that the random variable is assumed or specified for the maximum roll when the two dice are rolled simultaneously. Assume that the maximum rolls are 6.

Here,

X = Value on the First Die

Y = Outcome of the Second Die

Z = max (X, Y) is the Random Variable to consider.

The range of Z is given by,

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

The concept of the discrete random variable can be defined using:

- The probability mass function.
- The expected value of a random variable.
- The variance of a random variable.

1. For a discrete random variable, the probability mass function can be find using the equation,

   $$f(k) = P(X = k) \text{ for } k \in I$$

2. The expected value (expectation, mean) of a random variable can be defined using the probability mass function,

   $$\mu = E(X) = \sum_{k \in I} kP(X = k)$$

In this case, each outcome is subjective/weighted by its respective probability. The expected value function is specifically written using the notation E[.] or E(.), and for the random variable we can use the notation μ (mu) for mean.

3. The variance can be defined as,

   $\sigma^2 = \text{var}(X) = E[(X - \mu)^2]$

   This definition defines the expected value of the square of the difference of X from the mean μ. The var (.) function can be applied to any random variable, and the notation σ^2 defines which random variable is being evaluated.

Definition. Two discrete random variables X and Y are independent if for all x in the range of X and all y in the range of Y, we have

$$P(X = x, Y = y) = P(X = x)\, P(Y = y)$$

Here P denotes the probability. If X and Y are not independent, then this may not be true for all combinations of x and y and hence there will be some event $(X = x, Y = y)$ for which this equation is false.

Example 10: A coin is flipped only once. Find the number of heads and tails. Specify that the discrete random variables is independent random variable.

Solution: When the coin is flipped only once, then the probability of getting heads and tails on the same flip is zero, i.e., the events are mutually exclusive. Let X be the number of heads and Y be the number of tails. Then,

$$P(X = 1)\, P(Y = 1) = 0.25$$

It can be stated that the joint probability mass function of independent random variables X and Y factors into the probability mass functions for X and Y.

As per the definition of expected value the following fact can be proved for the discrete random variables:

If X and Y are independent random variables, then

$$E(XY) = E(X)\, E(Y)$$

Provided that both $E(X)$ and $E(Y)$ exist and are finite.

Check Your Progress

3. Why is sampling without replacement so called?
4. What do you mean by sampling with replacement?
5. What is the main difference between the above two types of sampling?
6. What are the two types of samples?
7. What is meant by a simple random sample?
8. Define the term discrete random variable.

8.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. It is the technique of assigning probabilities on the basis of personal judgement. Such assignment may differ from individual to individual and depends upon the expertise of the person assigning the probabilities.
2. When a random variate can take any value in the given interval $a \le x \le b$, it is a continuous variate and its distribution is a continuous probability distribution.
3. Sampling without replacement is so called because in this process, each person or element can be selected only once and not replaced for another selection.
4. Sampling with replacement means the process where a specific person or element can be selected once and can also be replaced for another selection.
5. Unlike in the case of sampling without replacement, the same element can get reselected in sampling with replacement.
6. The two types of samples are:
   (a) Probability Samples
   (b) Non-Probability Samples
7. A simple random sample is the one in which each and every unit of the population has an equal chance of being selected into the sample.
8. A discrete random variable is the variable that takes values in a finite or countable infinite subset of R. The value of the random variable for the specific range can be denoted by I.

8.6 SUMMARY

- A random variable takes on different values as a result of the outcomes of a random experiment.
- A probability cannot be less than zero or greater than one, i.e., $0 \le pr \le 1$, where pr represents probability.
- The sum of all the probabilities assigned to each value of the random variable must be exactly one.
- A continuous random variable can take all values in a given interval. A continuous probability distribution is represented by a smooth curve.
- The Cumulative Probability Function (CPF) shows the probability that x takes a value less than or equal to, say, z and corresponds to the area under the curve up to z:

$$p(x \le z) = \int_{-\infty}^{z} p(x)dx$$

- A sample is a portion of the total population that is considered for study and analysis.
- Sampling is the process of selecting a sample from the population. It is technically and economically not feasible to take the entire population for analysis.
- There are certain situations in which the piece of paper once selected and taken into consideration is put back into the container in such a manner that the same person has the same chance of being selected again as any other person.
- The third step in the primary data collection process is selecting an adequate sample. It is necessary to take a representative sample from the population, since it is extremely costly, time-consuming and cumbersome to do a complete census.
- If the resources available do not put a heavy constraint on the sample size, a larger sample would be desirable.
- A smaller sample could adequately represent the population, if the population consists of mostly homogeneous units. A heterogeneous universe would require a larger sample.
- Sampling is simply a process of learning about the population on the basis of a sample drawn from it. Thus, in any sampling technique, instead of every unit of the universe, only a part of the universe is studied and the conclusions are drawn on that basis for the entire population.
- Probability sampling methods are those in which every item in the universe has a known chance, or probability of being chosen for the sample.
- Non-probability sampling methods are those which do not provide every item in the universe with a known chance of being included in the sample. The selection process is, at least, partially subjective (dependent on the person making the study).
- Quota sampling is a type of judgement sampling and is perhaps the most commonly used sampling technique in non-probability category.
- The basic objective of a sample is to draw inferences about the population from which such sample is drawn. This means that sampling is a technique which helps us in understanding the parameters or the characteristics of the universe or the population by examining only a small part of it.
- It is understood that the larger the sample size, the closer the sample statistic will be to the population parameter. Hence, the degree of accuracy we require in our estimate would be one factor influencing our choice of sample size.

8.7 KEY WORDS

- **Sample:** A portion of the total population that is considered for study and analysis.
- **Sample size:** It is not possible to consider the entire population to conduct any study or to do any statistical analysis. Hence, the random representative samples are taken for the purpose of analysis. The sample size of 30 or more is considered as a large sample size while below 30 is considered as a small sample size.

8.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is empirical probability assignment?
2. What is meant by cumulative probability function?
3. Explain briefly the different types of sampling.
4. What is sampling distribution? Give examples.
5. Define population in statistical terms.
6. Explain briefly the terms sample and sampling.
7. What is standard error?

Long-Answer Questions

1. Discuss the techniques of assigning probabilities.
2. Describe the continuous probability distributions.
3. Give concrete examples of sampling with replacement and sampling without replacement.
4. A dealer of Toyota cars sold 20,000 Toyota Camry cars last year. He is interested to know if his customers are satisfied with their purchases. 3000 questionnaires were mailed at random to the purchasers. 1600 responses were received. 1440 of these responses indicated satisfaction.
   (a) What is the population of interest?
   (b) What is the sample?
   (c) Is the percentage of satisfied customer a parameter or a statistic?
5. Differentiate between probability samples and non-probability samples. Under what circumstances would non-probability types of samples be more useful in statistical analyses.

NOTES

6. How does sampling with replacement differ from sampling without replacement? Give some examples of situations where sampling has to be done without replacement.
7. Explain in detail the situations that would require:
   (a) Judgement sampling
   (b) Quota sampling
   (c) Stratified sampling
8. Differentiate between sampling errors and non-sampling errors. Under what circumstances would each type of error occur? What steps can be taken to minimize the impact of such errors upon statistical analyses?
9. Your college has a total population of 5000 students. It is desired to estimate the proportion of students who use drugs.
   (a) What type of sampling would be necessary to reach a meaningful conclusion regarding the drug use habits of all students?
   (b) What type of sampling would you select so that the sample is most representative of the population?
   (c) Drug use being a sensitive issue, what type of questions would you include in your questionnaire? What type of questions would you avoid? Give reasons.
10. The lottery method of sample selection is still the most often used method. Discuss this method in detail and give reasons as to why a sample selected by the lottery method would be representative of the population.
11. You are the chairman of the Department of Business Administration and you have been asked to make a report on the current status of students who graduated with B.S. in Business during the two years of 1989 and 1990. Records have indicated that a total of 425 students graduated from the department during these two years. The report is to include information regarding sex of the student, grade point average at the time of graduation, whether the students completed MBA degree or started a job after B.S. degree, current employment position and current annual salary. Prepare a proposal for this survey and include in this proposal:
    (a) Objectives of the survey.
    (b) Type of sampling technique.
    (c) Size of the sample.
12. At New Delhi airport, there is a green channel and a red channel. Passengers without any custom duty articles can go through the green channel. Some passengers are stopped for a random check. What type of random sampling would be appropriate in such situations? Would judgement sampling be more appropriate? Give reasons.

8.9 FURTHER READINGS

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

BLOCK - III
BETA, *t*, *F*, *X* AND ns^2/σ^2 DISTRIBUTIONS

UNIT 9 TRANSFORMATION OF VARIABLE OF THE CONTINUOUS TYPE

Structure

9.0 INTRODUCTION

A random variable has a probability distribution, which specifies the probability of Borel subsets of its range. Random variables can be discrete, that is, taking any of a specified finite or countable list of values (having a countable range), endowed with a probability mass function characteristic of the random variable's probability distribution; or continuous, taking any numerical value in an interval or collection of intervals (having an uncountable range), via a probability density function that is characteristic of the random variable's probability distribution; or a mixture of both types. Two random variables with the same probability distribution can still differ in terms of their associations with, or independence from, other random variables. The realizations of a random variable, that is, the results of randomly choosing values according to the variable's probability distribution function, are called random variates.

In this unit, you will study about the transformation of variable of the continuous type, beta distribution, *t* distribution, *F* Distribution and extensions of the change of variable techniques.

NOTES

9.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the transformation of variable of the continuous type
- Discuss about the beta distribution
- Briefly describe the *t* distribution and *F* distribution

9.2 TRANSFORMATION OF VARIABLE OF THE CONTINUOUS TYPE

A continuous variable is one which can take on an uncountable set of values. A variable over a non-empty range of the real numbers is continuous, if it can take on any value in that range. Methods of calculus are often used in problems in which the variables are continuous, for example in continuous optimization problems. In statistical theory, the probability distributions of continuous variables can be expressed in terms of probability density functions. In continuous-time dynamics, the variable time is treated as continuous, and the equation describing the evolution of some variable over time is a differential equation. The instantaneous rate of change is a well-defined concept.

9.2.1 The Beta Distribution

In probability theory and statistics, the **Beta distribution** is a family of continuous probability distributions defined on the interval [0, 1] parameterized by two positive shape parameters, denoted by α and β that appear as exponents of the random variable and control the shape of the distribution.

The beta distribution has been applied to model the behaviour of random variables limited to intervals of finite length in a wide variety of disciplines. For example, it has been used as a statistical description of allele frequencies in population genetics, time allocation in project management or control systems, sunshine data, variability of soil properties, proportions of the minerals in rocks in stratigraphy and heterogeneity in the probability of HIV transmission.

In Bayesian inference, the Beta distribution is the conjugate prior probability distribution for the Bernoulli, Binomial and Geometric distributions. For example, the Beta distribution can be used in Bayesian analysis to describe initial knowledge concerning probability of success, such as the probability that a space vehicle will successfully complete a specified mission. The Beta distribution is a suitable model for the random behavior of percentages and proportions.

The usual formulation of the Beta distribution is also known as the Beta distribution of the first kind, whereas Beta distribution of the second kind is an alternative name for the Beta prime distribution.

NOTES

The probability density function of the Beta distribution, for $0 \leq x \leq 1$, and shape parameters $\alpha, \beta > 0$, is a power function of the variable x and of its reflection $(1-x)$ like follows:

$$\begin{aligned} f(x;\alpha,\beta) &= \text{constant} \,.\, x^{\alpha-1}(1-x)^{\beta-1} \\ &= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,\mathrm{du}} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\ &= \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \end{aligned}$$

Where, $\Gamma(z)$ is the Gamma function. The Beta function, B, appears to be normalization constant to ensure that the total probability integrates to 1. In the above equations x is a realization — an observed value that actually occurred — of a random process X.

The Cumulative Distribution Function (CDF) is given below:

$$F(x;\alpha,\beta) = \frac{\mathrm{B}(x;\alpha,\beta)}{\mathrm{B}(\alpha,\beta)} = I_x(\alpha,\beta)$$

Where, B (x; α, β) is the incomplete beta function and I_x (α, β) is the regularized incomplete beta function.

The mode of a beta distributed random variable X with $\alpha, \beta > 1$ is given by the following expression:

$$\frac{\alpha-1}{\alpha+\beta-2}.$$

When both parameters are less than one ($\alpha, \beta < 1$), this is the anti-mode - the lowest point of the probability density curve.

The median of the beta distribution is the unique real number $x = I_{\frac{1}{2}}^{[-1]}(\alpha,\beta)$ for which the regularized incomplete beta function $\mathrm{I}_x(\alpha, \beta) = 1/2$, there are no general closed form expression for the median of the beta distribution for arbitrary values of α and β. Closed form expressions for particular values of the parameters a and b follow:

- For symmetric cases $\alpha = \beta$, median = 1/2.
- For $\alpha = 1$ and $\beta > 0$, median = $1-2^{-\frac{1}{\beta}}$ (this case is the mirror-image of the power function [0, 1] distribution).
- For $\alpha > 0$ and $\beta = 1$, median = $2^{-\frac{1}{\alpha}}$ (this case is the power function [0, 1] distribution).

- For $\alpha = 3$ and $\beta = 2$, median = 0.6142724318676105..., the real solution to the quartic equation $1-8x^3+6x^4 = 0$, which lies in [0, 1].
- For $\alpha = 2$ and $\beta = 3$, median = 0.38572756813238945... = 1–median (Beta (3, 2)).

The following are the limits with one parameter finite (non zero) and the other approaching these limits:

$$\lim_{\beta\to 0} \text{median} = \lim_{\alpha\to\infty} \text{median} = 1,$$

$$\lim_{\alpha\to 0} \text{median} = \lim_{\beta\to\infty} \text{median} = 0.$$

A reasonable approximation of the value of the median of the Beta distribution, for both α and β greater or equal to one, is given by the following formula:

$$\text{Median} \approx \frac{\alpha - \frac{1}{3}}{\alpha + \beta - \frac{2}{3}} \text{ for } \alpha, \beta \geq 1.$$

When $\alpha, \beta \geq 1$, the relative error (the absolute error divided by the median) in this approximation is less than 4% and for both $\alpha \geq 2$ and $\beta \geq 2$ it is less than 1%. The absolute error divided by the difference between the mean and the mode is similarly small.

The expected value (mean) (μ) of a beta distribution random variable X with two parameters α and β is a function of only the ratio β/α of these parameters:

$$\begin{aligned}\mu = E[X] &= \int_0^1 xf(x;\alpha,\beta)dx \\ &= \int_0^1 x\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}dx \\ &= \frac{\alpha}{\alpha+\beta} \\ &= \frac{1}{1+\frac{\beta}{\alpha}}\end{aligned}$$

Letting $\alpha = \beta$ in the above expression one obtains $\mu = 1/2$, showing that for $\alpha = \beta$ the mean is at the center of the distribution: it is symmetric. Also, the following limits can be obtained from the above expression:

NOTES

$$\lim_{\frac{\beta}{\alpha} \to 0} \mu = 1$$

$$\lim_{\frac{\beta}{\alpha} \to \infty} \mu = 0$$

Therefore, for $\beta/\alpha \to 0$, or for $\alpha/\beta \to \infty$, the mean is located at the right end, $x = 1$. For these limit ratios, the Beta distribution becomes a one-point degenerate distribution with a Dirac Delta function spike at the right end, $x = 1$, with probability 1 and zero probability everywhere else. There is 100% probability (absolute certainty) concentrated at the right end, $x = 1$.

Similarly, for $\beta/\alpha \to \infty$, or for $\alpha/\beta \to 0$, the mean is located at the left end, $x = 0$. The Beta distribution becomes a 1 point Degenerate distribution with a Dirac Delta function spike at the left end, $x = 0$, with probability 1 and zero probability everywhere else. There is 100% probability (absolute certainty) concentrated at the left end, $x = 0$. Following are the limits with one parameter finite (non zero) and the other approaching these limits:

$$\lim_{\beta \to 0} \mu = \lim_{\alpha \to \infty} \mu = 1$$

$$\lim_{\alpha \to 0} \mu = \lim_{\beta \to \infty} \mu = 0$$

While for typical unimodal distributions with centrally located modes, inflexion points at both sides of the mode and longer tails with Beta (α, β) such that $\alpha, \beta > 2$ it is known that the sample mean as an estimate of location is not as robust as the sample median, the opposite is the case for uniform or 'U-shaped' Bimodal distributions with beta(α, β) such that ($\alpha, \beta \leq 1$), with the modes located at the ends of the distribution.

The logarithm of the Geometric Mean (G_X) of a distribution with random variable X is the arithmetic mean of ln (X), or equivalently its expected value:

$$\ln G_x = E\,[\ln X]$$

For a Beta distribution, the expected value integral gives:

$$\begin{aligned}
\mathrm{E}[\ln X] &= \int_0^1 \ln x f(x;\alpha,\beta)dx \\
&= \int_0^1 \ln x \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} dx \\
&= \frac{1}{\mathrm{B}(\alpha,\beta)} \int_0^1 \frac{\partial x^{\alpha-1}(1-x)^{\beta-1}}{\partial \alpha} dx
\end{aligned}$$

$$= \frac{1}{B(\alpha,\beta)} \frac{\partial}{\partial\alpha} \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

$$= \frac{1}{B(\alpha,\beta)} \frac{\partial B(\alpha,\beta)}{\partial\alpha}$$

$$= \frac{\partial \ln B(\alpha,\beta)}{\partial\alpha}$$

$$= \frac{\partial \ln\Gamma(\alpha)}{\partial\alpha} - \frac{\partial \ln\Gamma(\alpha+\beta)}{\partial\alpha}$$

$$= \Psi(\alpha) - \Psi(\alpha+\beta)$$

Where, ψ is the Digamma Function.

Therefore, the geometric mean of a Beta distribution with shape parameters α and β is the exponential of the Digamma functions of α and β as follows:

$$G_X = e^{E[\ln X]} = e^{\psi(\alpha)-\psi(\alpha+\beta)}$$

While for a Beta distribution with equal shape parameters $\alpha = \beta$, it follows that Skewness = 0 and Mode = Mean = Median = 1/2, the geometric mean is less than 1/2: $0 < G_X < 1/2$. The reason for this is that the logarithmic transformation strongly weights the values of X close to zero, as ln (X) strongly tends towards negative infinity as X approaches zero, while ln (X) flattens towards zero as $X \to 1$.

Along a line $\alpha = \beta$, the following limits apply:

$$\lim_{\alpha=\beta\to 0} G_X = 0$$

$$\lim_{\alpha=\beta\to \infty} G_X = \frac{1}{2}$$

Following are the limits with one parameter finite (non zero) and the other approaching these limits:

$$\lim_{\beta\to 0} G_X = \lim_{\alpha\to\infty} G_X = 1$$

$$\lim_{\alpha\to 0} G_X = \lim_{\beta\to\infty} G_X = 0$$

The accompanying plot shows the difference between the mean and the geometric mean for shape parameters α and β from zero to 2. Besides the fact that the difference between them approaches zero as α and β approach infinity and that the difference becomes large for values of α and β approaching zero, one can observe an evident asymmetry of the geometric mean with respect to the shape parameters α and β. The difference between the geometric mean and the mean is larger for small values of α in relation to β than when exchanging the magnitudes of β and α.

NOTES

The inverse of the Harmonic Mean (H_X) of a distribution with random variable X is the arithmetic mean of $1/X$, or, equivalently, its expected value. Therefore, the Harmonic Mean (H_X) of a beta distribution with shape parameters α and β is:

$$\begin{aligned} H_X &= \frac{1}{\mathrm{E}\left[\frac{1}{X}\right]} \\ &= \frac{1}{\int_0^1 \frac{f(x;\alpha,\beta)}{x}\,dx} \\ &= \frac{1}{\int_0^1 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{x\mathrm{B}(\alpha,\beta)}\,dx} \\ &= \frac{\alpha-1}{\alpha+\beta-1} \text{ if } \alpha>1 \text{ and } \beta>0 \end{aligned}$$

The Harmonic Mean (H_X) of a beta distribution with $\alpha < 1$ is undefined, because its defining expression is not bounded in [0, 1] for shape parameter a less than unity.

Letting $\alpha = \beta$ in the above expression one can obtain the following:

$$H_X = \frac{\alpha-1}{2\alpha-1},$$

Showing that for $\alpha = \beta$ the harmonic mean ranges from 0, for $\alpha = \beta = 1$, to 1/2, for $\alpha = \beta \to \infty$.

Following are the limits with one parameter finite (non zero) and the other approaching these limits:

$$\lim_{\alpha\to 0} H_X = \text{undefined}$$

$$\lim_{\alpha\to 1} H_X = \lim_{\beta\to\infty} H_X = 0$$

$$\lim_{\beta\to 0} H_X = \lim_{\alpha\to\infty} H_X = 1$$

The Harmonic mean plays a role in maximum likelihood estimation for the four parameter case, in addition to the geometric mean. Actually, when performing maximum likelihood estimation for the four parameter case, besides the harmonic mean H_X based on the random variable X, also another harmonic mean appears naturally: the harmonic mean based on the linear transformation (1"X), the mirror image of X, denoted by $H_{(1-X)}$:

$$H_{(1-X)} = \frac{1}{\mathrm{E}\left[\frac{1}{(1-X)}\right]} = \frac{\beta-1}{\alpha+\beta-1} \text{ if } \beta>1, \& \alpha>0.$$

The Harmonic mean ($H_{(1-X)}$) of a Beta distribution with $\beta < 1$ is undefined, because its defining expression is not bounded in [0, 1] for shape parameter β less than unity.

Using $\alpha = \beta$ in the above expression one can obtain the following:

$$H_{(1-X)} = \frac{\beta - 1}{2\beta - 1},$$

This shows that for $\alpha = \beta$ the harmonic mean ranges from 0, for $\alpha = \beta = 1$, to 1/2, for $\alpha = \beta \rightarrow \infty$.

Following are the limits with one parameter finite (non zero) and the other approaching these limits:

$$\lim_{\beta \to 0} H_{(1-X)} = \text{undefined}$$

$$\lim_{\beta \to 1} H_{(1-X)} = \lim_{\alpha \to \infty} H_{(1-X)} = 0$$

$$\lim_{\alpha \to 0} H_{(1-X)} = \lim_{\beta \to \infty} H_{(1-X)} = 1$$

Although both H_X and $H_{(1-X)}$ are asymmetric, in the case that both shape parameters are equal $\alpha = \beta$, the harmonic means are equal: $H_X = H_{(1-X)}$. This equality follows from the following symmetry displayed between both harmonic means:

$$H_X = (\mathrm{B}(\alpha, \beta)) = H_{(1-X)}(\mathrm{B}(\beta, \alpha)) \text{ if } \alpha, \beta > 1.$$

9.2.2 *t* Distribution

Sir William S. Gosset (pen name Student) developed a significance test and through it made significant contribution in the theory of sampling applicable in case of small samples. When population variance is not known, the test is commonly known as Student's *t*-test and is based on the *t* distribution.

Like the normal distribution, *t* distribution is also symmetrical but happens to be flatter than the normal distribution. Moreover, there is a different *t* distribution for every possible sample size. As the sample size gets larger, the shape of the *t* distribution loses its flatness and becomes approximately equal to the normal distribution. In fact, for sample sizes of more than 30, the *t* distribution is so close to the normal distribution that we will use the normal to approximate the *t* distribution. Thus, when *n* is small, the *t* distribution is far from normal, but when *n* is infinite, it is identical with normal distribution.

For applying *t*-test in context of small samples, the *t* value is calculated first of all and, then the calculated value is compared with the table value of *t* at certain level of significance for given degrees of freedom. If the calculated value of *t* exceeds the table value (say $t_{0.05}$), we infer that the difference is significant at 5% level, but if calculated value is t_0, is less than its concerning table value, the difference is not treated as significant.

NOTES

The t-test is used when two conditions are fullfield,

(*i*) The sample size is less than 30, i.e., when $n \le 30$

(*ii*) The population standard deviation (σ_p) must be unknown.

In using the t-test, we assume the following:

(*i*) That the population is normal or approximately normal

(*ii*) That the observations are independent and the samples are randomly drawn samples;

(*iii*) That there is no measurement error;

(*iv*) That in the case of two samples, population variances are regarded as equal if equality of the two population means is to be tested

The following formulae are commonly used to calculate the t value:

(*i*) To Test the Significance of the Mean of a Random Sample

$$t = \frac{|\overline{X} - \mu|}{S \,|\, SE_{\bar{x}} \overline{X}}$$

Where, $\overline{X}$ = Mean of the sample

μ = Mean of the universe

$SE_{\bar{x}}$ = S.E. of mean in case of small sample and is worked out as follows:

$$SE_{\bar{x}} = \frac{\sigma_s}{\sqrt{n}} = \frac{\sqrt{\dfrac{\Sigma(x_i - \bar{x})^2}{\sqrt{n}}}}{\sqrt{n}}$$

and the degrees of freedom = $(n - 1)$

The above stated formula for t can as well be stated as under:

$$t = \frac{|\bar{x} - \mu|}{SE_{\bar{x}}}$$

$$= \frac{|\bar{x} - \mu|}{\dfrac{\dfrac{\sqrt{\Sigma(x - \bar{x})^2}}{n-1}}{\sqrt{n}}}$$

$$= \frac{|\bar{x} - \mu|}{\sqrt{\dfrac{\Sigma(x - \bar{x})^2}{n-1}}} \times \sqrt{n}$$

If we want to work out the probable or fiducial limits of population mean (μ) in case of small samples, we can use either of the following:

(a) Probable limits with 95% confidence level:

$$\mu = \overline{X} \pm SE_{\overline{x}}\ (t_{0.05})$$

(b) Probable limits with 99% confidence level:

$$\mu = \overline{X} \pm SE_{\overline{x}}\ (t_{0.01})$$

At other confidence levels, the limits can be worked out in a similar manner, taking the concerning table value of t just as we have taken $t_{0.05}$ in (a) and $t_{0.01}$ in (b) above.

(*ii*) To Test the Difference between the Means of the Two Samples

$$t = \frac{|\overline{X}_1 - \overline{X}_2|}{\text{SE}_{\overline{x}_1 - \overline{x}_2}}$$

Where, $\overline{X}_1$ = Mean of the sample 1

$\overline{X}_2$ = Mean of the sample 2

$SE_{\overline{x}_1 - \overline{x}_2}$ = Standard Error of difference between two sample means and is worked out as follows:

$$SE_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{\sum (X_{1i} - \overline{x}_1)^2 + \sum (X_{2\,i} - \overline{x}_2)^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

and the degrees of freedom = $(n_1 + n_2 - 2)$.

When the actual means are in fraction, then use of assumed means is convenient. In such a case, the standard deviation of difference, i.e.,

$$\sqrt{\frac{\Sigma (x_{1i} + x_1)^2 + \Sigma (x_{2i} - \overline{x}_2)^2}{n_1 + n_2 - 2}}$$

Can be worked out by the following short-cut formula:

$$= \frac{\sqrt{\Sigma (x_{1i} - A_1)^2} + \Sigma (x_{2i} - A_1)^2 - n_1 (x_{1i} - A_2)^2 - n_2 (x_{2i} - A_2)^2}{n_1 + n_2 - 2}$$

Where, A_1 = Assumed mean of sample 1

A_2 = Assumed mean of sample 2

X_1 = True mean of sample 1

X_2 = True mean of sample 2

NOTES

(*iii*) To Test the Significance of an Observed Correlation Coefficient

$$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{n-2}$$

Here, t is based on $(n-2)$ degrees of freedom.

(*iv*) In Context of the 'Difference Test'.

Difference test is applied in the case of paired data and in this context t is calculated as under:

$$t = \frac{\bar{x}_{Ditt} - 0}{6_{Diff}\sqrt{n}} = \frac{\bar{x}_{Diff} - 0}{6_{Diff}}\sqrt{n}$$

Where, $\bar{X}_{Diff}$ or $\bar{D}$ = Mean of the differences of sample items.

0 = the value zero on the hypothesis that there is no difference

$\sigma_{\text{Diff.}}$ = standard deviation of difference and is worked out as

$$\sqrt{\frac{\sum D - \bar{X}_{Diff})^2}{(n-1)}}$$

or

$$\sqrt{\frac{\Sigma D^2 - (\bar{D})^2 n}{(n-1)}}$$

D = differences

n = number of pairs in two samples and is based on $(n-1)$ degrees of freedom.

The following examples would illustrate the application of t-test using the above stated formulae.

Example 1: A sample of 10 measurements of the diameter of a sphere, gave a mean $X = 4.38$ inches and a standard deviation, $\sigma = 0.06$ inches. Find (*a*) 95% and (*b*) 99% confidence limits for the actual diameter.

Solution: On the basis of the given data the standard error of mean

$$= \frac{\sigma_s}{\sqrt{n-1}} = \frac{0.06}{\sqrt{10-1}} = \frac{0.06}{3} = 0.02$$

Assuming the sample mean 4.38 inches to be the population mean, the required limits are as follows:

(*i*) 95% confidence limits $= \bar{X} \pm SE_{\bar{x}}(t_{0.05})$ with degrees of freedom

$= 4.38 \pm .02(2.262)$

$$= 4.38 \pm .04524$$

i.e., 4.335 to 4.425

(*ii*) 99% confidence limits $= \overline{X} \pm SE_{\overline{x}}(t_{0.01})$ with 9 degrees of freedom

$$= 4.38 \pm .02(3.25) = 4.38 \pm .0650$$

i.e., 4.3150 to 4.4450.

Example 2: The specimen of copper wires drawn from a large lot have the following breaking strength (in kg. wt.):

578, 572, 570, 568, 572, 578, 570, 572, 596, 544

Test whether the mean breaking strength of the lot may be taken to be 578 kg. wt.

Solution: We take the hypothesis that there is no difference between the mean height of the sample and the given height of universe. In other words we can write, $H_0 : \mu = \overline{X},\ H_0 : \mu \neq \overline{X}$. Then on the basis of the sample data, the mean and standard deviation has been worked out as under:

S. No.	*X*	$(X - \overline{X})$	$(X_1 - \overline{X})^2$
1	578	6	36
2	572	0	0
3	570	–2	4
4	568	–4	16
5	572	0	0
6	578	6	36
7	570	–2	4
8	572	0	0
9	596	24	576
10	544	–28	784
$n = 10$	$\sum X_i = 5720$		$\sum (X_i - \overline{X})^2 = 1456$

$$\overline{X} = \frac{\Sigma x}{n} = \frac{5720}{10}$$

$$= 572$$

$$\sigma_s = \sqrt{\frac{\Sigma(x - \overline{x}_s)^2}{n-1}}$$

$$= \sqrt{\frac{1456}{10-1}} = \sqrt{\frac{1456}{9}}$$

$$= 12.72$$

NOTES

$$SE_x = \frac{\sigma_s}{\sqrt{n}} = \frac{12.72}{\sqrt{10}}$$

$$= \frac{12.72}{3.16} = 4.03$$

$$t = \frac{|\bar{x} - \mu|}{SE_x} = \frac{|572 - 578|}{4.03}$$

$$= 1.488$$

Degrees of freedom = $n - 1 = 9$

At 5% level of significance for 9 degrees of freedom, the table value of $t = 2.262$. For a two-tailed test.

The calculated value of t is less than its table value and hence the difference is insignificant. The mean breaking strength of the lot may be taken to be 578 Kg. wt. with 95% confidence level.

Example 3: Sample of sales in similar shops in two towns are taken for a new product with the following results:

	Mean sales	Variance	'Size of sample
Town *A*	57	5.3	5
Town *B*	61	4.8	7

Is there any evidence of difference in sales in the two towns?

Solution: We take the hypothesis that there is no difference between the two sample means concerning sales in the two towns. In other words, $H_0 : \bar{X}_1 = \bar{X}_2,\ H_0 : \bar{X}_1 \neq \bar{X}_2$. Then, we work out the concerning t value as follows:

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{SE_{\bar{x}_1 - \bar{x}_2}}$$

Where, $\bar{x}_1$ = Mean of the sample concerning Town *A*

$\bar{x}_2$ = Mean of the sample concerning Town *B*

$SE_{\bar{x}_1 - \bar{x}_2}$ = Standard Error of the difference between two means.

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\Sigma(x_{1i} - \bar{x}_1)^2 + \Sigma(x_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Hence,

$$t = \frac{|57 - 61|}{1.421} = \frac{4}{1.421}$$

$$= 2.82$$

Degrees of freedom = $(n_1 + n_2 - 2) = (5 + 7 - 2) = 10$

Table value of t at 5% level of significance for 10 degrees of freedom is 2.228, for a two-tailed test.

The calculated value of t is greater than its table value. Hence, the hypothesis is wrong and the difference is significant.

Example 4: The sales data of an item in six shops before and after a special promotional campaign are:

Shops	*A*	*B*	*C*	*D*	*E*	*F*
Before the promotional campaign	53	28	31	48	50	42
After the campaign	58	29	30	55	56	45

Can the campaign be judged to be a success? Test at 5% level of significance.

Solution: We take the hypothesis that the campaign does not bring any improvement in sales. We can thus write:

In order to judge this, we apply the 'difference test'. For this purpose we calculate the mean and standard deviation of differences in two sample items as follows:

Shops	*Sales before campaign* X_{Bi}	*Sales after campaign* X_{Ai}	*Difference* = *D* (*i.e., increase or decrease after the campaign*)	$(D - \bar{D})$	$(D - \bar{D})^2$
A	53	58	+5	+1.5	2.25
B	28	29	+1	–2.5	6.25
C	31	30	–1	–4.5	20.25
D	48	55	+7	+3.5	12.25
E	50	56	+6	+2.5	6.25
F	42	45	+3	–0.5	0.25
$n = 6$			$\Sigma D = 21$		$\Sigma(D - \bar{D})^2 = 47.50$

Mean of difference or $\bar{X}_{Diff} = \dfrac{\Sigma D}{n} = \dfrac{21}{6} = 3.5$

Standard deviation of difference

$$\sigma_{Diff} = \sqrt{\frac{\Sigma(D - \bar{D})^2}{n-1}} = \sqrt{\frac{47.50}{6-1}} = 3.08$$

$$t = \frac{\bar{X}_{Diff} - 0}{\sigma_{Diff}} = \sqrt{n}$$

$$= 1.14 \times 2.45 = 2.793$$

Degrees of freedom = $(n - 1) = (6 - 1) = 5$

Table value of t at 5% level of significance for 5 degrees of freedom = 2.015 for one-tailed test.

Since, the calculated value of t is greater than its table value, the difference is significant. Thus, the hypothesis is wrong and the special promotional campaign can be taken as a success.

Example 5: Memory capacity of 9 students was tested before and after training. From the following scores, state whether the training was effective or not.

Student	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Before (X_{Bi})	10	15	9	3	7	12	16	17	4
After (X_{Ai})	12	17	8	5	6	11	18	20	3

Solution: We take the hypothesis that training was not effective. We can write, $H_0 : \bar{x}_A = \bar{X}_B,\ H_0 : \bar{X} > \bar{X}_B$. We apply the difference test for which purpose first of all we calculate the mean and standard deviation of difference as follows:

Students	*Before* X_{Bi}	*After* X_{Ai}	*Difference* = *D*	D^2
1	10	12	2	4
2	15	17	2	4
3	9	8	–1	1
4	3	5	2	4
5	7	6	–1	1
6	12	11	–1	1
7	16	18	2	4
8	17	20	3	9
9	4	3	–1	1
$n = 9$			$\sum D = 7$	$\sum D^2 = 29$

$$\bar{D} = \frac{\Sigma D}{n} = \frac{7}{9} = 0.78$$

$$\sigma_{Diff} = \sqrt{\frac{\Sigma D^2 - (\bar{D})^2 n}{n-1}} = \sqrt{\frac{29 - (0.78)^2 \times 9}{9-1}} = 1.71$$

$$\therefore\ t = \frac{0.78}{1-71} = 1.369$$

Degrees of freedom = $(n - 1) = (9 - 1) = 8$

Table value of t at 5% level of significance for 8 degrees of freedom

= 1.860 for one-tailed test.

Since the calculated value of t is less than its table value, the difference is insignificant and the hypothesis is true. Hence it can be inferred that the training was not effective.

Example 6: It was found that the coefficient of correlation between two variables calculated from a sample of 25 items was 0.37. Test the significance of *r* at 5% level with the help of *t*-test.

Solution: To test the significance of *r* through *t*-test, we use the following formula for calculating *t* value:

$$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{n-2}$$

$$= \frac{0.37}{1-(0.37)^2} \times \sqrt{25-2}$$

$$= 1.903$$

Degrees of freedom = (n–2) = (25–2) = 23

The table value of α at 5% level of significance for 23 degrees of freedom is 2.069 for a two-tailed test.

The calculated value of *t* is less than its table value, hence *r is* insignificant.

Example 7: A group of seven week old chickens reared on high protein diet weigh 12, 15, 11, 16, 14, 14 and 16 ounces; a second group of five chickens similarly treated except that they receive a low protein diet weigh 8, 10, 14, 10 and 13 ounces. Test at 5% level whether there is significant evidence that additional protein has increased the weight of chickens. (Use assumed mean (or A_1) = 10 for the sample of 7 and assumed mean (or A_2) = 8 for the sample of 5 chickens in your calculation).

Solution: We take the hypothesis that additional protein has not increased the weight of the chickens. We can write, $H_o: X_1 > X_2$ $H_o: X_1 > X_2$.

Applying *t*-test, we work out the value of *t* for measuring the significance of two sample means as follows:

$$t = \frac{X_1 - X_2}{SE_{x_1 - x_2}}$$

Calculation of can be done as under:

X_1	$(X_{i1} - A_1)$ $A_1 = 10$	$(X_{i1} - A_1)^2$	X_2	$(X_{i2} - A_2)$ $A_2 = 8$	$(X_{i2} - A_2)^2$
12	2	4	8	0	0
15	5	25	10	2	4
11	1	1	14	6	36
16	6	36	10	2	4
14	4	16	13	5	25
14	4	16			
16	6	36			
n_1=7	$\Sigma(X_{1i} - A_1)$ =28	$\Sigma(X_{1i} - A_1)^2$ =134	n_2=5	$\Sigma(X_{2i} - A_2)$ =15	$\Sigma(X_{2i} - A_2)^2$ = 69

NOTES

$$X_1 = A_1 + \Sigma(x_{1i} - A_1)/n_1$$

$$\therefore \quad =10 + \frac{28}{7} = 14$$

Similarly, $$X_2 = A_1 + \frac{\Sigma(x_{2i} - A_2)}{n_2}$$

$$= 8 + \frac{15}{5} = 11$$

Hence,

$$SE_{X_1-X_2} = \sqrt{\frac{\Sigma(X_{1i} - A_1)^2 + \Sigma(X_{2i} - A_2)^2 - n_1(\bar{X}_1 - A_1)^2 - n_2(\bar{X}_2 - A_2)^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= \sqrt{\frac{134 + 69 - 7\,(14 - 10)^2 - 5(11-8)^2}{7+5-2}} \times \sqrt{\frac{1}{7} + \frac{1}{5}}$$

$$= (2.14)\,(.59) = 1.2626$$

We now calculate the value under *t*,

$$t = \frac{X_1 - X_2}{SE_{X_1-X_2}} = \frac{14-11}{1.2626} = 2.376$$

Degree of freedom = $(n_1 + n_2 - 2) = (7 + 5 - 2) = 10$

The table value of *t* at 5% level of significance for 10 degrees of freedom = 1.812 for one-tailed test.

The calculated value of *t* is higher than its table value and hence the difference is significant, which means the hypothesis is wrong. It can therefore be concluded that additional protein has increased the weight of chickens.

9.2.3 *F* Distribution

In business decisions, we are often involved in determining if there are significant differences among various sample means, from which conclusions can be drawn about the differences among various population means. In the previous chapters, we discussed and evaluated the differences between two sample means. But, what if we have to compare more than 2 sample means? For example, we may be interested to find out if there are any significant differences in the average sales figures of 4 different salesman employed by the same company, or we may be interested to find out if the average monthly expenditures of a family of 4 in 5 different localities are similar or not, or the telephone company may be interested in checking, whether there are any significant differences in the average number of requests for information received in a given day among the 5 areas of New York

City, and so on. The methodology used for such types of determinations is known as *Analysis of Variance.*

This technique is one of the most powerful techniques in statistical analysis and was developed by R.A. Fisher. It is also called the *F*-Test.

There are two types of classifications involved in the analysis of variance. The one-way analysis of variance refers to the situations when only one fact or variable is considered. For example, in testing for differences in sales for three salesman, we are considering only one factor, which is the salesman's selling ability. In the second type of classification, the response variable of interest may be affected by more than one factor. For example, the sales may be affected not only by the salesman's selling ability, but also by the price charged or the extent of advertising in a given area.

For the sake of simplicity and necessity, our discussion will be limited to *One-way Analysis of Variance*.

The null hypothesis, that we are going to test, is based upon the assumption that there is no significant difference among the means of different populations. For example, if we are testing for differences in the means of k populations, then,

$$H_0 = \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$$

The alternate hypothesis (H_1) will state that at least two means are different from each other. In order to accept the null hypothesis, all means must be equal. Even if one mean is not equal to the others, then we cannot accept the null hypothesis. The simultaneous comparison of several population means is called *Analysis of Variance or ANOVA.*

Assumptions

The methodology of ANOVA is based on the following assumptions.

(*i*) Each sample of size n is drawn randomly and each sample is independent of the other samples.

(*ii*) The populations are normally distributed.

(*iii*) The populations from which the samples are drawn have equal variances. This means that:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = ...= \sigma_k^2, \text{ for } k \text{ populations.}$$

The Rationale Behind Analysis of Variance

Why do we call it the *Analysis of Variance,* even though we are testing for *means*? Why not simply call it the *Analysis of Means*? How do we test for means by analysing the variances? As a matter of fact, in order to determine if the means of several populations are equal, we do consider the measure of variance, σ^2.

The estimate of population variance, σ^2, is computed by two different estimates of σ^2, each one by a different method. One approach is to compute an estimator of σ^2 in such a manner that even if the population means are not equal, it will have no effect on the value of this estimator. This means that, the differences in the values of the population means do not alter the value of σ^2 as calculated by a given method. This estimator of σ^2 is the average of the variances found within each of the samples. For example, if we take 10 samples of size n, then each sample will have a mean and a variance. Then, the mean of these 10 variances would be considered as an unbiased estimator of σ^2, the population variance, and its value remains appropriate irrespective of whether the population means are equal or not. This is really done by pooling all the sample variances to estimate a common population variance, which is the average of all sample variances. This common variance is known as variance within samples or σ^2_{within}.

The second approach to calculate the estimate of σ^2, is based upon the Central Limit Theorem and is valid only under the null hypothesis assumption that all the population means are equal. This means that in fact, if there are no differences among the population means, then the computed value of σ^2 by the second approach should not differ significantly from the computed value of σ^2 by the first approach.

Hence, If these two values of σ^2 are approximately the same, then we can decide to accept the null hypothesis.

The second approach results in the following computation.

Based upon the Central Limit Theorem, we have previously found that the standard error of the sample means is calculated by:

$$\sigma^2_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

or, the variance would be:

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$$

or, $\quad \sigma^2 = n\sigma^2_{\bar{x}}$

Thus, by knowing the square of the standard error of the mean $(\sigma_{\bar{x}})^2$, we could multiply it by n and obtain a precise estimate of σ^2. This approach of estimating σ^2 is known as $\sigma^2_{\text{between}}$. Now, if the null hypothesis is true, that is if all population means are equal then,

$\sigma^2_{\text{between}}$ value should be approximately the same as σ^2_{within} value. A significant difference between these two values would lead us to conclude that this difference is the result of differences between the population means.

But, how do we know that any difference between these two values is significant or not? How do we know whether this difference, if any, is simply due to random sampling error or due to actual differences among the population means?

NOTES

R.A. Fisher developed a Fisher test or F-test to answer the above question. He determined that the difference between $\sigma^2_{\text{between}}$ and σ^2_{within} values could be expressed as a ratio to be designated as the F-value, so that:

$$F = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}}$$

In the above case, if the population means are exactly the same, then $\sigma^2_{\text{between}}$ will be equal to the σ^2_{within} and the value of F will be equal to 1.

However, because of sampling errors and other variations, some disparity between these two values will be there, even when the null hypothesis is true, meaning that all population means are equal. The extent of disparity between the two variances and consequently, the value of F, will influence our decision on whether to accept or reject the null hypothesis. It is logical to conclude that, if the population means are not equal, then their sample means will also vary greatly from one another, resulting in a larger value of $\sigma^2_{\text{between}}$ and hence a larger value of F (σ^2_{within} is based only on sample variances and not on sample means and hence, is not affected by differences in sample means). Accordingly, the larger the value of F, the more likely the decision to reject the null hypothesis. But, how large the value of F be so as to reject the null hypothesis? The answer is that the computed value of F must be larger than the *critical* value of F, given in the table for a given level of significance and calculated number of degrees of freedom. (The F distribution is a family of curves, so that there are different curves for different degrees of freedom).

Degrees of Freedom

We have talked about the F-distribution being a family of curves, each curve reflecting the degrees of freedom relative to both $\sigma^2_{\text{between}}$ and σ^2_{within}. This means that, the degrees of freedom are associated both with the numerator as well as with the denominator of the F-ratio.

(*i*) **The Numerator:** Since the variance between samples, $\sigma^2_{\text{between}}$ comes from many samples and if there are k number of samples, then the degrees of freedom, associated with the numerator would be $(k-1)$.

(*ii*) **The Denominator:** It is the *mean variance* of the variances of k samples and since, each variance in each sample is associated with the size of the sample (n), then the degrees of freedom associated with each sample would be $(n-1)$. Hence, the total degrees of freedom would be the sum of degrees of freedom of k samples or

$df = k(n-1)$, when each sample is of size n.

NOTES

The *F*-Distribution

The major characteristics of the *F*-distribution are as follows:

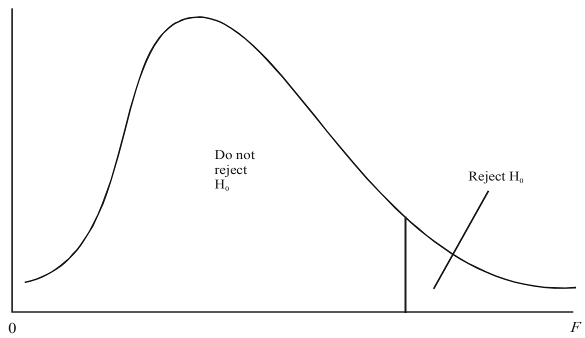
(*i*) Unlike normal distribution, which is only one type of curve irrespective of the value of the mean and the standard deviation, the *F* distribution is a *family* of curves. A particular curve is determined by two parameters. These are the degrees of freedom in the numerator and the degrees of freedom in the denominator. The shape of the curve changes as the number of degrees of freedom changes.

(*ii*) It is a continuous distribution and the value of *F* cannot be negative.

(*iii*) The curve representing the *F* distribution is positively skewed.

(*iv*) The values of *F* theoretically range from zero to infinity.

A diagram of *F* distribution curve is shown below.



The rejection region is only in the right end tail of the curve because unlike *Z* distribution and *t* distribution which had negative values for areas below the mean, *F* distribution has only positive values by definition and only positive values of *F* that are larger than the critical values of *F*, will lead to a decision to reject the null hypothesis.

Computation of *F*

Since *F* ratio contains only two elements, which are the variance between the samples and the variance within the samples, the concepts of which have been discussed before, let us recapitulate the calculation of these variances.

If all the means of samples were exactly equal and all samples were exactly representative of their respective populations so that all the sample means, were exactly equal to each other and to the population mean, then there will be no variance. However, this can never be the case. We always have variation, both between samples and within samples, even if we take these samples randomly and from the same population. This variation is known as the *total variation.*

The total variation designated by $\sum(X - \bar{\bar{X}})^2$, where *X* represents individual observations for all samples and $\bar{\bar{X}}$ is the grand mean of all sample means and equals (μ), the population mean, is also known as the *total sum of squares* or *SST,* and is simply the sum of squared differences between each observation and the overall mean. This total variation represents the contribution of two elements. These elements are:

(A) Variance between Samples: The variance between samples may be due to the effect of different *treatments,* meaning that the population means may be affected by the *factor* under consideration, thus, making the population means actually different, and some variance may be due to the inter-sample variability. This variance is also known as the *sum of squares* between samples. Let this *sum of squares* be designated as *SSB*.

Then, *SSB* is calculated by the following steps:

(*i*) Take k samples of size n each and calculate the mean of each sample, i.e., $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots \bar{X}_k$.

(*ii*) Calculate the grand mean $\bar{\bar{X}}$ of the distribution of these sample means, so that,

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{k} \bar{x}_i}{k}$$

(*iii*) Take the difference between the means of the various samples and the grand mean, i.e.,

$$(\bar{X}_1 - \bar{\bar{X}}),\ (\bar{X}_2 - \bar{\bar{X}}),\ (\bar{X}_3 - \bar{\bar{X}}),\ \ldots,\ (\bar{X}_k - \bar{\bar{X}})$$

(*iv*) Square these deviations or differences individually, multiply each of these squared deviations by its respective sample size and sum up all these products, so that we get;

$$\sum_{i=1}^{k} n_i (\bar{X}_i - \bar{\bar{X}})^2, \text{ where } n_i = \text{size of the } i\text{th sample.}$$

This will be the value of the *SSB*.

However, if the individual observations of all samples are not available, and only the various means of these samples are available, where the samples are either of the same size n or different sizes, $n_i, n_2, n_3, \ldots, n_k$, then the value of SSB can be calculated as:

$$SSB = n_i(\bar{X}_i - \bar{\bar{X}})^2 + n_2(\bar{X}_2 - \bar{\bar{X}})^2 + \ \ldots\ n_k(\bar{X}_k - \bar{\bar{X}})^2$$

where,

n_1 = Number of items in sample 1

n_2 = Number of items in sample 2

n_k = Number of items in sample k

NOTES

$\bar{X}_1$ = Mean of sample 1

$\bar{X}_2$ = Mean of sample 2

$\bar{X}_k$ = Mean of sample k

$\bar{\bar{X}}$ = Grand mean or average of all items in all samples.

(v) Divide SSB by the degrees of freedom, which are (k – 1), where k is the number of samples and this would give us the value of $\sigma^2_{\text{between}}$, so that,

$$\sigma^2_{\text{between}} = \frac{SSB}{(k-1)}.$$

(This is also known as mean square between samples or *MSB)*.

(B) Variance within Samples: Even though each observation in a given sample comes from the same population and is subjected to the same treatment, some chance variation can still occur. This variance may be due to sampling errors or other natural causes. This variance or sum of squares is calculated through the following steps:

(i) Calculate the mean value of each sample, i.e., $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots \bar{X}_k$.

(ii) Take one sample at a time and take the deviation of each item in the sample from its mean. Do this for all the samples, so that we would have a difference between each value in each sample and their respective means for all values in all samples.

(iii) Square these differences and take a total sum of all these squared differences (or deviations). This sum is also known as SSW or sum of squares within samples.

(iv) Divide this SSW by the corresponding degrees of freedom. The degrees of freedom are obtained by subtracting the total number of samples from the total number of items. Thus, if N is the total number of items or observations, and k is the number of samples, then,

$$df = (N - k)$$

These are the degrees of freedom within samples. (If all samples are of equal size n, then $df = k(n-1)$, since $(n-1)$ are the degrees of freedom for each sample and there are k samples).

(v) This figure SSW/df, is also known as σ^2_{within}, or MSW (mean of sum of squares within samples).

Now, the value of F can be computed as:

$$F = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}} = \frac{SSB/df}{SSW/df}$$

$$= \frac{SSB/(k-1)}{SSW/(N-k)} = \frac{MSB}{MSW}$$

This value of F is then compared with the critical value of F from the table and a decision is made about the validity of null hypothesis.

ANOVA Table

After various calculations for SSB, SSW and the degrees of freedom have been made, these figures can be presented in a simple table called *Analysis of Variance* table or simply ANOVA table, as follows:

ANOVA Table

Source of Variation	*Sum of Squares*	*Degrees of freedom*	*Mean Square*	*F*
Treatment	SSB	$(k-1)$	$MSB=\frac{SSB}{(k-1)}$	$\frac{MSB}{MSW}$
Within	SSW	$(N-k)$	$MSW=\frac{SSW}{(n-k)}$	
Total	SST			

Then,

$$F=\frac{MSB}{MSW}$$

A Short-Cut Method

The formula developed above for the computation of the values of F-statistic is rather complex and time consuming when we have to calculate the variance between samples and the variance within samples. However, a short-cut, simpler method for these *sum of squares* is available, which considerably reduces the computational work. This technique is used through the following steps:

(*i*) Take the sum of all the observations of all the samples, either by adding all the individual values, or by multiplying the mean of each sample by its size and then adding up all these products as follows:

$$\text{The Total Sum } TS=n_1\overline{X}_1+n_2\overline{X}_2+....n_k\overline{X}_k,\ \text{ for k samples}$$

(*ii*) Calculate the value of a *correction factor*. The correction factor (CF) value is obtained by squaring the total sum obtained above and dividing it by the total number of observations N, so that:

$$CF=\frac{(TS)^2}{N}$$

(*iii*) The total sum of squares is obtained by squaring all individual observations of all samples, summing up these values and subtracting from this sum, the correction factor (CF).

In other words:

$$\text{Total sum of squares } SST=\Sigma X_1^2+\Sigma X_2^2+....+\Sigma X_k^2-\frac{(TS)^2}{N}$$

NOTES

Where,

$\Sigma X_1^2 =$ Summation of squares for all X's in sample 1.

$\Sigma X_2^2 =$ Summation of squares for all X's in sample 2.

:

:

$\Sigma X_k^2 =$ Summation of squares for all X's in sample k.

(*iv*) The sum of squares between the samples (*SSB*) is obtained by the following formula:

$$SSB = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + ... + \frac{(\Sigma X_k)^2}{n_k} - \frac{(TS)^2}{N}$$

Where,

$(\Sigma X_1)^2 =$ Square of the total of all values in sample 1.

$(\Sigma X_2)^2 =$ Square of the total of all values in sample 2.

$(\Sigma X_k)^2 =$ Square of the total of all values in sample k.

(*v*) Then sum of squares within samples *SSW* can be calculated as:

$SSW =$ Total sum of squares minus the sum of squares between samples

$= SST - SSB$

(*vi*) The rest of the procedure is similar to the previous method.

Example 8: To test whether all professors teach the same material in different sections of the introductory statistics class or not, four sections of the same course were selected and a common test was administered to five students selected at random from each section. The scores for each student from each section were noted and are given below. We want to test for any differences in learning, as reflected in the average scores for each section.

Student #	*Section 1 Scores (X_1)*	*Section 2 Scores (X_2)*	*Section 3 Scores (X_3)*	*Section 4 Scores (X_4)*
1	8	12	10	12
2	10	12	13	15
3	12	10	11	13
4	10	8	12	10
5	5	13	14	10
Totals	$\Sigma X_1 = 45$	$\Sigma X_2 = 55$	$\Sigma X_3 = 60$	$\Sigma X_4 = 60$
Means	$\overline{X}_1 = 9$	$\overline{X}_2 = 11$	$\overline{X}_3 = 12$	$\overline{X}_4 = 12$

Solution: A. The Traditional Method

(*i*) State the null hypothesis. We are assuming that there is no significant difference among the average scores of students from these four sections and hence, all professors are teaching the same material with the same effectiveness, i.e.,

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

H_1: All means are not equal or at least two means differ from each other

(*ii*) Establish a level of significance. Let $a = 0.05$.

(*iii*) Calculate the variance between the samples, as follows:

(a) The mean of each sample is:

$$\overline{X}_1 = 9,\ \overline{X}_2 = 11,\ \overline{X}_3 = 12, \overline{X}_4 = 12$$

(b) The grand mean or $\overline{\overline{X}}$ is:

$$\overline{\overline{X}} = \frac{\Sigma \overline{x}}{n} = \frac{9+11+12+12}{4}$$
$$= 11$$

(c) Calculate the value of *SSB*:

$$SSB = \Sigma n(\overline{X} - \overline{\overline{X}})^2$$
$$= 5\ (9-11)^2 + 5\ (11-11)^2 + 5\ (12-11)^2 + 5\ (12-11)^2$$
$$= 20 + 0 + 5 + 5$$
$$= 30$$

(d) The variance between samples $\sigma^2_{\text{between}}$ or *MSB* is given by:

$$MSB = \frac{SSB}{df} = \frac{(30)}{(k-1)} = \frac{(30)}{3} = 10$$

(*iv*) Calculate the variance within samples, as follows:

To find the sum of squares within samples (*SSW*), we square each deviation between the individual value of each sample and its mean, for all samples and then sum these squared deviations, as follows:

Sample 1: $\overline{X}_1 = 9$

$$\Sigma(X_1 - \overline{X}_1)^2 = (8-9)^2 + (10-9)^2 + (12-9)^2 + (10-9)^2 + (5-9)^2$$
$$= 1 + 1 + 9 + 1 + 16$$
$$= 28$$

Sample 2: $\overline{X}_2 = 11$

$$\Sigma(X_2 - \overline{X}_2)^2 = (12-11)^2 + (12-11)^2 + (10-11)^2 + (8-11)^2 + (13-11)^2$$
$$= 1+1+1+9+4$$
$$= 16$$

NOTES

Sample 3: $\overline{X}_3 = 12$

$$\Sigma(X_3 - \overline{X}_3)^2 = (10-12)^2 + (13-12)^2 + (11-12)^2 + (12-12)^2 + (14-12)^2$$
$$= 4+1+1+0+4$$
$$= 10$$

Sample 4: $\overline{X}_4 = 12$

$$\Sigma(X_4 - \overline{X}_4)^2 = (12-12)^2 + (15-12)^2 + (13-12)^2 + (10-12)^2 + (10-12)^2$$
$$= 0+9+1+4+4$$
$$= 18$$

Then, $SSW = 28 + 16 + 10 + 18$
$= 72$

Now, the variance within samples, σ^2_{within}, or MSW is given by:

$$MSW = \frac{SSW}{df} = \frac{SSW}{(N-k)} = \frac{72}{20 \ - 4} = \frac{72}{16} = 4.5$$

Then, the F-ratio $= \dfrac{MSB}{MSW} = \dfrac{10}{4.5} = 2.22.$

Now, we check for the critical value of F from the table for $\alpha = 0.05$ and degrees of freedom as follows:

df (numerator) = $(k - 1) = (4 - 1) = 3$

df (denominator) = $(N - k) = (20 - 4) = 16$

This value of F from the table is given as 3.24. Now, since our calculated value of $F = 2.22$ is less than the critical value of $F = 3.24$, we cannot reject the null hypothesis.

B. The Short-Cut Method

Following the procedure outlined before for using the short-cut method, we get:

(*i*) Total sum (TS) = ΣX
= 220

(*ii*) Correction before $CF = \dfrac{(TS)^2}{N} = \dfrac{(220)^2}{20} = 2420$

(*iii*) Total sum of squares:

$$SST = \Sigma(X^2) - CF$$
$$= 2522 \ - \ 2420 \ - 102$$

NOTES

(iv) Sum of squares betwen the samples SSB is obtained by:

$$\begin{aligned} SSB &= \sum_{i=1}^{k} \frac{(X_i)^2}{n_i} - CF \\ &= \frac{(X_1)^2}{n_1} + \frac{(X_2)^2}{n_2} + \ .... \ + \frac{(X_k)^2}{n_k} - CF \\ &= \frac{(45)^2}{5} + \frac{(55)^2}{5} + \frac{(60)^2}{5} + \frac{(60)^2}{5} - (2420) \\ &= 405 + 605 + 720 + 720 - 2420 \\ &= 30 \end{aligned}$$

(v) SSW can be calculated by:

$$SST - SSB = 102 - 30 = 72$$

Now the F value can be calculated as:

$$F = \frac{SSB / df}{SSW / df} = \frac{30/(k-1)}{72/(n-k)} = \frac{30/3}{72/16} = \frac{10}{4.5}$$
$$= 2.22$$

As we see, we get the same value of F as obtained by the traditional method. So, we compare our value of F with the critical value of F from the table for $\alpha = 0.05$ and df (numerator = 3), and df (denominator = 16), and we get the critical value of F as 3.24. As before, we accept the null hypothesis.

The ANOVA Table

We can construct an ANOVA table for the problem solved above as follows:

ANOVA Table

Source of Variation	*Sum of Squares*	*Degrees of freedom*	*Mean Square*	*F*
Treatment	$SSB = 30$	$(k-1) = 3$	$MSB = \frac{SSB}{(k-1)}$	$\frac{MSB}{MSW}$
			$= \frac{30}{3} = 10$	$= \frac{10}{4.5}$
Within (or error)	$SSW = 72$	$(N-k) = 16$	$MSW = \frac{SSW}{(N-k)}$	$= 2.22$
			$= \frac{72}{16} = 4.5$	
Total	$SST = 102$			

NOTES

Check Your Progress

1. Give the value of the median of the beta distribution.
2. Who developed t-test? When it is used?
3. On what assumptions the ANOVA methodology is based?
4. What are the major characteristics of f-distribution?

9.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A reasonable approximation of the value of the median of the Beta distribution, for both α and β greater or equal to one, is given by the following formula:

   $$\text{Median} \approx \frac{\alpha - \frac{1}{3}}{\alpha + \beta - \frac{2}{3}} \text{ for } \alpha, \beta \geq 1.$$

2. Sir William S. Gosset (pen name Student) developed a significance test and through it made significant contribution in the theory of sampling applicable in case of small samples. When population variance is not known, the test is commonly known as Student's t-test and is based on the t distribution.
3. The methodology of ANOVA is based on the following assumptions.
   (*i*) Each sample of size n is drawn randomly and each sample is independent of the other samples.
   (*ii*) The populations are normally distributed.
   (*iii*) The populations from which the samples are drawn have equal variances. This means that:
   $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \text{.........} = \sigma_k^2$, for k populations.
4. The major characteristics of the F-distribution are as follows:
   (*i*) Unlike normal distribution, which is only one type of curve irrespective of the value of the mean and the standard deviation, the F distribution is a family of curves. A particular curve is determined by two parameters. These are the degrees of freedom in the numerator and the degrees of freedom in the denominator. The shape of the curve changes as the number of degrees of freedom changes.
   (*ii*) It is a continuous distribution and the value of F cannot be negative.
   (*iii*) The curve representing the F distribution is positively skewed.
   (*iv*) The values of F theoretically range from zero to infinity.

NOTES

9.4 SUMMARY

- In probability theory and statistics, the Beta distribution is a family of continuous probability distributions defined on the interval [0, 1] parameterized by two positive shape parameters, denoted by α and β that appear as exponents of the random variable and control the shape of the distribution.
- In Bayesian inference, the Beta distribution is the conjugate prior probability distribution for the Bernoulli, Binomial and Geometric distributions.
- The usual formulation of the Beta distribution is also known as the Beta distribution of the first kind, whereas Beta distribution of the second kind is an alternative name for the Beta prime distribution.
- The inverse of the Harmonic Mean (H_X) of a distribution with random variable X is the arithmetic mean of $1/X$, or, equivalently, its expected value.
- When population variance is not known, the test is commonly known as Student's *t*-test and is based on the *t* distribution.
- There are two types of classifications involved in the analysis of variance. The one-way analysis of variance refers to the situations when only one fact or variable is considered.
- In the second type of classification, the response variable of interest may be affected by more than one factor.
- The null hypothesis, that we are going to test, is based upon the assumption that there is no significant difference among the means of different populations.
- Each sample of size *n* is drawn randomly and each sample is independent of the other samples.
- The variance between samples may be due to the effect of different treatments, meaning that the population means may be affected by the factor under consideration, thus, making the population means actually different, and some variance may be due to the inter-sample variability.

9.5 KEY WORDS

- **The numerator:** Since the variance between samples, $\sigma^2_{\text{between}}$ comes from many samples and if there are *k* number of samples, then the degrees of freedom, associated with the numerator would be $(k-1)$.
- **The denominator:** It is the mean variance of the variances of *k* samples and since, each variance in each sample is associated with the size of the sample (n), then the degrees of freedom associated with each sample would be $(n-1)$.
- **The *F*-Distribution:** It is a continuous distribution and the value of *F* cannot be negative.

9.6 SELF-ASSESSMENT QUESTIONS AND EXERCISES

NOTES

Short-Answers Questions

1. What is beta distribution?
2. Give the two conditions to use t test.
3. What is the analysis of variance?
4. Define the term degree of freedom.

Long-Answer Questions

1. Briefly explain about the beta distribution.
2. Describe the formulae which are used to calculate the value of t.
3. Explain the F distribution and give their assumptions.
4. Discuss about the computation of F.

9.7 FURTHER READINGS

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

NOTES

UNIT 10 DISTRIBUTIONS OF ORDER STATISTICS

Structure

10.0 INTRODUCTION

In probability theory and statistics, the moment generating function of a real-valued random variable is an alternative specification of its probability distribution. Thus, it provides the basis of an alternative route to analytical results compared with working directly with probability density functions or cumulative distribution functions. There are particularly simple results for the moment-generating functions of distributions defined by the weighted sums of random variables. However, not all random variables have moment-generating functions. As its name implies, the moment generating function can be used to compute a distribution's moments: the *n*th moment about 0 is the *n*th derivative of the moment-generating function, evaluated at 0.

In this unit, you will study about the distributions of order statistics and the moment generating function techniques.

10.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the distributions of order statistics
- Explain the moment generating function techniques

10.2 DISTRIBUTIONS OF ORDER STATISTICS

In statistics, the *k*th order statistic of a statistical sample is equal to its *k*th-smallest value. Together with rank statistics, order statistics are among the most fundamental tools in non-parametric statistics and inference. Important special cases of the order statistics are the minimum and maximum value of a sample, and (with some qualifications discussed below) the sample median and other sample quintiles. When using probability theory to analyse order statistics of random samples from a continuous distribution, the cumulative distribution function is used to reduce the analysis to the case of order statistics of the uniform distribution.

10.2.1 The Moment Generating Function

A function that generates moments of a random variable is known as the Moment Generating Function (MGF) of the random variable. A moment generating function may or may not exist but if it exists it is unique.

For the random variable X, its moment generating function $M_X(t)$ or $\psi(t)$ is defined as follows:

$\psi(t)$ or $M_X(t) = E(e^{tx})$

$$= \begin{cases} \sum e^{tx_i} \; f_i \text{ when } X \text{ is discrete and } f_i = P(X = x_i) \\ \int_{-\infty}^{\infty} e^{tx} \; f(x)\,dx \text{ when } X \text{ is continuous and } f(x) \text{ is } p.d.f. \text{ of } X. \end{cases}$$

Note that $E(e^{tX})$ is a function of t. The rth raw moment of X is the coefficient of $\frac{t^r}{r!}$ in the power series expansion of $M_X(t)$. That is, if

$$M_X(t) = a_0 + a_1 t + a_2 t^2 + \ldots \text{ to } \infty,$$

Then, $$(r!)a_2 = \mu'_r.$$

Hence, $$\mu'_r = \left[\frac{d^r}{dt^r} M(t)\right]_{t=0} = M_X^{(r)}(0)$$

Notes: (*i*) We have $M_X(t) = E(e^{tX})$

Differentiating $M_X(t)$ with respect to t, k times, we get

$$M_X^k(t) = E(X^k e^{tX})$$

$\therefore \; M_X^k(0) = E(X^k) = \mu'_k$ or α_k, provided the moment exists. Thus, α_k is obtained by differentiating $M_X(t)$ k times and putting $t = 0$.

(*ii*) We assume that $M_X(t)$ or $\psi(t)$ can be expanded as a power series in t by Maclaurin's theorem. Then we have,

$$\psi(t) = \psi(0) + t\,\psi'(0) + \frac{t^2}{\lfloor 2}\psi''(0) + \ldots + \frac{t^k}{\lfloor k}\psi^k(0) + \ldots$$

$$= \sum_{k=0}^{\infty} \frac{t^k}{\lfloor k}\psi^k(0) \text{ where } \psi^0(0) = \psi(0) = E(e^0) = E(1) = 1$$

$$= \sum_{k=0}^{\infty} \frac{t^k}{\lfloor k}\alpha_k$$

NOTES

Thus, it is proved that the kth order moment about origin α_k if it exists, is equal to the coefficient of $\frac{t^k}{\underline{|k}}$ in the expansion of $\psi(t)$ as a power series in t for any positive integer k. This is why $\psi(t)$ or $M_X(t)$ is called moment generating function.

(*iii*) If a and b are constants, then

(a) $M_{X+a}(t) = E(e^{(X+a)t}) = e^{at}\, E(e^{Xt}) = e^{at}\, M_X(t)$

(b) $M_{aX}(t) = E(e^{aXt}) = M_X(at)$

(c) $M_{\frac{X+a}{b}}(t) = E\left(e^{\left(\frac{X+a}{b}\right)t}\right) = e^{\frac{at}{b}}\, E\left(e^{\frac{Xt}{b}}\right) = e^{\frac{at}{b}}\, M_X\left(\frac{t}{b}\right)$

Example 1: Find the moment generating function of the random variable X whose probability function is,

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } x = -1 \\ \frac{1}{2} & \text{if } x = 1 \\ 0 & \text{elsewhere} \end{cases}$$

Hence, obtain the first three raw and central moments of X.

Solution: We have,

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

Now, for the raw moments, we expand $M_X(t)$ in an infinite power series as follows:

$$M_X(t) = 1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \ldots \text{ to } \infty$$

Hence, $\mu_1' = 0,\ \mu_2' = 1,\ \mu_3' = 0,\ \mu_4' = 1.$

[Note $M_X'(0) = 0,\ M_X''(0) = 1,\ M^{iii}{}_X(0) = 0,\ M^{iv}{}_X(0) = 1$]

$\therefore$ $\mu_1 = 0,\ \mu_2 = \mu_2' - \mu_1'^2 = 1 - 0^2 = 1,\ \mu_3 = 0,\ \mu_4 = 6$

Example 2: Find the moment generating function of the random variable whose probability function is,

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the central moments also.

NOTES

Solution: We see,

$$M_X(t) = E(e^{tx})$$

$$= \frac{1}{2}\int_0^2 e^{tx}dx = \frac{1}{2}\left[\frac{e^{tx}}{t}\right]_0^2 = \frac{1}{2t}\left(e^{2t}-1\right)$$

Writing in a power series, we get $M_X(t) = 1 + \frac{2t}{2!} + \frac{4t^2}{3!} + \frac{8t^3}{4!} + \ldots \text{ to } \infty$

Or, $$M_X(t) = 1+\frac{t}{1!}+\frac{4}{3}\frac{t^2}{2!}+\frac{2t^3}{3!}+\ldots \text{to } \infty$$

Now, $$\mu'_1 = 1$$

$$\mu'_2 = \frac{4}{3}$$

$$\mu'_3 = 2$$

$$\mu'_4 = \frac{16}{5}$$

$\therefore$ $$\mu_1 = 0,$$

$$\mu_2 = \frac{4}{3} - 1^2 = \frac{1}{3},$$

$$\mu_3 = 2-3\left(\frac{4}{3}\right)(1) + 2\,(1)^3 = 0$$

$$\mu_4 = \frac{16}{5} - 4\,(2)\,(1) + 6\left(\frac{4}{3}\right)(1)^2 - 3(1)^4 = \frac{1}{5}$$

Definition: The characteristic function $\phi_X(t)$ of a random variable X is defined as:

$$\phi_X(t) = E(e^{itX})$$

Note that the characteristic function of a random variable always exists.

A characteristic function of a random variable also gives the moments of the random variable In fact, the coefficient of $\frac{(it)^r}{r!}$, is the rth raw moment.

Example 3: Obtain the characteristic function of the Laplace variate whose probability function is:

$$f\,(x) = \frac{1}{2\lambda}\,e^{-\frac{|x-\mu|}{\lambda}}, -\infty < x < \infty \;\; \lambda > 0.$$

Solution: We have

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx}\,\frac{1}{2\lambda}\,e^{-\frac{|x-\mu|}{x}}\,dx$$

$$= \frac{1}{2\lambda}\int_{-\infty}^{\mu} e^{itx+\frac{(x-\mu)}{\lambda}}\,dx + \frac{1}{2\lambda}\int_{\mu}^{\infty} e^{itx-\frac{x-\mu}{\lambda}}\,dx$$

NOTES

$$= \frac{1}{2\lambda} e^{-\frac{\mu}{\lambda}} \int_{-\infty}^{\mu} e^{\left(it - \frac{1}{\lambda}\right)x} dx + \frac{1}{2\lambda} e^{+\frac{\mu}{\lambda}} \int_{\mu}^{\infty} e^{\left(it + \frac{1}{\lambda}\right)x} dx$$

$$= \frac{1}{2\lambda} e^{-\frac{\mu}{\lambda}} \left[\frac{e^{\left(it - \frac{1}{\lambda}\right)x}}{it - \frac{1}{\lambda}} \right]_{-\infty}^{\mu} + \frac{1}{2\lambda} e^{+\frac{\mu}{\lambda}} \left[\frac{e^{\left(it + \frac{1}{\lambda}\right)x}}{it + \frac{1}{\lambda}} \right]_{\mu}^{\infty}$$

$$= \frac{1}{2\lambda} e^{-\frac{\mu}{\lambda}} \frac{e^{\left(it - \frac{1}{\lambda}\right)\mu}}{it - \frac{1}{\lambda}} - \frac{1}{2\lambda} e^{+\frac{\mu}{\lambda}} \frac{e^{\left(it + \frac{1}{\lambda}\right)\mu}}{it + \frac{1}{\lambda}}$$

$$= \frac{1}{2}\left(\frac{e^{it\mu}}{1 + it\lambda} - \frac{e^{it\mu}}{it\lambda - 1} \right)$$

$$= \frac{e^{it\mu}}{1 + t^2\lambda^2}$$

Probability Differential: Let X be a continuous random variable. If $\delta x > 0$, then

$$P(x < X \le x + \delta x) = F(x + \delta x) - F(x).$$

$$= \delta x\, F'(x + \theta\, \delta x) \text{ where } 0 < \theta < 1$$

[by Lagrange mean value theorem of differential calculus]

So, $$\frac{p(x < X \le x + \delta x)}{\delta x} = F'(x + \theta\, \delta x)$$

$\therefore \lim_{\delta x \to 0} \frac{p(x < X \le x + \delta x)}{\delta x} = \lim_{\delta x \to 0} F'(x + \theta\, \delta x) = F'(x)$ if x be a point of continuity of $F'(x) = f(x)$ where $f(x)$ is the probability density function of X.

Thus, we get

$$f(x) = \lim_{\delta x \to 0} \frac{p(x < X \le x + \delta x)}{\delta x}$$

$$= \lim_{dx \to 0} \frac{p(x < X \le x + dx)}{dx}$$

[$\because \delta x$ = differential of $x = dx$ for the independent variable x]

Henceforth, we shall write $f(x)\, dx$ for $P(x < X \le x + dx)$ which will actually mean $\lim_{dx \to 0} \frac{p(x < X \le x + dx)}{dx} = f(x)$.

The expression $P(x < X \le x + dx)$ will always be used in the above limiting sense and so there will be no ambiguity throughout our discussion. The expression $P(x < X \le x + dx)$ which is taken to be equal to $f(x)\, dx$ is called the *probability differential* for the continuous random variable X.

NOTES

Calculation of Mean and Variance from Moment Generating Functions

1. **Binomial Distribution:** Let X be a binomial (n, p) variate, then the M.G.F. of X is given by:

$$\psi(t) = (q + pe^t)^n \text{ where } q = 1 - p$$

$$\therefore \quad \psi'(t) = n(q + pe^t)^{n-1}\, pe^t$$

$$\psi''(t) = n(n-1)\,(q + pe^t)^{n-2}\,(pe^t)^2 + np\,(q + pe^t)^{n-1}\, e^t$$

$$\therefore \quad \psi'(0) = n(q + p)^{n-1} p = np = E\,(X) = m[\because\ (p + q) = 1]$$

And
$$\alpha_2 = \psi''(0) = n(n-1)\,(q + p)^{n-2}\, p^2 + np$$

$$= n(n-1)\, p^2 + np$$

$$\therefore \quad \text{var}(X) = \alpha_2 - m^2 = n(n-1)p^2 + np - n^2p^2$$

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

$$= npq$$

2. **Poisson Distribution:** Let X be a Poisson λ variate, then the M.G.F. of X is given by:

$$\psi(t) = e^{\lambda(e^t - 1)}$$

$$\therefore \quad \psi'(t) = e^{\lambda(e^t - 1)}\,.\,\lambda e^t$$

$$\psi''(t) = (\lambda e^t)^2\, e^{\lambda(e^t - 1)} + \lambda e^t\, e^{\lambda(e^t - 1)}$$

$$\therefore \quad \psi'(0) = E(X) = m = \lambda$$

And
$$\alpha_2 = \psi''(0) = \lambda^2 + \lambda$$

$$\therefore \quad \text{var}(X) = \alpha_2 - m^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

3. **Uniform Distribution:** Let X be a uniform random variable in $[a, b]$; then the M.G.F. of X is given by:

$$\psi(t) = \frac{1}{b-a}\left[\frac{e^{bt} - e^{at}}{t}\right]$$

$$= \frac{1}{t(b-a)}\left[\left(1 + bt + \frac{b^2t^2}{\lfloor 2} + \frac{b^3t^3}{\lfloor 3} + \ldots\right) - \left(1 + at + \frac{a^2t^2}{\lfloor 2} + \frac{a^2t^3}{\lfloor 3} + \ldots\right)\right]$$

$$= \frac{1}{t(b-a)}\left[(b-a)t + (b^2 - a^2)\frac{t^2}{2} + (b^3 - a^3)\frac{t^3}{6} + (b^4 - a^4)\frac{t^4}{24} + \ldots\right]$$

$$= 1 + (b + a)\frac{t}{2} + \frac{b^2 + a^2 + ab}{6}t^2 + \frac{(b^2 + a^2)\,(b + a)}{24}t^3 + \ldots$$

$$\therefore \quad \psi'(t) = \frac{b + a}{2} + \frac{(b^2 + a^2 + ab)}{3}t + \frac{(b^2 + a^2)\,(b + a)}{8}t^2 + \ldots.$$

$$\therefore \quad \psi''(t) = \frac{b^2 + a^2 + ab}{3} + \frac{(b^2 + a^2)\,(b + a)}{4}t + \ldots.$$

$$\therefore \quad \psi'(0) = E\,(X) = m = \frac{b + a}{2}$$

NOTES

And $\alpha_2 = \psi''(0) = \dfrac{b^2 + a^2 + ab}{3}$

$\therefore$ $\text{var}(X) = \alpha_2 - m^2 = \dfrac{b^2 + a^2 + ab}{3} - \left(\dfrac{b+a}{2}\right)^2$

$$= \frac{4b^2 + 4a^2 + 4ab - 3b^2 - 3a^2 - 6ab}{12}$$

$$= \frac{b^2 + a^2 - 2ab}{12} = \frac{(b-a)^2}{12}$$

4. **Normal Distribution:** Let X be a normal (m, σ) variate, then the M.G.F. of X is given by:

$$\psi(t) = e^{tm + \frac{1}{2}t^2\sigma^2}$$

$\therefore$ $$\psi'(t) = e^{tm + \frac{1}{2}t^2\sigma^2}(m + 2t\sigma^2)$$

$$\psi''(t) = (m + t\sigma^2)^2 e^{tm + \frac{1}{2}t^2\sigma^2} + \sigma^2 e^{tm + \frac{1}{2}t^2\sigma^2}$$

$\therefore$ $\psi'(0) = E(X) = \text{Mean} = m$

And $\psi''(0) = \alpha_2 = m^2 + \sigma^2$

$\therefore$ $\text{var}(X) = \alpha_2 - m^2 = m^2 + \sigma^2 - m^2 = \sigma^2$

5. **Exponential Distribution:** Let X be a random variable of the exponential distribution, then the M.G.F. of X is given by:

$$\psi(t) = \frac{\lambda}{\lambda - t} = \frac{1}{1 - \frac{t}{\lambda}} = \left(1 - \frac{t}{\lambda}\right)^{-1}$$

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

$\therefore$ $$\psi'(t) = \frac{1}{\lambda} + \frac{2t}{\lambda^2} + \frac{3t^2}{\lambda^3} + \ldots$$

$$\psi''(t) = \frac{2}{\lambda^2} + \frac{6t}{\lambda^3} + \ldots$$

$\therefore$ $\psi'(0) = \dfrac{1}{\lambda} = m = \text{Mean}$

And $\psi''(0) = \alpha_2 = \dfrac{2}{\lambda^2}$

$\therefore$ $\text{var}(X) = \alpha_2 - m^2 = \dfrac{2}{\lambda^2} - \dfrac{1}{\lambda^2} = \dfrac{1}{\lambda^2}$

Example 4: The probability mass function of a random variable X is $f_i = P(X = i) = 2^{-i}$ where $i = 1, 2, 3, \ldots$ Find the M.G.F. of X and hence find the mean and variance of X.

NOTES

Solution: Now $\psi(t)$ = M.G.F. of $X = E(e^{tx}) = \sum_{i=1}^{\infty} e^{ti} f_i$

$$= \sum_{i=1}^{\infty} e^{ti} 2^{-i} = \sum_{i=1}^{\infty}\left(\frac{e^t}{2}\right)^i = \frac{e^t}{2}\sum_{i=1}^{\infty}\left(\frac{e^t}{2}\right)^{i-1}$$

$$= \frac{e^t}{2}\sum_{k=0}^{\infty}\left(\frac{e^t}{2}\right)^k \qquad \text{where } i-1=k$$

$$= \frac{e^t}{2}\,[1+r+r^2+r^3+...] \text{ where } r=\frac{e^t}{2}$$

$$= \frac{e^t}{2}\left[\frac{1}{1-r}\right] = \frac{e^t}{2}\left[\frac{1}{1-e^t/2}\right] = \frac{e^t}{2-e^t}$$

$$\therefore \psi'(t) = \frac{e^t(2-e^t)-e^t(-e^t)}{(2-e^t)^2} = \frac{e^t[2-e^t+e^t]}{(2-e^t)^2} = \frac{2e^t}{(2-e^t)^2}$$

$$\psi''(t) = \frac{(2e^t)^2\,2e^t - 2e^t\,2(2-e^t)(-e^t)}{(2-e^t)^4} = \frac{2[e^t\,(2-e^t)+2e^te^t]}{(2-e^t)^3}$$

$$= \frac{2[e^t(2+e^t)]}{(2-e^t)^3}$$

$$\therefore \qquad \psi'(0) = m = \text{Mean } = E(X) = 2$$

And $\qquad \psi''(0) = \alpha_2 = \dfrac{2[1(2+1)]}{(2-1)^3} = 6$

$$\therefore \qquad \text{Var}(X) = \alpha_2 - m^2 = 6-4 = 2$$

Example 5: Find the M.G.F. of the following continuous distribution with probability density function:

$$f(x) = \begin{cases} \frac{1}{2}x^2e^{-x} & \text{for } x>0 \\ 0 & \text{otherwise}\end{cases}$$

and hence find the mean and variance.

Solution: The M.G.F. of this distribution is given by:

$$\psi(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \int_0^{\infty} e^{tx}\,\frac{1}{2}x^2e^{-x}dx$$

$$= \int_0^{\infty} e^{-x(1-t)}\,\frac{x^2}{2}\,dx$$

$$= \int_0^{\infty} e^{-z}\,\frac{z^2}{2(1-t)^2}\,\frac{dz}{(1-t)} \quad \text{where } (1-t)\,x = z,\ dx = \frac{dz}{1-t}$$

$$= \frac{1}{2(1-t)^3}\int_0^{\infty} e^{-z}z^2\,dz$$

$$= \frac{1}{2(1-t)^3}\left[\frac{z^2 e^{-z}}{-1} - 2ze^{-z} - 2e^{-z}\right]_0^\infty$$

$$= \frac{2}{2(1-t)^3} = \frac{1}{(1-t)^3}$$

NOTES

$$\therefore \quad \psi'(t) = \frac{3}{(1-t)^4},\ \psi''(t) = \frac{12}{(1-t)^5}$$

$$\therefore \quad \psi'(0) = 3 = E(X) = m = \text{Mean}$$

And $\quad \psi'(0) = \alpha_2 = 12$

$$\text{var}(X) = \alpha_2 - m^2 = 12 - 9 = 3$$

Check Your Progress

1. What is moment generating function?
2. What is the probability differential?
3. Calculate the variance from moment generating function of X, by Poisson distribution.

10.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A function that generates moments of a random variable is known as the Moment Generating Function (MGF) of the random variable. A moment generating function may or may not exist but if it exists it is unique.
2. The expression $P(x < X \leq x + dx)$ will always be used in the above limiting sense and so there will be no ambiguity throughout our discussion. The expression $P(x < X \leq x + dx)$ which is taken to be equal to $f(x)\, dx$ is called the probability differential for the continuous random variable X.
3. Poisson Distribution: Let X be a Poisson λ variate, then the M.G.F. of X is given by:

$$\psi(t) = e^{\lambda(e^t - 1)}$$

$$\therefore \quad \psi'(t) = e^{\lambda(e^t - 1)} \cdot \lambda e^t$$

$$\psi''(t) = (\lambda e^t)^2 e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)}$$

$$\therefore \quad \psi'(0) = E(X) = m = \lambda$$

And $\alpha_2 = \psi''(0) = \lambda^2 + \lambda$

$$\therefore \quad \text{var}(X) = \alpha_2 - m^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

NOTES

10.4 SUMMARY

- If a and b are constants, then
  (a) $M_{X+a}(t) = E(e^{(X+a)t}) = e^{at}\,E(e^{Xt}) = e^{at}\,M_X(t)$
  (b) $M_{aX}(t) = E(e^{aXt}) = M_X(at)$
  (c) $M_{\frac{X+a}{b}}(t) = E\left(e^{\left(\frac{X+a}{b}\right)t}\right) = e^{\frac{at}{b}}\,E\left(e^{\frac{Xt}{b}}\right) = e^{\frac{at}{b}}\,M_X\left(\frac{t}{b}\right)$
- The characteristic function $\phi_X(t)$ of a random variable X is defined as:
  $\phi_X(t) = E(e^{itX})$
- Let X be a continuous random variable. If $\delta x > 0$, then $P(x < X \leq x + \delta x) = F(x + \delta x) - F(x)$.

10.5 KEY WORDS

- **Moment Generating Function:** A function that generates moments of a random variable is known as the Moment Generating Function (MGF) of the random variable.
- **Probability differential:** The expression $P(x < X \leq x + dx)$ which is taken to be equal to $f(x)\,dx$ is called the *probability differential* for the continuous random variable X.

10.6 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define moment generating function for the random variable X.
2. What is probability differential?
3. Calculate the mean and variance form moment generating functions of binomial distribution.

Long-Answer Questions

1. Describe the distributions of order statistics.
2. Briefly explain about the moment generating function.
3. Give the calculation of mean and variance from moment generating functions.

10.7 FURTHER READINGS

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

NOTES

NOTES

UNIT 11 DISTRIBUTION OF X AND ns^2/σ^2

Structure

11.0 INTRODUCTION

In probability theory, the expected value of a random variable is a generalization of the weighted average and intuitively is the arithmetic mean of a large number of independent realizations of that variable. The expected value is also known as the expectation, mathematical expectation, mean, average, or first moment.

In probability theory and statistics, a probability distribution is the mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment. More specifically, the probability distribution is a mathematical description of a random phenomenon in terms of the probabilities of events. For instance, if the random variable X is used to denote the outcome of a coin toss experiment, then the probability distribution of X would take the value 0.5 for X = Heads, and 0.5 for X = Tails (assuming the coin is fair).

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

A probability distribution is a table or an equation that links each outcome of a statistical experiment with its probability of occurrence.

In this unit, you will study about the distributions of x and ns^2/σ^2 and expectations of functions of random variables.

11.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the distributions of x and ns^2/σ^2
- Analyse the expectations of functions of random variables

11.2 THE DISTRIBUTIONS OF X AND ns^2/σ^2

In probability theory and statistics, a probability distribution is the mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment. More specifically, the probability distribution is a mathematical description of a random phenomenon in terms of the probabilities of events. For instance, if the random variable X is used to denote the outcome of a coin toss experiment, then the probability distribution of X would take the value 0.5 for $X =$ Heads, and 0.5 for $X =$ Tails (assuming the coin is fair).

An example will make clear the relationship between random variables and probability distributions. Suppose you flip a coin two times. This simple statistical experiment can have four possible outcomes: HH, HT, TH, and TT. Now, let the variable X represent the number of Heads that result from this experiment. The variable X can take on the values 0, 1, or 2. In this example, X is a random variable; because its value is determined by the outcome of a statistical experiment. Therefore, a probability distribution is a table or an equation that links each outcome of a statistical experiment with its probability of occurrence.

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

In the test of independence, the row and column variables are independent of each other and this is the null hypothesis. The following are properties of the test for independence:

- The data are the observed frequencies.
- The data is arranged into a contingency table.
- The degrees of freedom are the degrees of freedom for the row variable times the degrees of freedom for the column variable. It is not one less than the sample size, it is the product of the two degrees of freedom.
- It is always a right tail test.
- It has a Chi-square distribution.

NOTES

- The expected value is computed by taking the row total times the column total and dividing by the grand total.
- The value of the test statistic does not change if the order of the rows or columns are switched.
- The value of the test statistic does not change if the rows and columns are interchanged (transpose of the matrix).

Contingency Tables

Suppose the frequencies in the data are classified according to attribute A into r classes (rows) and according to attribute B into c classes (columns) as follows:

Class	B_1	B_2	…	B_c	Total
A_1	O_{11}	O_{12}	…	O_{1c}	(A_1)
A_2	O_{21}	O_{22}	…	O_{2c}	(A_2)
…	…	…	…	…	…
A_r	O_{r1}	O_{r2}	…	O_{rc}	(A_r)
Table	(B_1)	(B_2)	…	(B_c)	N

The totals of row and column frequencies are (A_i), (B_j).

To test if there is any relation between A, B we set up the null hypothesis of independence between A, B.

The expected frequency in any cell is calculated by using the formula:

$$E_{ij} = \frac{(A_i)(B_j)}{N}$$

Use $\chi^2 = \frac{-(O_{ij} - E_{ij})^2}{E_{ij}}$ with degrees of freedom $= (r-1)(c-1)$

For example,

Observed Frequencies

	School	College	University	Total
Boys	10	15	25	50
Girls	25	10	15	50
Total	35	25	40	100

NOTES

$$\frac{35\times 50}{100}=17.5$$

$$\frac{25\times 50}{100}=12.5$$

$$\frac{40\times 50}{100}=20$$

Expected Frequencies

	School	**College**	**University**	**Total**
Boys	17.5	12.5	20	50
Girls	17.5	12.5	20	50
Total	35	25	40	100

Degrees of freedom = (2 – 1) (3 – 1) = 2

$$\chi^2=\sum(O-E)^2/E=9.9$$

This is greater than the table value. It is not true that education does not depend on sex, i.e., the two are not independent.

Concept of Test Statistics

In the test for given population variance, the variance is the square of the standard deviation, whatever you say about a variance can be, for all practical purposes, extended to a population standard deviation.

To test the hypothesis that a sample $x_1, x_2, ...x_n$ of size n has a specified variance $\sigma^2=\sigma_2^2$

$$H_0 : \sigma^2 = \sigma^2{}_0$$

Or,

Null hypothesis $\sigma = \sigma_0$

$$H_1 : \sigma^2 > \sigma_0^2$$

Test statistics $\chi^2=\dfrac{ns^2}{\sigma_0^2}=\dfrac{\sum(x-\overline{x})^2}{\sigma_0^2}$

If χ^2 is greater than the table value we reject the null hypothesis.

11.3 EXPECTATIONS OF FUNCTIONS OF RANDOM VARIABLES

If p happens to be the probability of the happening of an event in a single trial, then the expected number of occurrence of that event in n trials is given by $n, p,$ where n means the number of trials and p means the probability of happening of an event.

NOTES

Thus, the expectation may be regarded as the likely number of success in n trials. If probability p is determined as the Relative Frequency in n trials then the mathematical expectation in these n trials would be equal to the actual (observed) number of successes in these n trials. Mathematical expectation does in no way mean that the concerning event must happen the number of times given by the mathematical expectation; it simply gives the likely number of the happening of the event in n trials. Mathematical expectation can be explained with the help of following examples.

Example 1: In 12000 trials of a draw of two fair dice, what is the expected number of items that the sum will be less than 4?

Solution: With two fair dice, the total number of equally likely cases = 6 × 6 = 36

Number of cases favourable to the event in a single thrown of two dice = 3 viz., (1 + 1, 1 + 2, 2 + 1)

$\therefore$ The required $p = \dfrac{3}{36} = \dfrac{1}{12}$

Hence, the expected number of times the total will be less than 4 in 12000 trials.

$$= \frac{1}{12} \times 12000 = 1000$$

The concept of expectation is of great use in the analysis of all games of chance wherein an effort is made to evaluate the expectations of the players. If p represents the probability of a player in any game and M the sum of money which he will receive in case of success, the sum of money denoted by ($p.M$) is called his expectation. Thus, the expectation is calculated by finding the probability of success (by any of the methods stated so far) and then multiplying it by the money value which the player expects in case of success. The significance of this expectation lies in the fact that if a player pays more than this, by way of fair price, per game then he is sure to lose but if he plays long enough and if he pays less than his expectation per game he is certain to win in the long run. It is on this principle that speculators and businessmen take decisions in real life situations.

Example 2: A and B throw with one die for a stake of Rs 44 which is to be won by the player who first throws 2. If A has the first throw, what are their respective expectation?

Solution: The chance of throwing 2 with one die = $\dfrac{1}{6}$, A can win in the first, third, fifth .... throws.

His chance of throwing 2 is,

$$\left.\frac{1}{6} + \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} + ...\right]$$

NOTES

$$\text{Or} \quad \frac{1}{6}\left[1+\left(\frac{5}{6}\right)^2+\ldots\right]$$

B can win the second, fourth, sixth ... throw

His chance of throwing 2 is,

$$\frac{5}{6}\times\frac{1}{6}+\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{1}{6}+\ldots$$

$$\text{Or} \quad \frac{5}{36}\left[1+\left(\frac{5}{6}\right)^2+\ldots\right]$$

$\therefore$ A's chance to B's chance stands as 6 : 5. Hence their respective chances are $\frac{6}{11}$ and $\frac{5}{11}$. As such their expectations are as under:

$$\text{A's expectation} = 44\times\frac{6}{11}=\frac{264}{11}=24 \text{ Rs}$$

$$\text{B's expectation} = 44\times\frac{5}{11}=\frac{220}{11}=20 \text{ Rs}$$

Example 3: A person has ten coins which he throws down in succession. He is to receive one rupee if the first falls head, two rupees if the second also falls head, four rupees if the third also falls head and so on. The amount doubling each time but as soon as a coin falls tail he ceases to receive any thing. What is the value of his expectation?

Solution:

$$\text{Chance of falling head} = \frac{1}{2}$$

$$\text{Chance of falling tail} = \frac{1}{2}$$

$$\text{Chance of falling head in 1st trial} = \frac{1}{2}$$

$$\text{Expectation in 1st trial} = \frac{1}{2}\times 1 = \frac{1}{2} \text{ Re}$$

Now if he succeeds in getting head in the first trial then only he is allowed to do his second trial.

NOTES

$\therefore$ His chance of success in 2nd trial $= \frac{1}{2} \times \frac{1}{2} = \left(\frac{1}{2}\right)^2$

His expectation in the 2nd trial $= \left(\frac{1}{2}\right)^2 \times 2^1 = \frac{1}{2}$ Re

Similarly his expectation in third trial $= \left(\frac{1}{2}\right)^3 \times 2^2$

and so on upto the 10th trial.

His expectation in 10th trail $= \left(\frac{1}{2}\right)^{10} \times 2^9$

$\therefore$ Total expectation (or Expected Monetary Value or EMV)

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

$= 5$ Rs

Example 4: In a given business venture a man can make a profit of Rs 2000 with probability 0.8 or suffer a loss of Rs 800 with probability 0.2. Determine his expectation.

Solution: With probability 0.8, the expectation of profit is,

$$= \frac{8}{10} \times 2000 = 1600 \text{ Rs}$$

With probability 0.2, the expectation of loss is,

$$= \frac{2}{10} \times 800 = 160 \text{ Rs}$$

His overall net expectation in the venture concerned would then clearly be,

Rs 1600 – 160 = 1440

Thus in the above two examples the concept of mathematical expectation has been extended to discrete random variable which can assume the values X_1, X_2, ... X_k with respective probabilities $p_1, p_2 ...p_k$ where $p_1 + p_2 + ... p_k = 1$. The mathematical expectation of X denoted by $E(X)$ is defined as,

$E(X) = p_1X_1 + p_2X_2 + \ ... \ p_kX_k$

Check Your Progress

1. Give the two properties of the test for independence.
2. Explain the concept of test statistics.
3. What is expected number?

11.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

NOTES

1. The following are properties of the test for independence
   (a) The data are the observed frequencies.
   (b) The data is arranged into a contingency table.
2. In the test for given population variance, the variance is the square of the standard deviation, whatever you say about a variance can be, for all practical purposes, extended to a population standard deviation.
3. If p happens to be the probability of the happening of an event in a single trial, then the expected number of occurrence of that event in n trials is given by $n, p,$ where n means the number of trials and p means the probability of happening of an event.

11.5 SUMMARY

- The degrees of freedom are the degrees of freedom for the row variable times the degrees of freedom for the column variable. It is not one less than the sample size, it is the product of the two degrees of freedom.
- The expected value is computed by taking the row total times the column total and dividing by the grand total.
- The value of the test statistic doesn't change if the order of the rows or columns are switched.
- The value of the test statistic doesn't change if the rows and columns are interchanged (transpose of the matrix).
- If probability p is determined as the Relative Frequency in n trials then the mathematical expectation in these n trials would be equal to the actual (observed) number of successes in these n trials.

11.6 KEY WORDS

- **Probability distribution: I**n probability theory and statistics, a probability distribution is the mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment.
- **Null hypothesis:** $\sigma = \sigma_0$

  $H_1 : \sigma^2 > \sigma_0^2$
- **Test statistics:** $\chi^2 = \dfrac{ns^2}{\sigma_0^2} = \dfrac{\sum (x - \overline{x})^2}{\sigma_0^2}$

11.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Give the formula to calculate the expected frequency in any cell.
2. In 500 trials of a draw of two fair dice, what is the expected number of items that the sum will be less than 3?

Long-Answer Questions

1. Describe the distribution of X and ns^2/s^2.
2. Briefly explain the expectations of functions of random variables.

11.8 FURTHER READINGS

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

NOTES

BLOCK - IV
LIMITING DISTRIBUTIONS

UNIT 12 LIMITING DISTRIBUTIONS AND CONVERGENCE

Structure

12.0 INTRODUCTION

A 'Limiting Distribution' is also termed as an 'Asymptotic Distribution'. Fundamentally, it is defined as the hypothetical distribution or convergence of a sequence of distributions. Since it is hypothetical, it is not considered as a distribution as per the general logic. The asymptotic distribution theory is typically used to find a limiting distribution to a series of distributions. The mode of convergence for a sequence of random variables are defined on the basis of the convergence in probability and in distribution. The concept of convergence leads us to the two fundamental results of probability theory, the Law of Large Number and Central Limit Theorem (CLT). Limiting probability distributions are significantly used to find the appropriate sample sizes. When a sample size is large enough, then a statistic's distribution will form a limiting distribution, assuming that such a distribution exists. In probability theory, there exist several different notions of convergence of random variables. The convergence of sequences of random variables to some limit random variable is an important concept in probability theory, and its applications to statistics and stochastic processes. The same concepts are known in more general mathematics as 'Stochastic Convergence'.

In this unit, you will study about the limiting distributions, convergence in distribution and convergence in probability.

12.1 OBJECTIVES

NOTES

After going through this unit, you will be able to:

- Understand the basic concept of limiting distributions
- Analyse what convergence in distribution is
- Explain about the convergence in probability

12.2 LIMITING DISTRIBUTIONS

A 'Limiting Distribution' is also termed as an 'Asymptotic Distribution'. Fundamentally, it is defined as the hypothetical distribution or convergence of a sequence of distributions. Since it is hypothetical, it is not considered as a distribution as per the general logic. The asymptotic distribution theory is typically used to find a limiting distribution to a series of distributions.

Some of the limiting or asymptotic distributions are well-recognized and defined by different statisticians. For example, the sampling distribution of the *t*-statistic will converge to a standard normal distribution if the sample size is large enough.

In basic statistics, the process includes a random sample of observations and then fitting that data to a known distribution similar to the normal distribution or *t* distribution. Fitting the statistical data accurately to a known distribution is generally very challenging task because of the limited sample sizes. The accurate approximation is based on the estimation of 'presumptions and guesses' established on the nature of large sample statistics. The limiting/asymptotic distribution can be applied on small, finite samples for approximating the true distribution of a random variable.

Limiting probability distributions are significantly used to find the appropriate sample sizes. When a sample size is large enough, then a statistic's distribution will form a limiting distribution, assuming that such a distribution exists.

The Central Limit Theorem (CLT) uses the limit concept for describing the behaviour of sample means. The CLT states that the sampling distribution of the sampling means approaches a normal distribution as the sample size increases — irrespective of the shape of the population distribution. For example, if you consider large samples then the graph of the sample means will be similar to a normal distribution, even if the graph is skewed or otherwise non-normal. Alternatively, the 'Limiting Distribution' for a large set of sample means is the 'Normal Distribution'.

Definition: Suppose X_n is a random sequence with Cumulative Distribution Function (CDF) $F_n(X_n)$ and X is a random variable with CDF $F(x)$. If F_n converges to F as $n > \infty$ (for all points where $F(x)$ is continuous), then the distribution of x_n converges to x. This distribution is called the limiting distribution of x_n.

In simpler terms, it can be stated that the limiting probability distribution of X_n is the limiting distribution of some function of X_n.

Convergence of Random Variables

In probability theory, there exist several different notions of convergence of random variables. The convergence of sequences of random variables to some limit random variable is an important concept in probability theory, and its applications to statistics and stochastic processes. The same concepts are known in more general mathematics as 'Stochastic Convergence' and they formalize the idea that a sequence of essentially random or unpredictable events can sometimes be expected to settle down into a behaviour that is essentially unchanging when items far enough into the sequence are studied. The different possible notions of convergence relate to how such a behaviour can be characterized: two readily understood behaviours are that the sequence eventually takes a constant value, and that values in the sequence continue to change but can be described by an unchanging probability distribution.

'Stochastic Convergence' validates the notion that a sequence of essentially random or unpredictable events can sometimes be expected to settle into a pattern. The pattern may be,

- Convergence in the classical sense to a fixed value, perhaps itself coming from a random event.
- An increasing similarity of outcomes to what a purely deterministic function would produce.
- An increasing preference towards a certain outcome.
- An increasing 'aversion' against straying far away from a certain outcome.
- That the probability distribution describing the next outcome may grow increasingly similar to a certain distribution.

Some more theoretical patterns state that the,

- Series formed by calculating the expected value of the outcome's distance from a particular value may converge to 0.
- Variance of the random variable describing the next event grows smaller and smaller.

The above mentioned facts define the convergence of a single series to a limiting value, the notion of the convergence of two series towards each other is also significant and the sequence is defined as either the difference or the ratio of the two series.

If the average of n independent random variables Y_i, for $i = 1, ..., n$, all having the same finite mean and variance, is given by,

$$X_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Then as n tends to infinity, X_n converges in probability to the common mean, μ, of the random variables Y_i. This result is known as the 'Weak Law of Large Numbers'. Other forms of convergence are important in other useful theorems, including the Central Limit Theorem (CLT).

Here we assume that (X_n) is a sequence of random variables, where X is a random variable, and these are defined on the same probability space $(\Omega, \mathcal{F}, \Pr)$.

12.2.1 Convergence in Distribution

The convergence in distribution mode provide the required expectation to recognise and observe the next outcome in a sequence of random experiments which is well modelled by means of a given probability distribution.

Convergence in distribution is typically considered as the weakest form of convergence, since it is implied by all other types of convergence. However, convergence in distribution is very frequently used in practice; most often it arises from application of the Central Limit Theorem (CLT).

Definition: A sequence $X_1, X_2, \ldots$ of real-valued random variables is said to converge in distribution, or converge weakly, or converge in law to a random variable X if,

$$\lim_{n\to\infty} F_n(x) = F(x).$$

For every number $x \in \mathbb{R}$ at which F is continuous. Here F_n and F are the Cumulative Distribution Functions (CDFs) of random variables X_n and X, respectively.

Essentially, only the continuity points of F should be considered. For example, if X_n are distributed uniformly on intervals $(0, 1/n)$, then this sequence converges in distribution to degenerate a random variable $X = 0$. Certainly, we can state that $F_n(x) = 0$ for all n when $x \leq 0$, and $F_n(x) = 1$ for all $x \geq 1/n$ when $n > 0$. However, for this limiting random variable $F(0) = 1$, even though $F_n(0) = 0$ for all n.

Thus the convergence of CDFs fails at the point $x = 0$ where F is discontinuous.

Convergence in distribution may be denoted as,

$$X_n \xrightarrow{d} X,\ \ X_n \xrightarrow{\mathcal{D}} X,\ \ X_n \xrightarrow{\mathcal{L}} X,\ \ X_n \xrightarrow{d} \mathcal{L}_X,$$
$$X_n \rightsquigarrow X,\ \ X_n \Rightarrow X,\ \ \mathcal{L}(X_n) \to \mathcal{L}(X),$$

Where $\mathcal{L}_X$ is the probability distribution law of X. For example, if X is standard normal we can write $X_n \xrightarrow{d} \mathcal{N}(0,\ 1)$.

NOTES

For random vectors $\{X_1, X_2, ...\} \subset \mathbf{R}^k$ the convergence in distribution can be similarly defined. We say that this sequence converges in distribution to a random k-vector X if,

$$\lim_{n\to\infty} \Pr(X_n \in A) = \Pr(X \in A)$$

For every $A \subset \mathbf{R}^k$ which is a continuity set of X.

The definition of convergence in distribution may be extended from random vectors to more general random elements in arbitrary metric spaces, and even to the 'random variables' which are not measurable, for example a situation which occurs in the study of empirical processes. This is referred as the 'Weak Convergence of Laws without Laws Being Defined' except asymptotically.

In this case the term '**Weak Convergence**' is preferably used and we say that a sequence of random elements $\{X_n\}$ converges weakly to X (denoted as $X_n \Rightarrow X$) if,

$$E^* \, h(X_n) \to E \, h(X)$$

For all continuous bounded functions h. Here E* denotes the *outer expectation*, that is the expectation of a 'Smallest Measurable Function g that Dominates $h(X_n)$'.

Example of convergence in distribution can be explained by means of the newly built dice factory. Suppose a new dice factory has just been built. The first few dice come out quite biased, due to imperfections in the production process. The outcome from tossing any of them will follow a distribution markedly different from the desired uniform distribution.

As the production in the factory improves, the dice become less and less loaded, and the outcomes from tossing a newly produced die will follow the uniform distribution more and more closely.

12.2.2 Convergence in Probability

The basic notion behind this type of convergence is that the probability of an 'unusual' outcome becomes smaller and smaller as the sequence progresses.

The concept of convergence in probability is used very often in statistics. For example, an estimator is called consistent if it converges in probability to the quantity being estimated. Convergence in probability is also the type of convergence established by the 'Weak Law of Large Numbers'.

Definition: A sequence $\{X_n\}$ of random variables **Converges in Probability** towards the random variable X if for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr\left(|X_n - X| > \varepsilon\right) = 0.$$

More explicitly, let P_n be the probability that X_n is outside the ball of radius ε centred at X. Then X_n is said to converge in probability to X if for any $\varepsilon > 0$ and any $\delta > 0$ there exists a number N (which may depend on ε and δ) such that for all $n \geq N$, $P_n < \delta$ (the Definition of Limit).

Notice that for the condition to be satisfied, for each n the random variables X and X_n cannot be independent and thus convergence in probability is a condition on the joint CDF's, as opposed to convergence in distribution, which is a condition on the individual CDF's, unless X is deterministic like for the 'Weak Law of Large Numbers'. At the same time, the case of a deterministic X cannot, whenever the deterministic value is a discontinuity point (not isolated), be handled by convergence in distribution, where discontinuity points have to be explicitly excluded.

Convergence in probability is denoted by adding the letter 'p' over an arrow indicating convergence, or using the 'plim' probability limit operator:

$$X_n \xrightarrow{p} X, \quad X_n \xrightarrow{P} X, \quad \operatorname*{plim}_{n\to\infty} X_n = X.$$

For random elements $\{X_n\}$ on a separable metric space (S, d), convergence in probability is defined similarly by,

$$\forall \varepsilon > 0, \Pr\left(d(X_n, X) \geq \varepsilon\right) \to 0.$$

Properties

- 'Convergence in Probability' implies 'Convergence in Distribution'.
- In the opposite direction, convergence in distribution implies convergence in probability when the limiting random variable X is a constant.
- Convergence in probability does not imply almost sure convergence.
- The continuous mapping theorem states that for every continuous function $g(\bullet)$, if $X_n \xrightarrow{p} X$, then also $g(X_n) \xrightarrow{p} g(X)$.

Check Your Progress

1. Explain the basic concept of limiting distribution?
2. When the limiting/asymptotic distribution can be applied?
3. Why the limiting probability distributions are used in statistics?
4. Elucidate on the convergence in distribution in probability distribution.
5. Explain the concept of convergence in probability. Why it is often used in statistics?

12.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

NOTES

1. A 'Limiting Distribution' is also termed as an 'Asymptotic Distribution'. Fundamentally, it is defined as the hypothetical distribution or convergence of a sequence of distributions. The asymptotic distribution theory is typically used to find a limiting distribution to a series of distributions. Some of the limiting or asymptotic distributions are well-recognized and defined by different statisticians. For example, the sampling distribution of the *t*-statistic will converge to a standard normal distribution if the sample size is large enough.
2. The limiting/asymptotic distribution can be applied on small, finite samples for approximating the true distribution of a random variable.
3. Limiting probability distributions are significantly used to find the appropriate sample sizes. When a sample size is large enough, then a statistic's distribution will form a limiting distribution, assuming that such a distribution exists. In statistics, the process includes a random sample of observations and then fitting that data to a known distribution similar to the normal distribution or *t* distribution. Fitting the statistical data accurately to a known distribution is generally very challenging task because of the limited sample sizes.
4. The convergence in distribution mode provide the required expectation to recognise and observe the next outcome in a sequence of random experiments which is well modelled by means of a given probability distribution. Convergence in distribution is typically considered as the weakest form of convergence, since it is implied by all other types of convergence.
5. The basic notion behind the convergence in probability type is that the probability of an 'unusual' outcome becomes smaller and smaller as the sequence progresses.

   The concept of convergence in probability is used very often in statistics. For example, an estimator is called consistent if it converges in probability to the quantity being estimated.

12.4 SUMMARY

- A 'Limiting Distribution' is also termed as an 'Asymptotic Distribution'. Fundamentally, it is defined as the hypothetical distribution or convergence of a sequence of distributions.
- The asymptotic distribution theory is typically used to find a limiting distribution to a series of distributions.

NOTES

- Some of the limiting or asymptotic distributions are well-recognized and defined by different statisticians.
- In basic statistics, the process includes a random sample of observations and then fitting that data to a known distribution similar to the normal distribution or *t* distribution.
- The accurate approximation is based on the estimation of 'presumptions and guesses' established on the nature of large sample statistics.
- The limiting/asymptotic distribution can be applied on small, finite samples for approximating the true distribution of a random variable.
- Limiting probability distributions are significantly used to find the appropriate sample sizes. When a sample size is large enough, then a statistic's distribution will form a limiting distribution, assuming that such a distribution exists.
- The Central Limit Theorem (CLT) uses the limit concept for describing the behaviour of sample means. The CLT states that the sampling distribution of the sampling means approaches a normal distribution as the sample size increases — irrespective of the shape of the population distribution.
- In simpler terms, it can be stated that the limiting probability distribution of X_n is the limiting distribution of some function of X_n.
- The different possible notions of convergence relate to how such a behaviour can be characterized: two readily understood behaviours are that the sequence eventually takes a constant value, and that values in the sequence continue to change but can be described by an unchanging probability distribution.
- 'Stochastic convergence' validates the notion that a sequence of essentially random or unpredictable events can sometimes be expected to settle into a pattern.
- The probability distribution describing the next outcome may grow increasingly similar to a certain distribution.
- Series formed by calculating the expected value of the outcome's distance from a particular value may converge to 0.
- Variance of the random variable describing the next event grows smaller and smaller.
- The convergence in distribution mode provide the required expectation to recognise and observe the next outcome in a sequence of random experiments which is well modelled by means of a given probability distribution.
- Convergence in distribution is typically considered as the weakest form of convergence, since it is implied by all other types of convergence.

- However, convergence in distribution is very frequently used in practice; most often it arises from application of the Central Limit Theorem (CLT).
- The definition of convergence in distribution may be extended from random vectors to more general random elements in arbitrary metric spaces, and even to the 'random variables' which are not measurable, for example a situation which occurs in the study of empirical processes.
- The basic notion behind this type of convergence is that the probability of an 'unusual' outcome becomes smaller and smaller as the sequence progresses.
- The concept of convergence in probability is used very often in statistics. For example, an estimator is called consistent if it converges in probability to the quantity being estimated.
- Convergence in probability is denoted by adding the letter '*p*' over an arrow indicating convergence, or using the 'plim' probability limit operator:

$$X_n \xrightarrow{p} X, \quad X_n \xrightarrow{P} X, \quad \operatorname*{plim}_{n\to\infty} X_n = X.$$

- 'Convergence in Probability' implies 'Convergence in Distribution'.
- In the opposite direction, convergence in distribution implies convergence in probability when the limiting random variable X is a constant.

12.5 KEY WORDS

- **Limiting distribution:** A 'Limiting Distribution' is also termed as an 'Asymptotic Distribution', it is defined as the hypothetical distribution or convergence of a sequence of distributions.
- **Limiting probability distributions:** The limiting probability distributions are significantly used to find the appropriate sample sizes.
- **Central Limit Theorem (CLT):** The Central Limit Theorem (CLT) uses the limit concept for describing the behaviour of sample means, it states that the sampling distribution of the sampling means approaches a normal distribution as the sample size increases — irrespective of the shape of the population distribution.
- **Stochastic convergence:** Stochastic convergence validates the notion that a sequence of essentially random or unpredictable events can sometimes be expected to settle into a pattern.
- **Convergence in distribution:** Convergence in distribution mode provide the required expectation to recognise and observe the next outcome in a sequence of random experiments which is well modelled by means of a given probability distribution.

12.6 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is limiting distribution?
2. Define the term convergence.
3. What is stochastic convergence?
4. Give definition for convergence in distribution.
5. Define the basic concept of convergence in probability.
6. State the properties of convergence.

Long-Answer Questions

1. Briefly discuss the significance of limiting distributions in the field of probability and statistics.
2. Explain about the convergence of random variables giving appropriate examples.
3. Discuss the concept of convergence in distribution giving definition and appropriate examples.
4. Why the Central Limit Theorem (CLT) uses the limit concept for describing the behaviour of sample means? Explain giving appropriate examples.
5. Analyse and discuss the significance of convergence in probability giving definition and appropriate examples.
6. Is $N(0, 1/n)$ close to the $N(1/n, 1/n)$ distribution? Explain.
7. 'X_n converges in distribution to X'. Justify the statement with appropriate proof.

12.7 FURTHER READINGS

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

NOTES

UNIT 13 LIMITING MOMENT GENERATING FUNCTION AND CENTRAL LIMIT THEOREM

Structure

13.0 INTRODUCTION

In probability theory, the Central Limit Theorem (CLT) establishes that, in some situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution (informally a *bell curve*) even if the original variables themselves are not normally distributed. The theorem is a key concept in probability theory because it implies that probabilistic and statistical methods that work for normal distributions can be applicable to many problems involving other types of distributions.

In this unit, you will study about the limiting moment generating function and central limit theorem.

13.1 OBJECTIVES

After going through this unit, you will be able to:

- Explain about the limiting moment generating function
- Understand the central limit theorem

13.2 LIMITING MOMENT GENERATING FUNCTIONS

Binomial, Poisson, negative binomial and uniform distribution are some of the discrete probability distributions. The random variables in these distributions assume

NOTES

a finite or enumerably infinite number of values but in nature these are random variables which take infinite number of values i.e., these variables can take any value in an interval. Such variables and their probability distributions are known as continuous probability distributions.

A random variable X is the said to be normally distributed if it has the following probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ for } -\infty \le x \le \infty$$

Where μ and $\sigma > 0$ are the parameters of distribution.

Normal Curve: A curve given by,

$$y_x = y_0 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Which is known as the normal curve when origin is taken at mean.

Then, $$y_x = y_0 e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}.$$

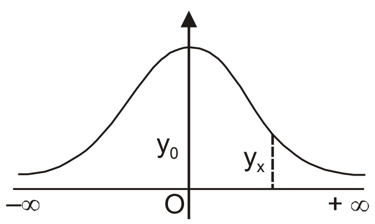


Fig. 13.1 *Normal Curve*

Standard Normal Variate : A normal variate with mean zero and standard deviation unity, is called a standard normal variate.

That is; if X is a standard normal variate then $E(X) = 0$ and $V(X) = 1$.

Then, $X \sim N(0, 1)$

The moment generating function or MGF of a standard normal variate is given as follows:

$$M_X(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2}\Big]_{\substack{\mu=0\\ \sigma=1}} = e^{\frac{1}{2}t^2}$$

NOTES

Frequently the exchange of variable in the integral:

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}$$

Is used by introducing the following new variable:

$$Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$$

This new random variable Z simplifies calculations of probabilities etc. concerning normally distributed variates.

Standard Normal Distribution: The distribution of a random variable $Z = \dfrac{X-\mu}{\sigma}$ which is known as standard normal variate, is called the standard normal distribution or unit normal distribution, where X has a normal distribution with mean μ end variance σ^2.

The density function of Z is given as follows:

$$\phi(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}, -\infty < Z < \infty$$

with mean O variance one of MGF $e^{\frac{1}{2}t^2}$. Normal distribution is the most frequently used distribution in statistics. The importance of this distribution is highlighted by central limit theorem, mathematical properties, such as the calculation of height, weight, the blood pressure of normal individuals, heart diameter measurement, etc. They all follow normal distribution if the number of observations is very large. Normal distribution also has great importance in statistical inference theory.

Examples of Normal Distribution:

1. The height of men of matured age belonging to same race and living in similar environments provide a normal frequency distribution.
2. The heights of trees of the same variety and age in the same locality would confirm to the laws of normal curve.
3. The length of leaves of a tree form a normal frequency distribution. Though some of them are very short and some are long, yet they try to tend towards their mean length.

Example 1: X has normal distribution with $\mu = 50$ and $\sigma^2 = 25$. Find out

(i) The approximate value of the probability density function for $X = 50$

(ii) The value of the distribution function for $x = 50$.

Solution: (*i*) $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, -\infty \le x \le \infty.$$

for $X = 50, \sigma^2 = 25, \quad \mu = 50$, you have

$$f(x) = \frac{1}{5\sqrt{2\pi}} \equiv 0.08.$$

Distribution function $f(x)$

$$= \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu^2)/2\sigma^2}.dx$$

$$= \int_{-\infty}^{\left(\frac{x-\mu}{\sigma}\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} dZ, \text{ where } Z = \frac{x-\mu}{\sigma}$$

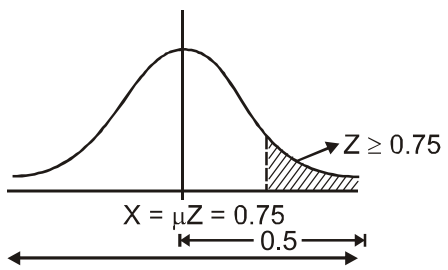
$$\therefore \quad F(50) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}.dZ = 0.5.$$

Example 2: If X is a normal variable with mean 8 and standard deviation 4, find

(i) $P[X \le 5]$

(ii) $P[5 \le X \le 10]$

Solution: (i) $$P[X \le 5] = P\left(\frac{X-\mu}{\sigma} \le \frac{5-8}{4}\right)$$



$= P\ (Z \le -0.75)$

$= P\ (Z \ge\ 0.75)$

[By Symmetry]

$= 0.5 - P(0 \le Z \le 0.75)$

[To use relevant table]

$= 0.5 - 0.2734$ [See Appendix for value of '2']

$= 0.2266.$

NOTES

(ii) $P[5 \leq X \leq 10] = P\left(\frac{5.8}{4} \leq Z \leq \frac{10-8}{4}\right)$

$= P(-0.75 \leq Z \leq 0.5)$

$= P(-0.75 \leq Z \leq 0) + P(0 \leq Z \leq 0.5)$

$= P(-0 \leq Z \leq 0.75) + P(0 \leq Z \leq 0.5)$

$= 0.2734 + 0.1915$

[See Appendix]

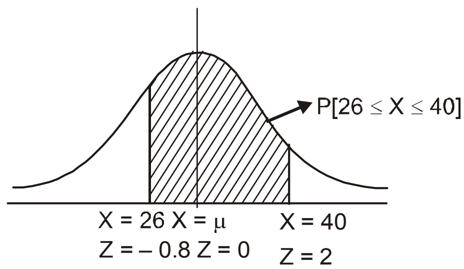
$= 0.4649.$

Example 3: X is a normal variate with mean 30 and S.D. 5. Find

(i) $P[26 \leq X \leq 40]$

(ii) $P[|X - 30| > 5]$

Solution: Here $\mu = 30$, $\sigma = 5$.



(i) When $X = 26$, $Z = \frac{X - \mu}{\sigma} = -0.8$

And for $X = 40$, $Z = \frac{X - \mu}{\sigma} = 2$

$\therefore \quad P[26 \leq X \leq 40] = P[-0.8 \leq Z \leq 2]$

$= P[0 \leq Z \leq 0.8] + P[0 \leq Z \leq 2]$

$= 0.2881 + 0.4772 = 0.7653$

(*ii*) $P[|X - 3| > 5] = 1 - P[|X - 3| \leq 5]$

$P[|X - 3| \leq 5] = P[25 \leq X \leq 35]$

$= P\left(\frac{25-30}{5} \leq Z \leq \frac{35-30}{5}\right)$

$= 2.P\ (0 \leq Z \leq 1) = 0.$

$= 2 \times 0.3413 = 0.6826.$

So $\quad P[|X-3|>5] = 1 - P[|X-3| \leq 5]$

$= 1 - 0.6826 = 0.3174.$

NOTES

13.3 THE CENTRAL LIMIT THEOREM (CLT)

The Central Limit Theorem (CLT) has several variants. In its common form, the random variables must be identically distributed. In variants, convergence of the mean to the normal distribution also occurs for non-identical distributions or for non-independent observations, if they comply with certain conditions.

The earliest version of this theorem, that the normal distribution may be used as an approximation to the binomial distribution, is the de Moivre–Laplace theorem.

Let $X_1, X_2, \ldots. X_n$ be n independent random variables all of which have the same distribution. Let the common expectation and variance be μ and σ_2, respectively.

Let $$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

Then, the distribution of $\bar{X}$ approches the normal distribution with mean m and variance $\frac{\sigma^2}{n}$ as $n \to \infty$

That is, the variate $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ has standard normal distribution.

Proof: Moment generating function of Z about origin is given as follows:

$$M_Z(t) = E(e^{tZ}) = E\left[e^{t\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)}\right]$$

$$= e^{-\mu t\sqrt{n}/\sigma} E(e^{t \times \sqrt{n}/\sigma})$$

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

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

NOTES

$$= e^{-\mu t \frac{\sqrt{n}}{\sigma}} . M(X_1 + X_2 + ... + X_n) . \frac{t}{\sigma\sqrt{n}}$$

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

This is because the random variables are independent and have the same MGF by using logarithms, you have:

$$\log M_z(t) = \frac{-\mu t\sqrt{n}}{\sigma} + n \log M_x \left(\frac{t}{\sigma\sqrt{n}} \right)$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + n \log \left[1 + \frac{\mu'_1 t}{\sigma\sqrt{n}} + \frac{\mu'_2 t}{2!}\left(\frac{i}{\sigma\sqrt{n}}\right)^2 + ... \right]$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + n \left[\left(\frac{\mu'_1 t}{\sigma\sqrt{n}} + \frac{\mu'_2 t}{2!} . \frac{t^2}{n\sigma^2} + ... \right) - \frac{1}{2}\left(\frac{\mu'_1 t}{\sigma\sqrt{n}} + ... \right)^2 + ... \right]$$

$$= \frac{-\mu t\sqrt{n}}{\sigma} + \frac{\mu'_1 t\sqrt{n}}{\sigma} + \frac{\mu'_2 t^2}{2\sigma^2} - \frac{\mu'^2_1 t^2}{2\sigma^2} + ...$$

$$= \frac{t^2}{2} + O(n^{-1/2}) \qquad [\because \mu'_2 - \mu'^2_1 = \sigma^2 \mu'_1 = \mu]$$

Hence, as $n \to \infty$

$$\log(M_z)(t) \to \frac{t^2}{2} \quad \text{i.e.} \quad M_2(t) = e^{t^2/2}$$

However, this is the M.G.F. of a standard normal random variable. Thus, the random variable Z converges to N.

This follows that the limiting distribution of $\sqrt{x}$ as normal with mean μ and variance $\frac{\sigma^2}{n}$.

Check Your Progress

1. What is standard normal variate?
2. Define the term standard normal distribution.

13.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

NOTES

1. A normal variate with mean zero and standard deviation unity, is called a standard normal variate.
2. The distribution of a random variable $Z = \frac{X - \mu}{\sigma}$ which is known as standard normal variate, is called the standard normal distribution or unit normal distribution, where X has a normal distribution with mean μ end variance σ^2.

13.5 SUMMARY

- Binomial, Poisson, negative binomial and uniform distribution are some of the discrete probability distributions.
- A normal variate with mean zero and standard deviation unity, is called a standard normal variate.
- The height of men of matured age belonging to same race and living in similar environments provide a normal frequency distribution.
- The length of leaves of a tree form a normal frequency distribution. Though some of them are very short and some are long, yet they try to tend towards their mean length.

13.6 KEY WORDS

- **Variants:** In variants, convergence of the mean to the normal distribution also occurs for non-identical distributions or for non-independent observations, if they comply with certain conditions.
- **Standard normal variate:** A normal variate with mean zero and standard deviation unity, is called a standard normal variate.

13.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is meant by normal curve?
2. Give the density function of Z.
3. Write some examples of normal distribution.

Long-Answer Questions

1. Briefly discuss about the limiting moment generating functions.
2. Explain about the moment generating function of a standard normal variate.
3. Discuss the central limit theorem. Give examples.

13.8 FURTHER READINGS

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 14 SOME THEOREMS ON LIMITING DISTRIBUTIONS

Structure

14.0 INTRODUCTION

A general name for a number of theorems in probability theory that give conditions for the appearance of some regularity as the result of the action of a large number of random sources. The first limit theorems, established by J. Bernoulli (1713) and P. Laplace (1812), are related to the distribution of the deviation of the frequency μ_n/n of appearance of some event EE in n independent trials from its probability p, 0<p<10<p<1(exact statements can be found in the articles Bernoulli theorem; Laplace theorem). S. Poisson (1837) generalized these theorems to the case when the probability p_k of appearance of E in the k- th trial depends on k, by writing down the limiting behaviour, as $n \to \infty$, of the distribution of the deviation of μ_n/n from the arithmetic mean $\overline{p} = \left(\sum_{k=1}^{n} p_k\right)/n$ of the probabilities p_k, $1 \le k \le n$ (cf. Poisson Theorem).

In this unit, you will study about the some theorems of limiting distributions.

14.1 OBJECTIVES

After going through this unit, you will be able to:

- Analyse the some theorems of limiting distributions
- Understand some laws of large numbers

14.2 SOME THEOREMS ON LIMITING DISTRIBUTIONS

NOTES

Before starting the laws of large numbers, let us define an inequality named as Kolmogorov's inequality. The set of Kolmogorov's inequalities was defined by Kolmogorov in 1928.

Suppose $X_1, X_2, ..., X_n$ is a set of independent random variables having mean O and variances $\sigma_1^2, \sigma_2^2, ... \sigma_n^2$.

Let $\qquad C_n^2 = \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2.$

Then the probability that all of the inequalities

$| x_1 + x_2 + ... + x_\alpha| < \lambda Cn, \alpha = 1, 2, ...n$ hold, is at least $\left(1 - \frac{1}{\lambda^2}\right)$.

14.2.1 Weak Law of Large Numbers

If you put, $x = \dfrac{x_1 + x_2 + ... + x_n}{n}, \bar{x} = \dfrac{\overline{x_1} + \overline{x_2} + ... + \overline{x_n}}{n}$

And $\lambda = \sqrt{\dfrac{En}{n^2}} = \alpha.$

where $\propto$ is an arbitrary positive number and En is the mathematical expectation of the variate

$$u = (x_1 + x_2 + ... + x_n - \overline{x_1} - \overline{x_2} - ... - \overline{x_n})^2$$

i.e. $\qquad En = E\,[(x_1 + x_2 + ... + x_n - \overline{x_1} - \overline{x_2} - ... - \overline{x_n})^2]$

Then, $\qquad P\left[\left|\dfrac{x_1 + x_2 + ... + x_n}{n} - \dfrac{\overline{x_1} - \overline{x_2} - ... - \overline{x_n}}{n}\right| < \alpha\right]$

$$\geq 1 - \frac{En}{n^2\,\alpha^2} = 1 - \eta \text{ provided } \frac{En}{n^2\,\alpha^2} < \eta\,.$$

This is known as the Weak Law of Large Numbers. This can also be stated as:

With the probability approaching unity or certainty as near as you please, you can expect that the arithmetic mean of the values actually assumed by n variates will differ from the mean by their expectations by less than any given number however small, provided the number of variates can be taken sufficiently large and provided the condition $\dfrac{E_n}{n^2} \to 0\ as \to \infty$ is fulfilled.

NOTES

In other words, a sequence $X_1, X_2, \ldots X_n$ of random variables is said to satisfy the weak law of large numbers if

$$\lim_{n \to \infty} P\left[\left|\frac{Sn}{n} - E\left(\frac{Sn}{n}\right)\right| < \varepsilon\right] = 1$$

for any $\varepsilon > 0$ where $Sn = X_1 + X_2 + \ldots + X_n$. The law holds provided $\frac{\beta_n}{n^2} \to 0$ as $n \to \infty$ where Bn = Var. $(Sn) < \infty$.

14.2.2 Strong Law of Large Numbers

Consider the sequence $X_1, X_2, \ldots X_n$ of independent random variables with expectation $\overline{x_k}$ or $\mu_k = E(X_k)$ and variance σ^2.

If $Sn = X_1 + X_2 + \ldots + X_n$ and $E(S_n) = m_n$ then it can be said that the sequence S_1, S_2, ... S_n obeys the strong law of large numbers if every pair $\varepsilon > 0$, $\delta > 0$ corresponds to N such that there is a probability $(1 - \delta)$ or better that for every $r > 0$ all $r + 1$ inequalities,

$$\frac{|S_n - m_n|}{n} < \varepsilon$$

$n = N, N + 1, \ldots N + r$ will be satisfied.

Example 1: Examine whether the weak law of large numbers holds for the sequence $\{X_k\}$ of independent random variables defined as follows:

$$P(X_k = \pm 2^k) = 2^{-(2k+1)}$$

$$P(X_k = 0) = 1 - 2^{-2k}$$

Solution:

$$E(X_k) = \Sigma X_k p_k$$
$$= 2^k \times 2^{-(2k+1)} + (-2^k) \times 2^{-(2k+1)} + 0 \times (1 - 2^{-2k})$$
$$= 2^{-(2k+1)} [2^k - 2^k] = 0.$$

$$E(X_k^2) = \Sigma x_k^2 \cdot p_k$$
$$= (2^k)^2 \times 2^{-(2k+1)} + (-2^k)^2 \times 2^{-(2k+1)} + 0^2 \times (1 - 2^{-2k})$$
$$= 2^{-(2k+1)} [2^{2k} + 2^{2k}] = 2^{-1} + 2^{-1} = 1.$$

$$\therefore \quad \text{Vari}(X_k) = E(X_k^2)\ [E(X_k)]^2 = 1 - 0 = 1$$

$$Bn = \sum_{i=1}^{n} Var\ (X_k) = \sum_{i=1}^{n} 1 = n$$

$$\therefore \quad \lim_{n \to \infty} \frac{Bn}{n^2} = \lim_{x \to \infty} \frac{n}{n^2} = \lim_{x \to \infty} \frac{1}{n} = 0.$$

NOTES

Hence, weak law of large numbers holds for the sequence $\{X_k\}$ of independent random variables.

Example 2: Examine whether the laws of large numbers holds for the sequence $\{X_k\}$ independent random variables which are defined as follows:

$$P(X_k = \pm k^{-1/2}) = \frac{1}{2}$$

Solution:

$$E(X_k) = \Sigma X_k p_k$$

$$= k^{-\frac{1}{2}} \times \frac{1}{2} + (-k^{-\frac{1}{2}}) \times \frac{1}{2} = 0$$

$$E(X_k^2) = \Sigma x_k^2 \,.\, p_k = (k^{-1/2})^2 \times \frac{1}{2} + (-k^{-1/2})^2 \times \frac{1}{2}$$

$$= \frac{1}{2} \times k^{-1} + \frac{1}{2} k^{-1} = k^{-1}$$

$$\text{Var.}\,(X_k) = E(X_k^2) - [E(X_k)]^2 = k^{-1} - 0 = k^{-1}$$

$$Bn = \sum_{i=1}^{n} \text{Var}\,(X_k) = \sum_{i=1}^{n} k^{-1} = n\,.\,k^{-1}$$

$$\therefore \quad \lim_{n \to \infty} \frac{Bn}{n_2} = \frac{n\,.\,k^{-1}}{n^2} = \lim_{n \to \infty} \frac{k^{-1}}{n} = 0$$

Hence, the laws of large numbers hold for the sequence $\{X_k\}$ of independent random variables.

Check Your Progress

1. Who defined the set of Kolmogrov's inequalities and when?
2. Under what circumstances normal approximation can be applied to binomial and Poisson distribution.

14.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The set of Kolmogorov's inequalities was defined by Kolmogorov in 1928.
2. When number of trials is large and probability p is close to 1/2, normal approximation can be used to for binomial as well as Poisson distribution.

14.4 SUMMARY

- The set of Kolmogorov's inequalities was defined by Kolmogorov in 1928.

- Suppose $X_1, X_2, ..., X_n$ is a set of independent random variables having mean O and variances $\sigma_1^2, \sigma_2^2, ... \sigma_n^2$.

  Let $C_n^2 = \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2$.

  Then the probability that all of the inequalities

  $|x_1 + x_2 + ... + x_\alpha| < \lambda Cn$, $\alpha = 1, 2, ...n$ hold, is at least $\left(1 - \frac{1}{\lambda^2}\right)$.

- A sequence $X_1, X_2, ... X_n$ of random variables is said to satisfy the weak law of large numbers if

$$\lim_{n \to \infty} P\left[\left|\frac{Sn}{n} - E\left(\frac{Sn}{n}\right)\right| < \varepsilon\right] = 1$$

14.5 KEY WORDS

- **Strong law of large numbers:** If $Sn = X_1 + X_2 + ... + X_n$ and $E(S_n) = m_n$ then it can be said that the sequence $S_1, S_2, ... S_n$ obeys the strong law of large numbers

14.6 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define Kolmogorov's inequality.
2. What is strong law of large numbers?

Long-Answer Questions

1. Briefly describe the weak law of large numbers.
2. Discuss about the strong law of large numbers.

14.7 FURTHER READINGS

Hogg, Robert. V., Allen T. Craig and Joseph W. McKean. 2005. *Introduction to Mathematical Statistics*. New Delhi: Pearson Education.

Fisz, Marek. 1963. *Probability Theory and Mathematical Statistics*, 3rd Edition. New York: John Wiley & Sons, Inc.

Rohatgi, Vijay K. 1985. *An Introduction to Probability Theory and Mathematical Statistics*. New Delhi: Wiley Eastern Limited.

NOTES

Gupta, C.B. and Vijay Gupta. 2004. *An Introduction to Statistical Methods,* 23rd Edition. New Delhi: Vikas Publishing House.

Levin, Richard I. and David S. Rubin. 1998. *Statistics for Management*. New Jersey: Prentice Hall.

Chandan, J. S. 1998. *Statistics for Business and Economics*. New Delhi: Vikas Publishing House Pvt. Ltd.

Campbell, Stephen K. 1987. *Applied Statistics*. New York: Harper & Row.

Chance, William A. 1969. *Statistical Methods for Decision Making*. Illinois: Richard D Irwin.

Hooda, R. P. 2013. *Statistics for Business and Economics*, 5th Edition. New Delhi: Vikas Publishing House Pvt. Ltd.